2016-06-01

Student Growth Trajectories with Summer Achievement Loss Using Hierarchical and Growth Modeling

Sara Bernice Chapman
Brigham Young University

Student Growth Trajectories with Summer Achievement Loss

Using Hierarchical and Growth Modeling

Sara Bernice Chapman

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard R Sudweeks, Chair
Ross Larsen
Robert V. Bullough
Bryant T. Jensen
Joseph A. Olsen
Lane Fischer

Educational Inquiry, Measurement, and Evaluation

Brigham Young University

June 2016

ABSTRACT

Student Growth Trajectories with Summer Achievement Loss
Using Hierarchical and Growth Modeling

Sara Bernice Chapman
Educational Inquiry, Measurement, and Evaluation, BYU
Doctor of Philosophy

Using measures of student growth has become more popular in recent years—especially in the context of high stakes testing and accountability. While these methods have advantages over historical status measures, there is still much evidence to be gathered on patterns of growth generally and in student subgroups. To date, most research studies dealing with student growth focus on the effectiveness of specific interventions or examine growth in a few urban areas. This project explored math, reading, and English language arts (ELA) growth in the students of two rural school districts in Utah. The study incorporated hierarchical and latent growth methods to describe and compare these students' growth in third, fourth and fifth grades. Additionally, student characteristics were tested as predictors of growth.

Results showed student growth as complex and patterns varied across grade levels, subjects and student subgroups. Growth generally declined after third grade and students experienced summer loss in the second summer more than the first. Females began third grade ahead of their male peers in ELA and reading and began at a similar level in math. Male students narrowed the gap in reading and ELA in fourth and fifth grade and pulled ahead of their female peers in math in third grade. Low SES students were the most similar to their peers in math and ELA growth but were ahead of their peers in reading. Hispanic and Native American students started consistently behind white students in all subjects. Hispanic students tended to grow faster during the school year but lost more over the summer months. Native American students had more shallow growth than white students with a gradual decline in growth in fourth and fifth grades. ELA and reading growth were more closely related to each other than with math growth. Initial achievement estimates were more highly correlated with subsequent growth than previous years' growth. A cross-classified model for teacher-level effects was attempted to account for students changing class groupings each school year but computational limits were reached. After estimating subjects and grade levels separately, results showed variance in test scores was primarily due to student differences. In ELA and reading, school differences accounted for a larger portion of the overall variance than teacher differences.

Keywords: student growth, teacher effects, principal effects, hierarchical linear modeling, growth modeling, summer loss

ACKNOWLEDGEMENTS

It has taken many hands and hearts to accomplish this work, starting with my committee. I would like to first thank my chair, Dr. Richard Sudweeks, for his continued enthusiasm for this project, for his example of true scholarship, for his constant guidance, and for his confidence in me. This journey would not have even started, let alone finished, without your encouragement. Many thanks to Dr. Joseph Olsen for the many hours he spent with me slaving over messy data and data analyses to push the boundaries of this research. My gratitude also goes to Dr. Lane Fischer for his profound example in teaching and learning, and for his wise counsel to "never do educational research," even though I did not listen. I also thank Dr. Bryant Jensen for his mentoring and for his contagious passion that helped me stay focused on the purpose of this research—to help students. I would also like to thank Dr. Robert Bullough for pushing me to test the limits--my own limits as well as the limits of what we know and accept in our world. And to Dr. Ross Larsen, many thanks for the added rigor and thoughtful suggestions you brought to this project.

Of course, I also express gratitude for those in my life who have supported my efforts along this journey and encouraged me to finish. The list of names would be too long to include here, but I would especially like to thank my parents, siblings, mentors, coworkers, and friends who continually gave feedback and encouragement throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

School administrators and other interested stakeholders have long relied on standardized test scores to measure students' progress in achieving state-mandated curriculum standards. Policies at federal, state and local levels have typically used student achievement information to inform high-stakes student-level decisions such as student graduation or grade retention. In more recent years, education policy at various levels of government has supported the use of student achievement scores at teacher and administrative levels as an indicator of the effectiveness of schools and school systems as a whole (e.g., National Assessment of Educational Progress). Recent research efforts have focused on establishing evidence for or against the use of achievement results for these purposes.

Status Measures

Historically, tests have provided a glimpse into students' performance relative to their peers and curriculum standards with which the test was assumed to align. These so called "status measures" (Betebenner, 2011, p. 2) provided a snapshot of student achievement derived from a single data point from one test administration towards the end of the school year. Students' scores were then categorized as proficient, above proficient, or below proficient in relation to predetermined cut scores. Over time several pitfalls have surfaced from the widespread use of status measures to describe students' academic progress.

Among the problems with the reliability of a measure with only one data point is what Harris (2011) describes as a reliance on a "single snapshot" (p. 28). He argues that one measurement point is insufficient to reliably measure a student's academic progress since testing experiences can be influenced by so many external and unrelated factors generally. Test performance is assumed to be influenced primarily by the student's level of knowledge or competence in the construct being measured, but it can also be influenced by hunger, fatigue, or other distractions unrelated to test content. For example, a student who takes a test while hungry or tired may not do as well as when he or she has eaten and is rested, or a student may be distracted during a test by a recent adverse life event. With only one snapshot, such sources of error are confounded with the measure of a student's ability level.

Status measures also fail to consider what level of proficiency or ability a student started from at the beginning of the academic year (Harris, 2011). Ignoring this "starting-gate" problem (Harris, 2011, p. 24) perpetuates a false assumption that all students within a grade start the school year with the same content and background knowledge. In reality, some students may have already learned some of the current year's content while others may be struggling to master content from the previous year upon which the current year's content may build. This issue is particularly relevant to underperforming students who are generally behind at the beginning of the year and who must catch up. The starting gate problem means underperforming students are unduly penalized from the start and face inequitable and often unreasonable expectations in order to improve their status.

Using status measures for high-stakes decisions like teacher and school accountability was seen as unfair based on these limitations as well. In addition to the error introduced based on a single measurement of student performance, teachers and schools would be held accountable for student factors that were out of their control. Regardless of where students began the school year or what learning limitations they may face, teachers and schools would be held responsible for the growth of their students to meet proficiency standards.

Growth Measures

In response to these and similar issues with status measures, growth measures have become accepted as an alternative way to describe student progress (Goldschmidt, Choi, & Beaudoin, 2012; Harris, 2011; Holt, 2006). In 2009, student growth models became particularly important with the announcement of the Race to the Top competition sponsored by the federal government, which shifted the focus from student achievement levels to student growth. Race to the Top also emphasized the use of student growth in evaluating teacher performance. Most states, including Utah (Senate Bill 64), have incorporated growth measures when assessing student achievement. States also use growth measures in when making decisions about teacher and principal effectiveness and pay.

The use of growth models allows for the inclusion of multiple data points in an effort to describe change over time. This approach addresses the snapshot issue by describing students' progress through academic content over time. The majority of growth models (e.g., value added models) also address the starting-gate weaknesses of status measures by incorporating previous achievement into the models. Students' trajectories as described through growth models include elements of both initial achievement and growth over time.

Achievement growth over time can be calculated in many ways. At the simplest level, gain scores can be computed as the difference between subsequent scores. The use of gain scores is limiting based on the necessity of comparable pre- and post-tests. In addition, a simple difference between subsequent scores does not take into account potential external influences on score differences. Score changes attributable to tragic life events such as the death of a close relative, or factors external to the classroom such as the influence of a tutor or supplemental educational program, should be accounted for, especially when growth can be tied to high stakes decisions. Therefore computing simple gain scores as an indication of growth is problematic, especially in accountability contexts such as teacher effectiveness.

In settings where teacher accountability is emphasized, value-added models (VAMs) have been implemented at a policy level in many states and large cities. In many places, student growth percentiles (SGPs) have been adopted as an alternative method for estimating growth. Although SGPs were not originally designed to be used for accountability purposes, they are used for that purpose. Since the algorithm used in calculating SGPs is protected, the "black box" of SGP is less useful for research purposes in describing and comparing student trajectories. In addition, the data input process in SGP analysis provides little added flexibility in incorporating theorized variables or factors that may influence the starting achievement or growth of achievement of students over time. While these methods are considered an improvement over reporting status measures, the use of these approaches in teacher and school accountability is still heavily debated since estimating teacher and principal effects with any confidence requires a large number of variables which are measured with impressive accuracy (American Educational Research Association, 2016).

Imbedded in the challenge of improving education is the challenge of balancing the data needs for accurate measurement versus practical means for collecting and using the data. For example, more data is often useful in establishing reliability of results; however collecting more data is often costly, difficult, infeasible, or even at times unethical. The estimation of student achievement growth over time requires the use of more data points over time, which can be difficult to collect. This study makes use of a unique dataset with more frequent data points to further validate what we already claim to know about the growth of student achievement while pushing the limits of common statistical practice in this context. With added data points, we can examine on a deeper level patterns of growth in third through fifth grades in math, reading and English language arts (ELA), and can bring to the forefront methodological questions in using multilevel growth modeling in real educational contexts. Such a dataset is unusually rich and is unlikely to be replicated based on the practical and political implications of such wide-spread and frequent test administrations.

While growth measures and value-added models are increasingly used for high stakes decision making, there is still much that needs to be understood about patterns of student growth and the factors that influence both initial achievement and achievement growth over time. This study seeks to contribute to the ongoing conversation on student growth through two main goals. The first goal is to add empirical evidence on the patterns of growth over time for different subgroups of students to this growing body of literature. Differences in students' growth trajectories will be compared by student demographic variables such as gender, ethnicity and free or reduced lunch status. The second focus of this study is to contribute to conversations on methodological issues that arise in estimating growth trajectories. In this effort, commentary will be specifically related to issues in hierarchical linear modeling approaches to measuring student growth.

Statement of Purpose

The purpose of this study was to describe the achievement trajectories of elementary school students in two rural Utah school districts on three dependent variables: math, reading, and language. The dataset included scores from three test occasions within each school year in grades 3 through 5. This study focused on the following research questions.

1. How do student growth trajectories change across measurement occasions within years and across years?

   a. What proportion of variation in student achievement at the time students enter the third grade is associated with differences between classrooms?
   b. How do the intercepts and slopes of growth trajectories for the same sample of students vary between grade levels (third, fourth, and fifth)?
      i. What evidence is there that those with the largest gains in the first year tend to have large gains in subsequent years, and those with smallest gains in the first year will have the smallest gains in subsequent years?
      ii. How are students' learning rates correlated across school subjects (math, reading, and English language arts)?
   c. What may be the advantages of using three measures per year rather than one annual spring score to model growth trajectories?
2. How do students' growth trajectories vary across ethnic, gender, and socioeconomic subgroups?
3. To what extent does summer learning loss vary between grade levels?
   a. To what extent is the summer loss for individual students associated with the slopes and intercepts of their growth trajectory for the previous academic year?
   b. How do the summer loss patterns vary across the three school subjects?
   c. How do the summer loss patterns vary across demographic subgroups?
4. What are the advantages and challenges of modeling the nested student-teacher structure using cross-classified effects?
5. What are the advantages and disadvantages of using a multivariate versus a univariate approach to multilevel modeling in the context of the data structure available in this study?

Chapter 2: Literature Review

The literature regarding student achievement growth can be summarized by two main bodies of literature. The first includes studies conducted on student achievement growth in the last 20 years. Searches in ERIC, PsychInfo, PsychArticles and Google Scholar included keywords such as *academic achievement*, *mathematics achievement*, *reading achievement*, *educational assessment*, *achievement tests* in combination with keywords such as *longitudinal studies, growth*, *hierarchical linear modeling* and *structural equation modeling*. The resulting studies cover a variety of content areas with added variety in the demographic covariates used in the model. The studies included in this review were selected based on their relevance to the current study and dataset.

The second body of literature focuses on studies dealing with summer achievement loss as an important component of multi-year growth estimation for students. Searches in ERIC and Google Scholar included keywords such as *summer achievement loss, summer slide,* and *summer loss*. These studies show the importance of incorporating summer setback in growth models. Both bodies of literature contain important insights into current methodologies and theories associated with student achievement growth.

Growth Modeling

Growth modeling is used to describe student growth trajectories and predict patterns in student achievement growth based on demographic or other variables of interest without attributing causality (Briggs, 2011). Several statistical methods can be used in estimating growth models to describe change over time. Three of the most common methods used in recent literature are (a) hierarchical linear modeling, (b) latent growth curve modeling, and (c) growth mixture modeling.

Hierarchical Linear Modeling (HLM; Bryk & Raudenbush, 1988) comes from a regression framework and accounts for correlated error within nested contexts. In educational research, HLM is used extensively to account for the common variance within classrooms or schools. HLM is also used in longitudinal contexts to account for the common variation within participants across time. For example, a student's math scores will show distinct patterns over time when compared with another student's scores. Students within a classroom can have common variance in their scores based on the similar classroom setting they were in. Additionally, students' scores are likely to be more similar within a school than across different schools.

HLM has some analytic advantages. Since the approach is based on regression, it is conceptually intuitive to many researchers and readers (Holt, 2008). From a more technical perspective, HLM is considered to be a more robust approach when the data have missing or unbalanced data (Kline, 2011; Singer & Willet, 2003). In addition, HLM models are well equipped to estimate specified fixed effects of time-varying, person-varying, group-varying variables at each level as well as the specified random effects.

Traditional HLM analysis is typically a linear, univariate approach. Data from real world contexts in education show evidence that there are very few instances where data are linear, and multivariate analysis is becoming more popular. A variety of approaches to HLM have been used to compensate for nonlinear trends, multivariate outcomes, and incomplete nesting.

Since developmental processes are rarely linear (Grimm, Ram & Hamagami, 2011), quadratic or piecewise models are often constructed using manipulations of the time variable (e.g., months, years, age) to approximate a nonlinear shape. A quadratic variable is often incorporated into HLM models and tested for significance as a test of linearity. For nonlinear trajectories with multiple curves, higher-order polynomial term can be used (i.e., cubic or quartic), or a discontinuous or piecewise model can be estimated. A discontinuous model allows for a break or knot in the trajectory which can account for a sudden shift in growth or different developmental phases. These shifts in developmental phases can be modeled as a constant across participants, or can be estimated freely for each participant in more complex multilevel modeling (Hoffman, 2015).

In many cases, nesting is not as clean in real world contexts as it is in theory. For example, students often switch classrooms or schools within or between school years. Typically, research studies exclude participants who do not follow a clean nesting pattern (e.g., students who switch schools or teachers); however, student mobility is a large and growing phenomenon, especially in some subgroups (Grady & Beretvas, 2010). Failing to model mobile students can affect the estimates and limit the generalizability of conclusions (Grady & Beretvas, 2010). Cross-classification or multiple membership models can be estimated to account for these and similar issues, although these approaches are not common in most research studies.

Multiple membership is used to model data where lower level units (i.e., students) belong to more than one higher level cluster (i.e., classrooms or schools). Multiple membership is largely ignored in recent literature (Grady & Beretvas, 2010). Cross-classified models occur in three-level data where the two higher levels are not nested within each other. For example, graduating high school seniors from the same high school are likely to attend different colleges the following year. Cross-classification is also rarely used in educational research studies.

Latent growth curve modeling (LGCM) can also be addressed from a structural equation modeling framework. The strengths of LGCM lie in its ability to estimate latent variables (e.g., math or reading ability) and associated error. In the context of growth modeling, latent variables are typically the initial status and the subsequent change over time of a participant. The scores for each data collection event is driven or influenced by the latent starting point and the latent growth estimate.

As an extension of structural equation modeling, LGCM is well equipped to estimate and preserve measurement error in the analyses. In addition, taking an SEM approach allows for the estimation overall model-fit indices such as RMSEA, CFI, TLI and others.

An LGCM approach in longitudinal growth modeling has added flexibility in estimating nonlinear effects (Grimm et al., 2011) and estimating the relationships between predictor, latent, and growth variables within a model (Newsom, 2015). LGCM can also easily model multivariate data.

Growth mixture modeling (GMM) is an extension of the SEM and LGCM framework used to estimate differences in growth among subgroups of participants. As in LGCM, latent intercept and slope variables are modeled to estimate growth trajectories of participants. In addition, a predetermined number of latent classes are modeled. These latent classes or subgroups are empirically derived, and the results describe the differences in averages between groups on the variables modeled.

GMM has similar advantages to LGCM. Since GMM also comes from an SEM framework, measurement models can be included to estimate measurement error, improving the precision of results. In addition to the flexibility of nonlinear and multivariate analyses added in LGCM, GMM has added flexibility in assumptions of normality (Holt, 2008). One disadvantage in using GMM is the inherent complexity of the model, which increases the probability of convergence problems (Hipp & Bauer, 2006).

Overall, both SEM and HLM approaches produce similar parameters to describe growth trajectories (Kline, 2011). Each has a different underlying framework making some aspects of modeling more intuitive in one approach than the other. For example, HLM has a more intuitive nesting component, but SEM can also account for the correlated errors associated with nesting. Both approaches are appropriate to use in modeling growth, and both can be used to describe growth trajectories in very similar ways.

Advances in the statistical complexity of both approaches allows for better accuracy in modeling growth over time as measured by model fit statistics. Often model-data fit is most dramatically improved by modeling random effects, which can be done through HLM or SEM techniques. Modeling the random effects has little to no effect on the estimates of fixed effects (Signer & Willett, 2003); however, and modeling with increasing complexity requires more data and increases the complexity of interpretation (Grimm et al., 2011). Therefore, studies on growth rarely focus on the random effects of growth trajectories.

Modeling Growth in an Educational Context

The two educational outcomes variables most commonly investigated in current research are reading and mathematics achievement. In the last 20 years, the studies addressing student achievement trajectories have focused mainly on reading achievement growth--especially for immigrants and English language learners. Studies compare the growth of students from different language backgrounds (Roberts, Mohammed & Vaughn, 2010), or English proficiency levels (Guglielmi, 2012). Language impairments have been shown to influence initial achievement levels but have not been shown to be sources of significant differences in growth (Catts, Bridges, & Little, 2008). Other studies also have focused on other types of learning disabilities as factors affecting student growth trajectories (Judge & Watson, 2011). Fewer studies in the past 20 years of achievement studies have dealt with factors that influence math achievement growth trajectories.

Studies most commonly include some basic influential demographics such as gender, ethnicity, or socioeconomic status in addition to various socioemotional, academic or intervention variables of interest.

Gender is usually included as a covariate and tends to be a significant predictor of academic growth. Gender is believed to be more of a factor for immigrants (Suarez-Orozco & Gaytán, 2010) and at earlier ages (McCoach, O'Connell, Reis, & Levitt, 2006). For example, males are believed to accelerate in math faster than females in first through third grades (Holt, 2006; Judge & Watson, 2011) although evidence suggests both male and female students start at comparable levels in math when entering Kindergarten (Judge & Watson, 2011). Females are expected to show higher achievement at the beginning of Kindergarten and steeper growth during their Kindergarten year (McCoach et al., 2006).

Strong evidence shows ethnicity influences growth; however, studies focus mainly on Hispanic or Black minority groups (Judge & Watson, 2011; McCoach et al., 2006). Only a few studies look at growth trends for Asian and Pacific Islander or American Indian and Alaska Native students.

The achievement gap between Black and white students is a commonly known and discussed issue in education. Some disagreement exists of when the gap in mathematics starts (Entwisle & Alexander, 1992; Judge & Watson, 2011; McCoach et al., 2006); however, there is agreement that the gap grows in early elementary grade levels as the growth of Black students is slower than white students (McCoach et al., 2006). Black students perform around .5 standard deviations below white students by third grade (Entwisle & Alexander, 1992).

A study on children of immigrants by Han (2008) suggests that children of immigrants begin kindergarten with lower scores in math and reading than their white peers. In both reading and math, children from Latin American origins have steeper growth than white children while East Asian and Indian students have steeper growth in math but decreased growth in reading. The study found that Latin American, East Asian and Indian students are generally achieving at similar levels with their white peers by the end of third grade. Judge and Watson (2011) and McCoach et al. (2006) confirmed that Asian and Pacific Islander students, on average, have a faster rate of growth in math than their white peers. However, the findings from the study by McCoach et al. (2006) conflicted with the Han (2008) results suggesting that Asian students start Kindergarten with higher reading scores than white students. This disparity may be due to the fact that the sample included in the study by Han (2008) was limited to students with immigrant parents. The McCoach et al. (2006) study makes no distinction in their sample of students of Asian ethnicity. The proximity of the students to their family's immigration may influence mediating factors that would in turn influence achievement.

American Indian and Alaska Native (AIAN) students are underrepresented in longitudinal achievement studies generally. Marks and Garcia Coll (2007) conducted a study showing that AIAN students are more likely to live in poverty and in rural areas than other minority students and from parents with lower educational achievement. These students start Kindergarten with lower reading and math achievement than white students, but are comparable to Hispanic and Black students. Poverty and parental education were the largest predictors of achievement and rurality was a larger influence on academic achievement of AIAN students than other minority students. The gap between AIAN students and their white peers grew from kindergarten through third grade.

Students' socioeconomic status is also a predictor of achievement growth. Family income or eligibility for free or reduced lunch are most commonly used as indicators of SES. For studies using national datasets, other family background variables such as parental education or English language proficiency can be combined with SES indicators to create a summary of family or home environment (Guglielmi, 2012; Johnson, McGue, & Iacono, 2006).

Although recent research continues to investigate the role of SES as a predictor of growth or slope parameters in growth models, SES has been consistently significant as a predictor of differential starting achievement for students (Judge & Watson, 2011; Roberts et al., 2010). In addition, studies show that SES is one of the largest predictors of variance in achievement growth (Entwisle & Alexander, 1992; McCoach et al., 2006; Roberts & Bryant, 2011; Roberts et al., 2010). Kieffer (2012) used piecewise latent growth modeling to detect growth differences before and after third grade. Growth analyses showed that students from low SES backgrounds experienced their highest reading growth before third grade and a slower rate after third grade compared with other students from higher SES backgrounds. Based on this and other similar findings (McCoach et al., 2006), SES may be a greater predictor of achievement in earlier grades.

Some studies investigate the interaction effects between initial achievement and achievement growth. The Matthew effect is a common theory that the rich get richer while the poor get poorer, meaning those students who start with high achievement are more likely to experience high growth (Bast & Reitsma, 1997). Empirical studies show inconclusive evidence for and against the Matthew effect (Baumert, Nagy, & Lehmann, 2012; Kim, Petscher, Schatschneider, & Foorman, 2010; Mancilla-Martinez & Leaux, 2000; McCoach et al., 2006). Additionally, the Matthew effect may be more prevalent in specific contexts. For example, the correlation between initial achievement and subsequent growth is larger for minority groups than for white students (Marks & Garcia Coll, 2007). Holt (2006) argues that the evidence for a Matthew Effect in early grades may be attributable to the floor effect mentioned earlier, meaning that there is naturally less variation possible in Kindergarten than in third grade (or higher).

Summer Loss

In addition to the factors that influence the change of student achievement over the course of a year or more, considerations should be given to the factors that influence where students start academically at the beginning of a school year. While summer breaks were historically an effort to accommodate a mainly agrarian society, evidence suggests that a large portion of school-aged children are not using summers well and even lose academic standing (Cooper, 2004; Entwisle & Alexander, 1992; Von Drehle, 2010). From the early 20th century, studies have shown that students of all ages lose knowledge and understanding over summer breaks (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Therefore, in trying to accurately describe student growth, excluding summer loss in a growth model can produce biased estimates of within-school year growth (Kim et al., 2010).

Several studies show that summer loss is income sensitive with students from lower income backgrounds losing more academic footing than their middle or higher SES peers and a major contributor to the achievement gap (Alexander, Entwisle, & Olson, 2007; Allington et al., 2010; Burkam, Ready, Lee, & LoGerfo, 2004; Chaplin & Capizzano, 2006; Cooper, 2004; Downey, von Hippel, & Broh, 2004; Entwisle & Alexander, 1992). Cooper et al. (1996) suggested an average fall of .1 standard deviations between spring and fall measures, although actual summer loss may be greater since measures were rarely administered the last and first days of the school year. Summer loss was reported to be more dramatic in math than in reading, which was attributed to a combination of more readily available reading practice through summer vacation as well as potential memory loss of factual or conceptual knowledge more than procedural knowledge.

Additionally, Cooper et al. tested for moderating variables that influence the degree of summer loss such as IQ, gender, grade and socio-economic status. While the combined evidence of several studies were inconclusive for the influence of IQ, one study described greater summer loss for students eligible for special education services. Gender and race were not significant moderators; however, higher grades (fourth and fifth grades) experienced a greater summer loss than younger grades (second or third grades). Cooper et al. suggest the difference between grades as a byproduct of a “floor effect” (p 263) limiting the possible variance of achievement scores in younger grades.

Differences in summer loss by socio-economic status were well substantiated by Cooper et al. Family income level seemed to moderate reading and language achievement loss, but did not suggest differential loss in math. Students from lower-class backgrounds showed greater losses in reading and language than middle-class students who even showed gains in some areas. This finding exacerbated the achievement gap between lower and middle classes. Cooper et al. argued that reading may take an especially hard hit for lower-class students as they may have less access to reading materials and reading practice over the summer months than middle-class students. Math achievement, however, may not show such large differences between classes because few students practice math processes and concepts during summer vacation even though other studies included in the literature review found that math achievement loss was greater for low-income students.

Since Cooper et al.'s (1996) meta-analysis, several other studies have examined summer loss. Many try to resolve some of the discrepancies that surfaced in the meta-analysis while others expand on the work that was done previously. Generally, studies agree that students from differing income level backgrounds tend to experience summer loss differently. These studies use national datasets as well as researcher-collected data from local cities to illustrate that reading scores are particularly susceptible to differential loses by socio-economic status.

Burkam et al. (2004) studied summer loss in literacy, math and general knowledge between SES quintiles and found significant summer growth differences specifically between SES extremes—the highly advantaged students and exceptionally less advantaged students—when compared with median SES students. Students from high SES backgrounds actually grew slightly over the summer while lower SES students experienced significant achievement loss. These patterns were observed in reading, math and general knowledge of the Kindergarten and first graders included in the national dataset.

Downey et al. (2004) also looked at student growth during the school year in comparison with growth over summer months. Their research highlights the finding that growth during the school year is relatively comparable between groups but that achievement gaps appear as groups grow differentially over the summer. By their estimation, schools are doing a good job of equalizing students of differing backgrounds but the disparity comes when students are no longer in school.

Alexander et al. (2007) show that the gaps in reading comprehension are greatest in the summer. They also illustrate how these gaps then translate into differential dropout, graduation, and college attendance rates in high school by socio-economic status. Over half of the variation in achievement gap between SES groups in ninth grade can be accounted for by summer reading loss in the previous nine summers.

Limitations of the Current Literature

In most cases, longitudinal data used in studies come from national datasets such as the National Education Longitudinal Study or the Early Childhood Longitudinal Survey—Kindergarten. Due to the collection designs of these datasets, studies relying on them rarely look at growth over subsequent grades. Achievement data from national datasets typically include achievement scores for kindergarten, first, third, fifth and eighth grades. While this approach may capture a larger time span of growth for students in general, the unequal intervals and sparse observation points may decrease the accuracy and sensitivity to growth nuances over time.

Most of the existing studies in the literature are concerned with reading loss over the summer months. The few studies that have been conducted on math yield controversial results. Cooper et al. (1996) suggest math scores decrease over summer for all students without a significant difference in background variables. They also, however, acknowledge the work done by Entwisle and Alexander (1992) where significant differences in math achievement were found. More research is needed in this area—especially as society continues to place higher value on science and math education goals.

Additionally, the literature on summer loss is limited to studies using national datasets or data from large urban school districts as other districts rarely test their students more than once a year. Similarly, the studies on student growth models mainly model student growth with the use of annual test administrations and lack the power to incorporate potential summer learning loss. This study merges these two fields to compare student growth trajectories from rural Utah school districts while accounting for summer loss.

Chapter 3: Method

Design

The study was comprised of longitudinal achievement data from two rural school districts in Utah. The data included five cohorts of students with data for various numbers of test occasions (Table 1). For example, a cohort included nine data points for students who entered the study in third grade and continued through all three years. Some students were already in fifth grade when data collection began, leaving them with three data points through fifth grade. Cohorts were labeled as the year the students were in the first grade.

All students were included in the analysis since multilevel modeling is equipped to handle missing data due to waves of data collection as is present in this study. Student test scores were compared based on grade levels although not all the participants in the study were in third grade during the same calendar year.

Table 1

Cohort Labels by Grade and Data Collection Year

Grade Level	2010-2011	2011-2012	2012-2013
Third Grade	08	09	10
Fourth Grade	07	08	09
Fifth Grade	06	07	08

Participants

Of five Utah districts that elected to use the NWEA MAP tests and were invited to participate in this study, the Sevier and Uintah school districts agreed to participate in this study. Both districts are considered rural districts with and. Both districts are predominantly Caucasian with minorities mostly belonging to Hispanic and Native American ethnic categories. Over the course of the study, data were collected from more than 5,000 students in 13 different schools as is shown in Table 2.

The Sevier School District is a relatively small, rural district located in Sevier County in southcentral Utah. Richfield, the county seat, is located approximately 160 miles south of Salt Lake City. In 2014, the county population was estimated by the U.S. Census Bureau to be 20,870 with a median family income of $36,327. Of the families with children 18 years old or younger, 15.7% have an income below the poverty level. Last year, approximately 36% of the population lived in the city of Richfield while the remaining 64% lived in 10 other cities and in unincorporated areas of the county. The population consists primarily of Caucasians, but includes about 2% who are Native Americans. The school district is coterminous with the boundaries of the county and includes all public elementary and secondary schools in the county. Sevier's fall 2010 enrollment was 4,533. The economy of the county is primarily agricultural, but manufacturing, mining and tourism are also important.

The Uintah School District is a relatively small, rural district located in Uintah County in the eastern part of Utah adjacent to the Colorado border. Uintah's fall enrollment was 6,684. The 2015 county population was estimated to be 37,928 with a median family income of $62,363. Ten percent of families with children under 18 in the Uintah district live below the poverty line. Close to 30% of the population lives in county seat of Vernal, while the remainder live in smaller communities and in the unincorporated parts of the county. The district includes seven elementary schools. The main ethnic groups are Caucasians, but approximately 7% of the population is Native American. The economy of the county is heavily dependent on oil drilling, tourism and agriculture.

Table 2

Data Provided by Sevier and Uintah Districts by Subject

Data Source	Math	ELA	Reading	Total
Data Points	25,079	24,998	25,058	75,135
Students	5,236	5,220	5,228	5,240
Teachers	233	233	233	233
Schools	13	13	13	13
Districts	2	2	2	2

Instruments

The Measures of Academic Progress (MAP) assessment is a computer-adaptive test that produces separate scores for each student on mathematics, reading, and language usage for students in grades 2 through 12 for use three to four times a year. As a computer-adaptive test, the test length can vary, but there are approximately 40-50 items per test, and examinees generally take less than one hour to complete it. Examinee scores are vertically scaled across grade levels. Test score averages by grade level and subject for students in this study are found in Table 3.

Table 3

Mean Scores by Grade Level and Subject

Grade Level	Math	ELA	Reading
Third	201.053	196.449	193.619
Fourth	211.272	204.133	202.619
Fifth	221.364	210.802	209.957
Combined	214.073	206.278	203.607

Reliability estimates obtained through item response theory analysis range from .94 to .95 in third- through fifth-grade examinees in all three subject areas. Scores from the MAP test were correlated with scores from other standardized measures such as the *Stanford Achievement Test*

9^{th} *Edition* (SAT9), the *Iowa Test of Basic Skills* and several state assessments. Correlations with relevant subjects and grade levels range from .70 to .88. MAP scores were also linked to the proficiency categories on the Smarter Balanced Assessment Consortium assessment with 84%-88% accuracy.

The assessment was designed for formative assessment endeavors of teachers and administrators to monitor the growth of students throughout the school year. The test was not designed for teacher or principal evaluation purposes but was used in many states and school districts (e.g., Seattle, Chicago, North Carolina) to inform evaluation decisions. As a result, many states and schools have abandoned the MAP for tests that are more clearly aligned with the purposes for which the test were being used in local school settings.

Procedure

Each district administered the Northwest Evaluation Association's (NWEA) *Measures of Academic Progress* (MAP) test in Math, Language Use, and Reading three times a year. Both districts provided NWEA generated data of student test and demographic data at the test occasion level. Each of the files was imported into a secure database.

A flat file was generated with the test and demographic data for third through fifth graders in both districts over three consecutive years from which data analyses were performed. Data were organized in long format with each row consisting of data for a single test occasion in a single subject. Unique IDs were generated for students, teachers and schools since the IDs provided were internal district IDs with potential for duplication.

Student-level variables were also included for each test occasion. Student gender, ethnicity, and free or reduced lunch status were collected on student characteristics for each test occasion. In addition, one of the districts included a dichotomous variable labeled as "Low

Income." Students who were classified as either "Low Income" or "Free/Reduced Lunch" in the data were classified as "Low SES" in the current study. While there are limitations in using these variables to approximate socioeconomic status, these variables are widely used and were the only proxies available in the dataset. Since a large proportion of students' low SES categorization changed over time, this variable was left as a time-varying, student-level variable. Fewer than 10 students had changes in gender or ethnicity over the duration of the study. These students were eliminated from the dataset. Separate data files were created by subject for the univariate analyses. Table 4 shows the number of students who were incorporated into analyses for each subject at each test occasion after cleaning and prepping the files.

Table 4

Number of Students Tested by Subject and Test Occasion

	Third Grade			Fourth Grade			Fifth Grade		
	Fall	Winter	Spring	Fall	Winter	Spring	Fall	Winter	Spring
Math	2,903	2,900	2,894	2,799	2,811	2,798	2,655	2,664	2,655
ELA	2,891	2,887	2,891	2,785	2,798	2,795	2,646	2,654	2,651
Reading	2,899	2,897	2,894	2,795	2,808	2,797	2,655	2,660	2,653

Analysis

This study used piecewise, longitudinal, multilevel modeling with test occasions nested within students (Level 1), students nested within teachers (Level 2), and teachers nested within schools (Level 3). The model was developed using SPSS, with variables added first as fixed effects estimating an average effect for all students. Random effects for the same variables were then added to account for variability in student trajectories around the average fixed effects. The resulting model produced estimates of variability in student intercepts and slopes over time.

In Level 1, variables were added to the model, which vary over time. The time variable was coded as a function of test occasions rather than the passage of calendar days, weeks or months; therefore, growth was measured as the amount of change that occurs between the first, second and third test occasions in each school year. Linear growth was estimated as a baseline slope across all time points, after which additional slope variables (i.e., fourth- and fifth-grade slope) were added to estimate growth in subsequent time periods, Additional summer offset variables were added to estimate the average drop in achievement after the summer months. Setting up the piecewise model this way tested the statistical significance of subsequent growth (i.e., fourth and fifth grade) in comparison to the baseline year (i.e., third grade). Using a piecewise in this way allowed for estimating each grade level's intercept and slope parameters separately.

Demographic variables were added as covariates at Level 2. Previous research on summer loss suggests that race and gender are not significant predictors. These findings are based on urban datasets with large black or Hispanic minorities. Rural Utah districts have different demographic profiles with a larger Native American constituency. Based on the difference in demographics from previous studies, race and gender were also incorporated as Level 2 predictors.

In addition to race, students' economic background was incorporated into the model as a second-level variable as represented by free or reduced lunch status and low-income status. These variables are dichotomous variables reported by the respective districts. Gender was also included in the model. Table 5 shows available student-level variables and their frequencies.

The third level of the model groups students within teachers. The model at this level is cross-classified since students typically change teachers from year to year. In the literature on student growth trajectories, few studies estimate cross-classified models. If a study includes a three-level model, the three levels are most often test occasion, student, and school. It is possible that the common exclusion of teacher or class as a level is due to the complexity of a cross-classified model required to isolate a teacher effect and variability. The current study seeks to further investigate analytic options of dealing with cross-classified, longitudinal, multilevel models.

Table 5

Distribution of Students Classified by Level-1 Covariates

Variable/Category	Number of Data Points	Number of Students
Gender		
Female	36,685	2,557
Male	38,440	2,682
Ethnicity		
Caucasian	64.942	4,437
Hispanic	4,292	347
Native American	4,569	333
Asian/Pacific Islander	834	67
Black	346	33
Low Income	40,520	2,744

A cross-classified model can be conceptualized as a consideration of all possible combinations of the higher level. For example, a cross-classified study would result in the investigation of neighborhood versus school effects. Not all students from the same neighborhood go to the same school, and not all schools are made up of students from the same neighborhood, producing cross-classification. In this scenario, we could map the possible combinations of school and neighborhood and classify the students by their combination of neighborhood and school. In the current study, students are classified by their combination of teachers over the four grade levels included in the study (Table 6). Studies with cross-classified models typically have two variables crossed in nesting (e.g., neighborhood and school). The current study includes three crossed variables for each grade level (i.e., third-grade teacher effects, fourth-grade teacher effects and fifth-grade teacher effects).

Cross-classified analysis estimates the effect of each of the crossed variables, but can also incorporate the interaction effects of cross-classification. The current study further investigates the advantages and disadvantages of including the interaction effects of a cross-classified model.

Table 6

An Example of Cross-classification of Third- and Fourth-Grade Teachers

Fourth-Grade Teachers	Third-Grade Teachers			
	Teacher A	Teacher B	Teacher C	Teacher D
Teacher E	S_3, S_8	S_6, S_7, S_{18}		
Teacher F		S_1, S_{12}	S_2, S_{20}, S_{11}	
Teacher G	S_{10}, S_{16}, S_{19}			S_9, S_{13}
Teacher H	S_4, S_{17}		S_5, S_{14}	

Equations 1.1 through 3.4 represent the theoretical model for this study with $a_{ti(j1,j2,j3)}$ representing time-varying factors of test occasion, grade level, student SES, and an interaction between occasion and grade. The term $X_{i(j1,j2,j3)}$ represents student-level variables of gender, and ethnicity as dummy coded into categories of Hispanic, Native American and Black and other minorities. The inclusion of $_{(j1,j2,j3)}$ denotes the cross-classification of three classes of teachers: third-, fourth-, and fifth-grade teachers.

Students and teachers are also nested within schools. Since there are only 15 schools, it is difficult to precisely estimate school-level random effects. However, there is concern that the exclusion of this fourth-level impacts teacher-level effects. This study will therefore also explore the school-level of analysis in comparison with teacher effects.

Level 1

$$Y_{ti(j1,j2,j3)} = \pi_{0i(j1,j2,j3)} + \pi_{1i(j1,j2,j3)} a_{ti(j1,j2,j3)} + e_{ti(j1,j2,j3)} \qquad (1.1)$$

Level 2

$$\pi_{0i(j1,j2,j3)} = \beta_{00(j1,j2,j3)} + \beta_{01(j1,j2,j3)} X_{i(j1,j2,j3)} + u_{0i(j1,j2,j3)} \qquad (2.1)$$

$$\pi_{1i(j1,j2,j3)} = \beta_{10(j1,j2,j3)} + \beta_{11(j1,j2,j3)} X_{i(j1,j2,j3)} + u_{1i(j1,j2,j3)} \qquad (2.2)$$

Level 3

$$\beta_{00j} = \gamma_{000} + v_{00j1} + v_{00j2} + v_{00j3} \qquad (3.1)$$

$$\beta_{01j} = \gamma_{010} + v_{01j1} + v_{01j2} + v_{01j3} \qquad (3.2)$$

$$\beta_{10j} = \gamma_{100} + v_{10j1} + v_{10j2} + v_{10j3} \qquad (3.3)$$

$$\beta_{11j} = \gamma_{110} + v_{11j1} + v_{11j2} + v_{11j3} \qquad (3.4)$$

Chapter 4: Results

Results are presented in two sections that reflect the two purposes of this study. Section I answers research questions 1-3 regarding significant predictors and patterns of academic growth that may inform educational practice, using typical analytic techniques. Section II answers questions 4 and 5 regarding methodological issues in the measurement of student growth. The results will be presented in that order.

Preliminary Models

We started with an unconditional linear model with the scores predicted by an intercept and the overall linear slope across all occasions (Overall Slope). Both the slope and intercept fixed effects were significant estimates. We then added slope terms for fourth and fifth grade as fixed effects in the model. The fourth-grade slope was not significantly different from the overall slope in math and reading, while both the fourth- and fifth-grade slopes were significantly different from the overall or third-grade slope for ELA. In both models the residual variance and log likelihood estimates were similar.

After adding the summer offset terms, the log likelihood estimate dropped for both the ELA and Math models. In each model growth slowed after third grade. Additionally, the summer offsets showed large and statistically significant achievement loss in each summer compared to the score that students were predicted to achieve if their rate of growth during the previous year had continued.

Past studies investigating student growth suggest that student growth is nonlinear, which is often accounted for by a quadratic term. With three data points per school year, quadratic terms were added in a similar format as the piecewise linear terms. All quadratic terms were significant except in the case of the quadratic term for fifth-grade reading. We also added a random effect for the intercept, which reduced the log likelihood ratio but did not substantially change the fixed effects estimates.

Table 7

Model of Time Parameters without Student-Level Covariates

Parameter	Math	ELA	Reading
Intercept	190.080*	188.788*	185.883*
Overall Linear Slope	7.338*	8.148*	9.613*
Overall Quadratic	1.768*	-0.509*	-1.350*
Fourth-Grade Linear Slope	-14.547*	-0.262	5.510*
Fourth-Grade Quadratic	0.976*	0.696*	0.600*
Fifth-Grade Linear Slope	-24.649*	2.221	12.531*
Fifth-Grade Quadratic	0.540*	0.516*	0.243
Summer 1 Offset	-12.523*	-4.995*	-3.027*
Summer 2 Offset	-34.592*	-5.684*	3.135*
Residual Variance	28.904*	30.600*	39.140
Intercept Variance	137.677*	207.118*	236.307*
Overall Slope Variance	1.121*	0.497*	0.724*
-2 Log Likelihood	174367.911	173958.194	180101.431

*$p < .05$.

A structural equation modeling approach was used to confirm nonlinear trends. The SEM approach confirmed the nonlinear trends of math and reading. The factor loadings for the middle test occasion for math were consistently higher than 1.0, but were consistently lower than 1.0 for reading. The math results suggested students learn more in the second half of the school year while in reading, students seem to have higher gains during the first half of the academic year.

Growth in Mathematics

Initial third-grade achievement. On average, students scored 191.62 in math when they entered third grade. Unlike ELA scores, math scores when entering third grade were not significantly different for students who were classified as low SES; however, female students score an average of .85 points lower than their male peers. Hispanic students started third grade at an average of 7.35 points lower than their White peers. On average, American Indian students scored 10.60 points lower than their White peers at the beginning of third grade. The mean score for other minority groups were not significantly different at the beginning of third grade. Table 8 reports the model parameters for the intercept as well as for growth. These parameters are highlighted below. **Growth trends**. The overall trajectory for growth in math scores had both a significant linear effect and a significant quadratic effect. Positive quadratic effects in each grade level and for the overall quadratic term suggested that growth increased on average after the second test occasion. In this scenario, the negative linear terms diminished the exponential increase of scores over time, but the overall trends continued to show exponential growth within each grade. Additionally, the quadratic and linear effects were significantly different in each grade level. Based on predicted scores for a student in the reference group (White, male, middle to high SES), students' math scores increased by 22.43 points in third grade, 18.36 points in fourth grade and 17.71 points in fifth grade. **Summer loss**. After the first summer, students scored 12.91 points lower on average than they would have if they continued their third-grade trajectory through the summer months. In the second summer, on average, students experienced a drop of 34.51 points from their predicted scores on entry into fifth grade based on the fourth-grade trajectory.

Gender differences. As is shown in Figure 1, male students in third grade grew faster in comparison to their female peers. The gender gap (favoring males) widened through the third grade from 0.85 points to 2.13 points by the end of the third grade. In fourth and fifth grades, the growth of female students is close to parallel to male students' growth, with a small gap maintained.

Table 8

Math Growth Model Parameters with Student-Level Covariates

Parameter	Estimate	Standardized Estimate	Std. Error	*p*
Intercept	191.621	-0.030	0.295	< .001
Overall Linear Slope	7.835	1.062	0.275	< .001
Grade 4 Linear Slope	-14.064	-0.730	0.591	< .001
Grade 5 Linear Slope	-23.744	-0.858	1.298	< .001
Overall Quadratic Slope	1.69	1.988	0.126	< .001
Grade 4 Quadratic Slope	0.943	0.099	0.178	< .001
Grade 5 Quadratic Slope	0.55	0.037	0.182	0.002
Summer 1 Offset	-12.91	-0.649	0.269	< .001
Summer 2 Offset	-34.514	-0.869	1.169	< .001
Low SES	0.353	0.018	0.273	0.197
Overall Linear Slope * Low SES	0.154	0.010	0.168	0.358
Grade 4 Linear Slope * Low SES	-0.292	-0.007	0.234	0.211
Grade 5 Linear Slope * Low SES	-0.608	-0.010	0.243	0.012
Summer 1 Offset * Low SES	0.155	0.004	0.21	0.461
Summer 2 Offset * Low SES	0.082	0.001	0.369	0.823
Female	-0.853	-0.035	0.387	0.027
Overall Linear Slope * Female	-0.636	-0.045	0.147	< .001
Grade 4 Linear Slope * Female	0.465	0.012	0.206	0.024
Grade 5 Linear Slope * Female	0.445	0.008	0.211	0.035
Summer 1 Offset * Female	0.883	0.023	0.191	< .001
Summer 2 Offset * Female	0.984	0.013	0.335	0.003
Hispanic	-7.352	-0.104	0.814	< .001
Overall Linear Slope * Hispanic	-0.391	-0.013	0.326	0.231
Grade 4 Linear Slope * Hispanic	-0.644	-0.008	0.457	0.159
Grade 5 Linear Slope * Hispanic	0.644	0.005	0.462	0.163
Summer 1 Offset * Hispanic	0.532	0.006	0.426	0.211
Summer 2 Offset * Hispanic	1.576	0.009	0.74	0.035
Native American	-10.601	-0.203	0.793	< .001
Overall Linear Slope * Native American	-1.449	-0.049	0.306	< .001
Grade 4 Linear Slope * Native American	-0.836	-0.010	0.431	0.052
Grade 5 Linear Slope * Native American	-0.678	-0.006	0.456	0.137
Summer 1 Offset * Native American	1.028	0.013	0.396	0.009
Summer 2 Offset * Native American	1.757	0.011	0.729	0.016
Other Minority	-1.92	-0.023	1.382	0.165
Overall Linear Slope * Other Minority	-0.257	-0.005	0.548	0.639
Grade 4 Linear Slope * Other Minority	-0.359	-0.002	0.807	0.657
Grade 5 Linear Slope * Other Minority	0.442	0.002	0.822	0.59
Summer 1 Offset * Other Minority	-0.545	-0.004	0.727	0.453
Summer 2 Offset * Other Minority	2.149	0.007	1.385	0.121

The first summer's slope of 0.88 for female students compensates for the negative overall slope, narrowing the gap between genders over those summer months. On average, male students gain 3.62 points while female students gain an average of 3.58 points during the first summer. The second summer, after fourth grade, has a similar slope for females of 0.98. Females, then tend to grow slightly less in math achievement during the school year but narrow the achievement gap over the summer months as they lose less achievement than their male peers.

SES differences. Figure 2 shows that overall, SES was not a significant predictor of math growth. In fifth grade, however, low SES students grew an average of 0.61 less than their peers. All other growth trends for math are roughly equivalent to their peers.

Ethnic differences. In general Hispanic students do not experience significantly different growth patterns in comparison to their White peers as is shown in Figure 3. The only exception is during the second summer, after fourth grade, when Hispanic students add an additional 1.58 points to their scores. The growth of American Indian students is significantly less than their White peers in third grade, and through each of the three summers. Math scores for American Indian students decrease 1.45 points for each test occasion in third grade. American Indian students gain 1.03 points more than White students when returning to fourth grade and 1.76 more points when returning to fifth grade. This suggests that the achievement gap between Whites and Native Americans narrows between academic years, and growth within the school year is similar. Students who classified themselves as Black or another ethnicity did not have significantly different growth patterns from their peers.

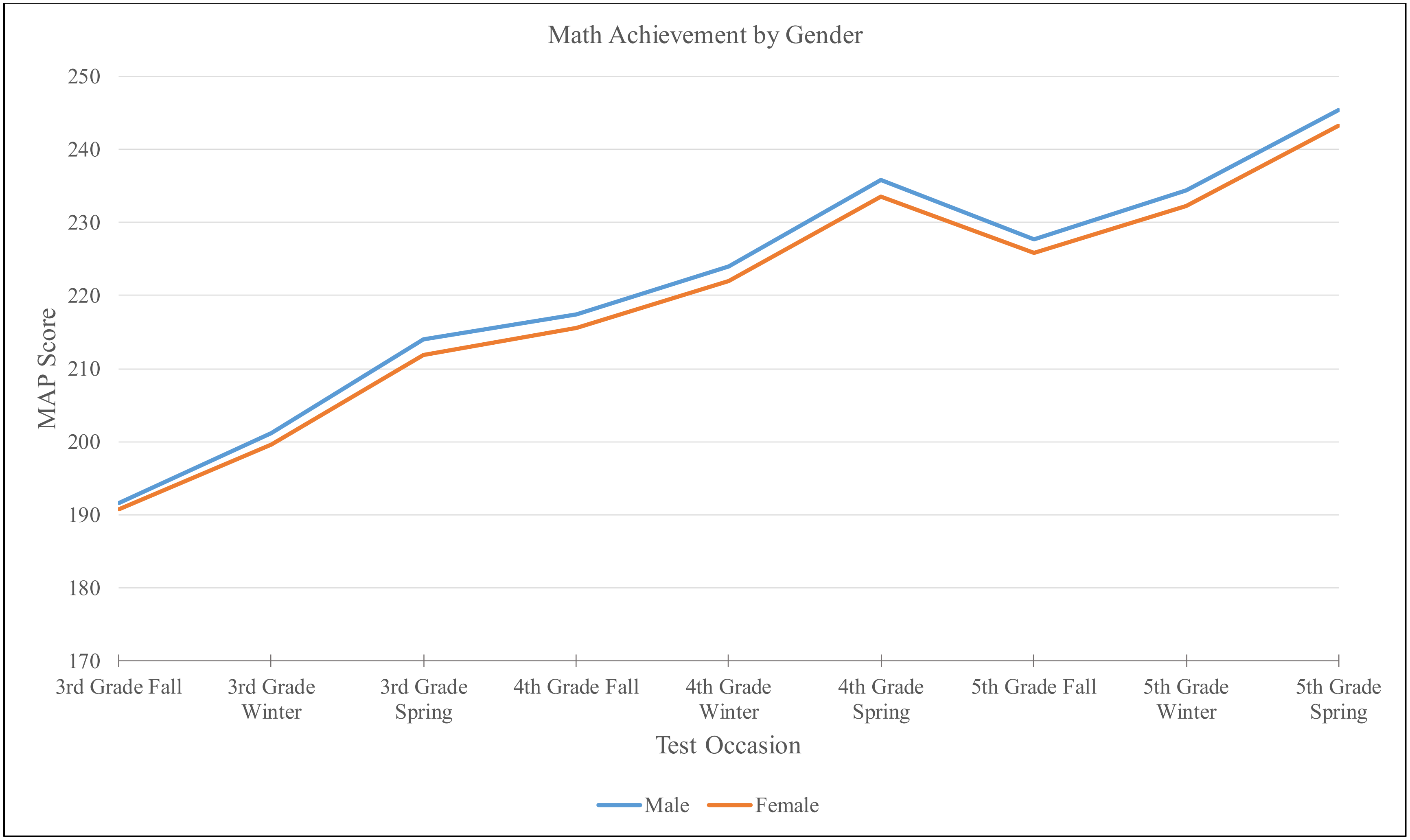

Figure 1. Math achievement by gender and grade level.

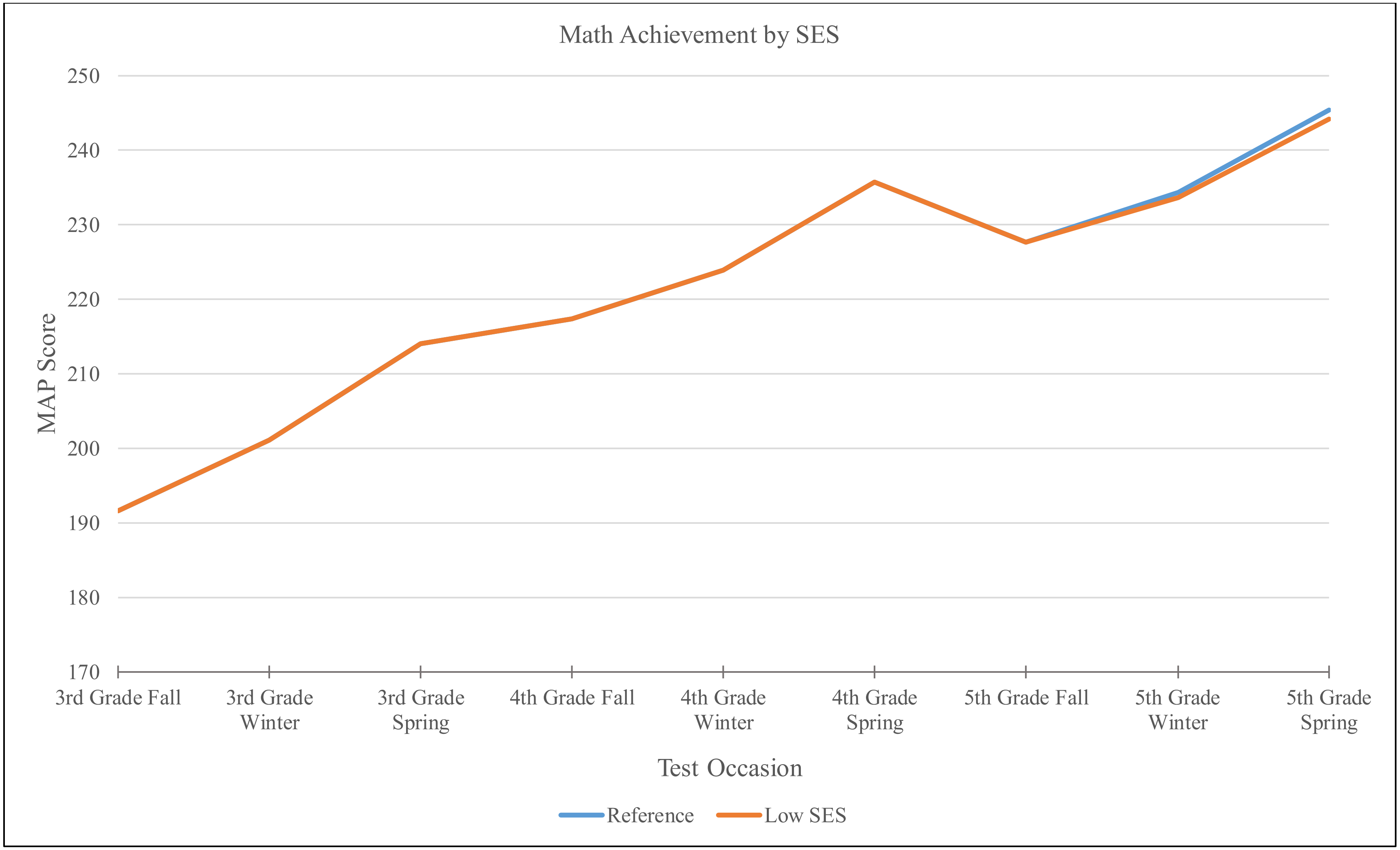

Figure 2. Math achievement by SES and grade level.

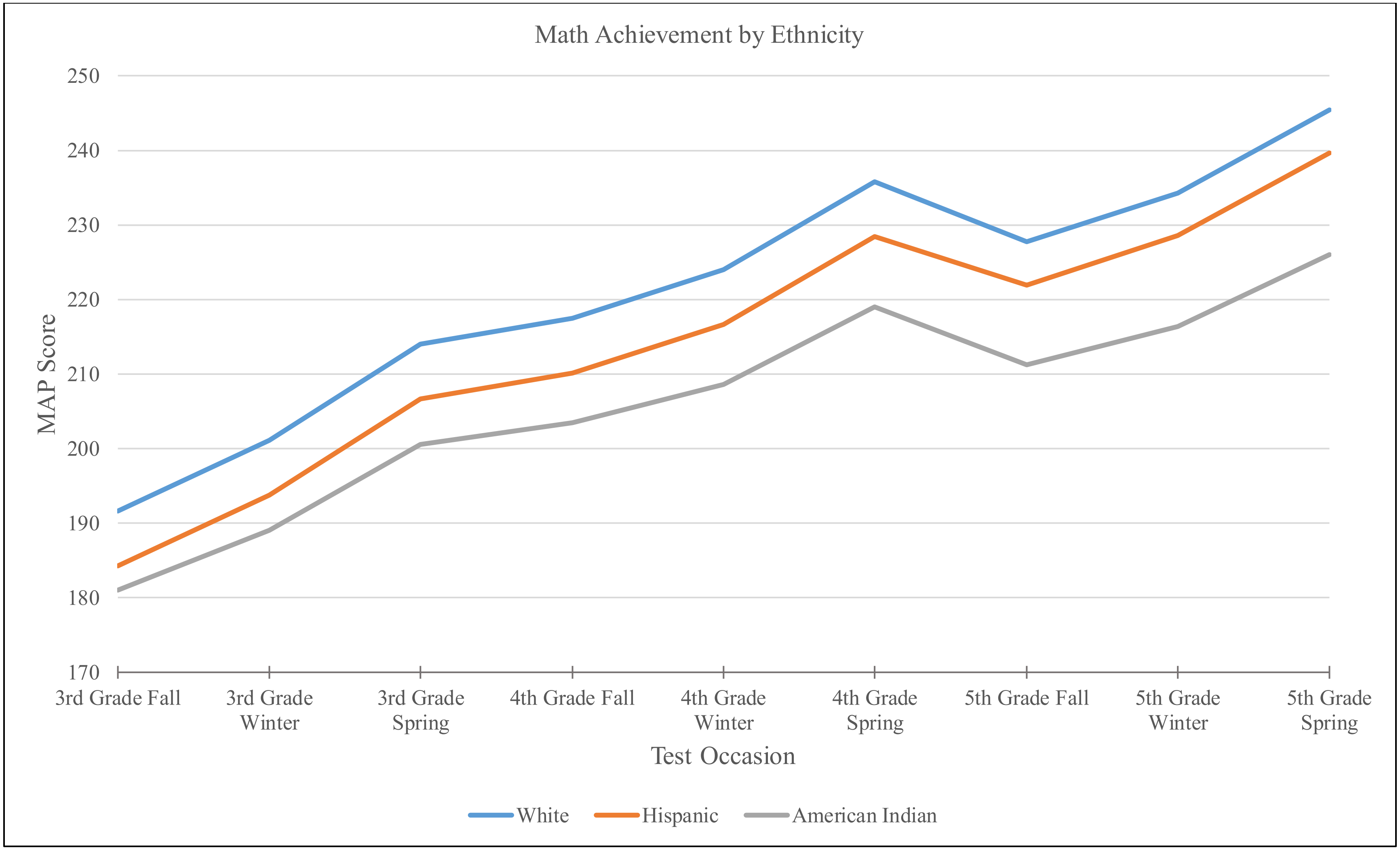

Figure 3. Math achievement by ethnicity and grade level.

Growth in English Language Arts

Initial third-grade achievement. Females scored about three points higher than males on average when entering third grade. Low SES students scored one point lower on average in the beginning of third grade. Hispanic students on average scored about 10 points lower than their White peers, and American Indian students typically scored more than 13 points lower than White students. Students identified as any other minority (e.g., Black, Asian) or as multiple minorities scored an average of four points lower than their White peers. Table 9 shows the parameters estimated to model ELA growth.

Growth trends. Slopes for fourth- and fifth-grade were not significantly different from the third-grade slope. Differences between grades were accounted for in quadratic terms. The overall quadratic term suggests that within third grade, the score gain between test occasions decreases by around 0.5 points on average per test occasion. Each of the quadratic terms is significant suggesting that the polynomial shape in fourth and fifth grades are significantly different (although a potentially small difference) from the polynomial shape of third-grade scores.

Summer loss. The negative slopes for both the first and second summers confirm our hypothesis of summer achievement loss. These estimates can be interpreted as the average offset in student achievement from the predicted trajectory of the previous grade slope. The significance of the estimates for both the first and second summers in the presence of other student-level characteristics suggests that all students experience summer loss between third and fourth grades as well as fourth and fifth grades of around five achievement points each. This is a substantial loss considering the average growth (in the presence of all other variables) from fall to spring in the third grade is 13.95 and is 10.28 from fall to spring of fourth grade.

Gender differences. The significant gender effect in third grade is negative, suggesting female students in third grade lose their advantage over their male peers in third grade. Language growth in fourth and fifth grades is not significantly different between genders; however, males lose more during each summer. Females score 0.94 to 1.44 points higher than their male peers in fall test administrations following the summer months. Differences in growth by gender are depicted in Figure 4.

SES differences. Figure 5 shows that students classified as low income or eligible for free or reduced lunch were found to make more progress than their peers in the third grade, closing the small gap estimated at the beginning of third grade. Low SES students continue to climb faster than their peers in fourth and fifth grade. By the end of fourth grade, the low SES students have pulled ahead of their peers but lose more in the second summer. This finding confirms other findings in the literature that suggest the commonly discussed income-based achievement gap is perpetuated by differential summer loss. In the first summer after third grade, low SES students score an additional 0.65 points lower than their peers on average. Low SES students' scores fall an additional 1.85 points in the summer after fourth grade. These offsets are small in magnitude but statistically significant.

Ethnic differences. Growth patterns for Hispanic students mirror the trend for low SES students. Hispanic students increase their scores by an additional 1.35 points above the average growth experienced by their White peers in third grade. Hispanic students maintain their steeper growth rate in fourth and fifth grade. Summer loss for Hispanic students is more severe, with the second summer (after fourth grade) seeing the greatest loss of 3.04 points lower compared with White students. American Indian students experience similar growth in third and fourth grades. In fifth grade, American Indian students increase their scores by an additional 2.38 points each

Table 9

ELA Growth Model Parameters with Student-Level Covariates

Parameter	Estimate	Standardized Estimate	Std. Error	*p*
Intercept	189.189	-0.032	0.337	< .001
Overall Linear Slope	8.020	1.336	0.282	< .001
Grade 4 Linear Slope	-0.051	-0.006	0.607	0.933
Grade 5 Linear Slope	2.237	0.099	1.335	0.094
Overall Quadratic Slope	-0.521	-0.706	0.129	< .001
Grade 4 Quadratic Slope	0.645	0.078	0.183	< .001
Grade 5 Quadratic Slope	0.521	0.041	0.186	0.005
Summer 1 Offset	-4.995	-0.298	0.275	< .001
Summer 2 Offset	-5.110	-0.163	1.200	< .001
Low SES	-0.999	0.001	0.278	< .001
Overall Linear Slope * Low SES	0.751	0.057	0.171	< .001
Grade 4 Linear Slope * Low SES	-0.361	-0.010	0.237	0.129
Grade 5 Linear Slope * Low SES	-0.410	-0.008	0.245	0.094
Summer 1 Offset * Low SES	-0.649	-0.018	0.214	0.002
Summer 2 Offset * Low SES	-1.852	-0.025	0.373	< .001
Female	2.969	0.097	0.446	< .001
Overall Linear Slope * Female	-0.488	-0.040	0.149	0.001
Grade 4 Linear Slope * Female	0.240	0.007	0.210	0.253
Grade 5 Linear Slope * Female	0.246	0.005	0.214	0.251
Summer 1 Offset * Female	0.943	0.029	0.194	< .001
Summer 2 Offset * Female	1.443	0.021	0.336	< .001
Hispanic	-9.895	-0.111	0.931	< .001
Overall Linear Slope * Hispanic	1.353	0.052	0.329	< .001
Grade 4 Linear Slope * Hispanic	-0.137	-0.002	0.466	0.769
Grade 5 Linear Slope * Hispanic	0.179	0.002	0.469	0.702
Summer 1 Offset * Hispanic	-1.421	-0.020	0.432	0.001
Summer 2 Offset * Hispanic	-3.041	-0.021	0.746	< .001
Native American	-13.748	-0.221	0.916	< .001
Overall Linear Slope * Native American	0.581	0.023	0.311	0.062
Grade 4 Linear Slope * Native American	-0.359	-0.005	0.441	0.416
Grade 5 Linear Slope * Native American	2.377	0.024	0.463	< .001
Summer 1 Offset * Native American	-1.324	-0.019	0.403	0.001
Summer 2 Offset * Native American	-5.505	-0.039	0.728	< .001
Other Minority	-3.926	-0.019	1.592	0.014
Overall Linear Slope * Other Minority	1.057	0.023	0.555	0.057
Grade 4 Linear Slope * Other Minority	-0.947	-0.007	0.829	0.253
Grade 5 Linear Slope * Other Minority	-0.689	-0.004	0.835	0.409
Summer 1 Offset * Other Minority	-1.024	-0.008	0.744	0.169
Summer 2 Offset * Other Minority	-0.130	-0.001	1.384	0.925

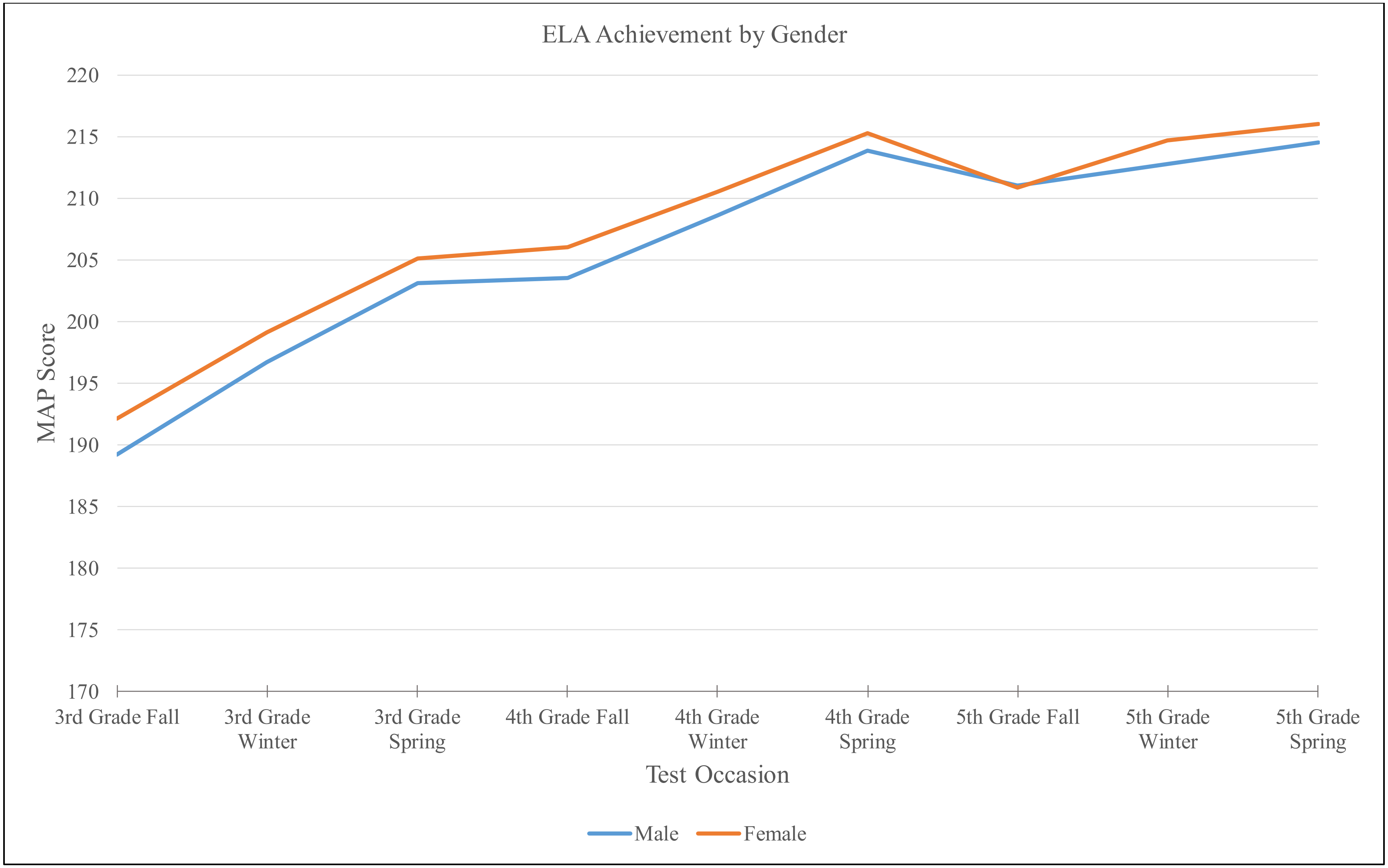

Figure 4. ELA achievement by gender and grade level.

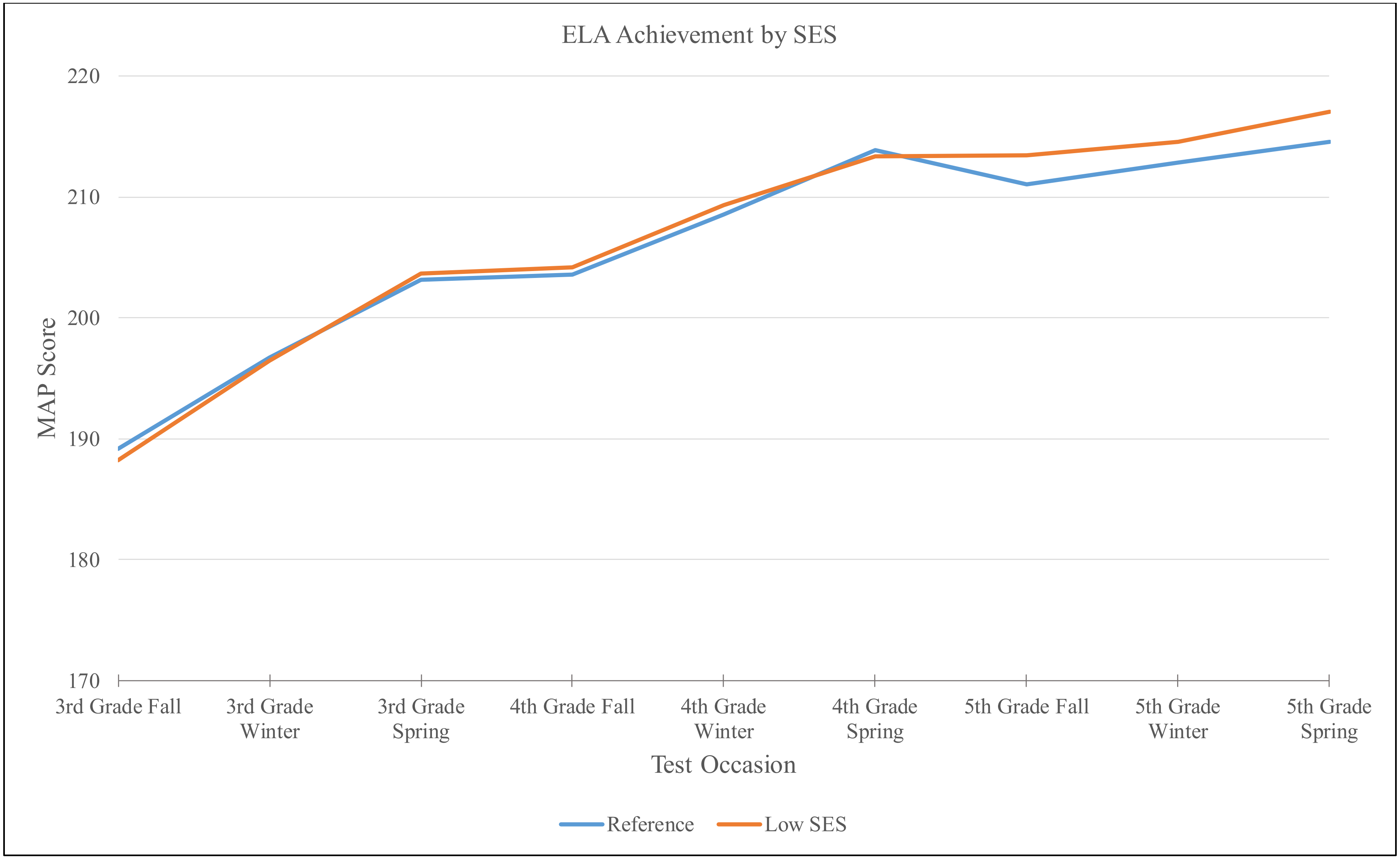

Figure 5. ELA achievement by SES and grade level.

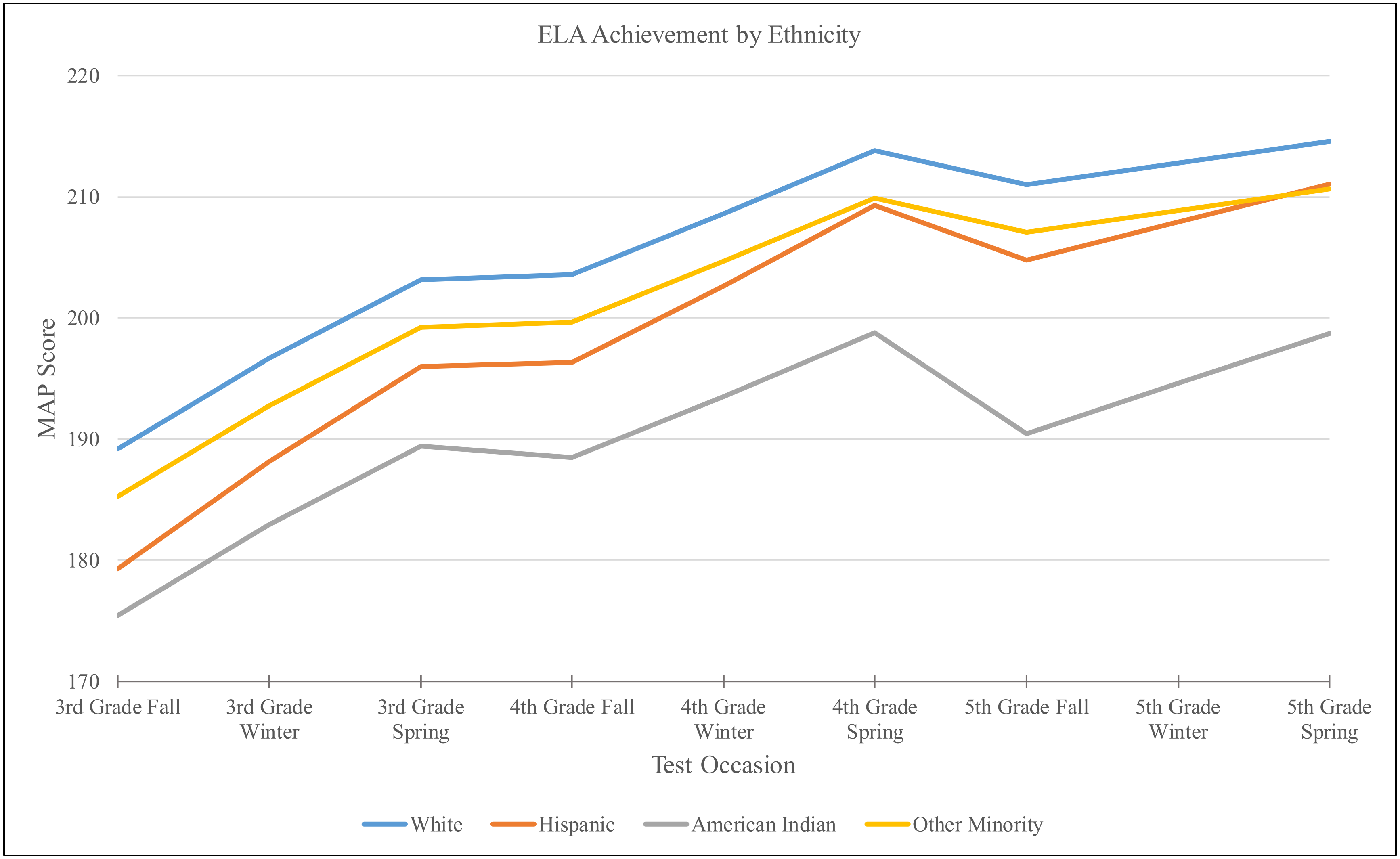

Figure 6. ELA achievement by ethnicity and grade level.

test occasion. Summer loss is also more severe for American Indian. They lose an additional 1.32 points in the first summer and 5.50 points in the second summer. Students classified as any other minority show similar patterns of growth in all grades and during both summers. These growth patterns across ethnic categories are shown in Figure 6.

Growth in Reading

Initial third grade achievement. On average, students started third grade with a reading score of 186.16 points. Low SES students had similar starting averages in third grade. Female students scored 2.86 points higher in reading than their male peers when entering third grade. Hispanics scored an average of 9.95 points lower on their first reading test, and American Indians scored an average of 13.23 points lower than their White peers. Students identified as black, Asian or other minorities scored 4.57 points lower, on average, when entering third grade. These estimated parameters as well as the estimated reading growth parameters are reported in Table 10.

Growth trends. Overall reading growth has both linear and quadratic components. Reading scores increase over time by 9.92 points each test occasion; however, the negative quadratic term indicates that this increase lessens over time. Fourth- and fifth-grade slopes are significantly different than the third-grade slope estimate. These additions in slope are overcome by the quadratic effect, and the slope ends up decreasing across grade levels.

Summer loss. Overall, students score an average of 3.43 points lower entering fourth grade due to summer loss. In the second summer, reading scores grow an additional 3.28 points, above their expected scores in the subsequent fall test occasion. As can be seen in the graph, students still lose achievement in the second summer. The positive estimate most likely corrects for the negative quadratic term.

Table 10

Reading Growth Model Parameters with Student-Level Covariates

Parameter	Estimate	Standardized Estimate	Std. Error	*p*
Intercept	186.163	-0.029	0.369	< .001
Overall Linear Slope	9.923	1.477	0.319	< .001
Grade 4 Linear Slope	5.988	0.326	0.687	< .001
Grade 5 Linear Slope	12.813	0.505	1.510	< .001
Overall Quadratic Slope	-1.398	-1.764	0.146	< .001
Grade 4 Quadratic Slope	0.583	0.066	0.207	0.005
Grade 5 Quadratic Slope	0.301	0.022	0.211	0.153
Summer 1 Offset	-3.426	-0.163	0.311	< .001
Summer 2 Offset	3.280	0.100	1.358	0.016
Low SES	-0.391	0.004	0.317	0.217
Overall Linear Slope * Low SES	0.510	0.036	0.194	0.008
Grade 4 Linear Slope * Low SES	-0.457	-0.012	0.269	0.089
Grade 5 Linear Slope * Low SES	-0.152	-0.003	0.278	0.583
Summer 1 Offset * Low SES	-0.346	-0.009	0.242	0.154
Summer 2 Offset * Low SES	-1.582	-0.020	0.423	< .001
Female	2.856	0.061	0.487	< .001
Overall Linear Slope * Female	-0.965	-0.074	0.169	< .001
Grade 4 Linear Slope * Female	0.119	0.003	0.238	0.618
Grade 5 Linear Slope * Female	0.350	0.007	0.243	0.149
Summer 1 Offset * Female	1.574	0.044	0.219	< .001
Summer 2 Offset * Female	2.231	0.031	0.380	< .001
Hispanic	-9.949	-0.115	1.022	< .001
Overall Linear Slope * Hispanic	0.417	0.015	0.374	0.265
Grade 4 Linear Slope * Hispanic	-0.048	-0.001	0.528	0.928
Grade 5 Linear Slope * Hispanic	0.394	0.004	0.532	0.459
Summer 1 Offset * Hispanic	-0.032	0.000	0.490	0.947
Summer 2 Offset * Hispanic	-0.203	-0.001	0.846	0.81
Native American	-13.236	-0.222	0.998	< .001
Overall Linear Slope * Native American	0.399	0.015	0.351	0.256
Grade 4 Linear Slope * Native American	-0.708	-0.009	0.497	0.155
Grade 5 Linear Slope * Native American	0.751	0.007	0.524	0.152
Summer 1 Offset * Native American	-1.863	-0.025	0.455	< .001
Summer 2 Offset * Native American	-3.503	-0.023	0.824	< .001
Other Minority	-4.568	-0.021	1.738	0.009
Overall Linear Slope * Other Minority	0.688	0.014	0.626	0.272
Grade 4 Linear Slope * Other Minority	-0.603	-0.004	0.929	0.516
Grade 5 Linear Slope * Other Minority	0.936	0.005	0.941	0.32
Summer 1 Offset * Other Minority	-0.124	-0.001	0.835	0.882
Summer 2 Offset * Other Minority	-1.087	-0.004	1.559	0.485

Gender differences. Female students, as shown in Figure 7, start third grade with higher reading scores than male students. During third grade, female students' reading growth drops 0.96 points each test occasion, lessening the gap between the genders. In fourth and fifth grade, the slope of female students' reading growth is similar to that of third grade.

Females consistently outperform their male peers after summer months, although there is still some achievement loss for both genders from their expected fall scores. Female students gain an additional 1.57 points over their male peers over the first summer time period and 2.23 points over the second summer.

SES differences. Overall, low SES students experience more growth than their peers by 0.51 points each test occasion. This slope for low SES students is maintained in fourth, and fifth grade as is shown in Figure 8. Low SES students lose an additional 1.58 points on average during the second summer. Summer loss is roughly equivalent during the first summer.

Ethnic differences. Figure 9 illustrates that during fourth and fifth grades, the reading growth of Hispanic, American Indian and Black, Asian or other minorities is not significantly different from White students. Hispanic students as well as students classified as Black, Asian, or other minorities show parallel growth patterns to White students over each summer. American Indian students' growth, however, is parallel to White students during each school year, but American Indian students lose an average of 1.69 points more than White students in the first summer and 3.53 points in the second summer.

Modeling with Annual Scores

Current student growth measurement practices typically only include one test score per school year, usually in the spring. Using the same data from the above analysis, the spring test scores were extracted and modeled for comparison purposes. These estimated parameters are

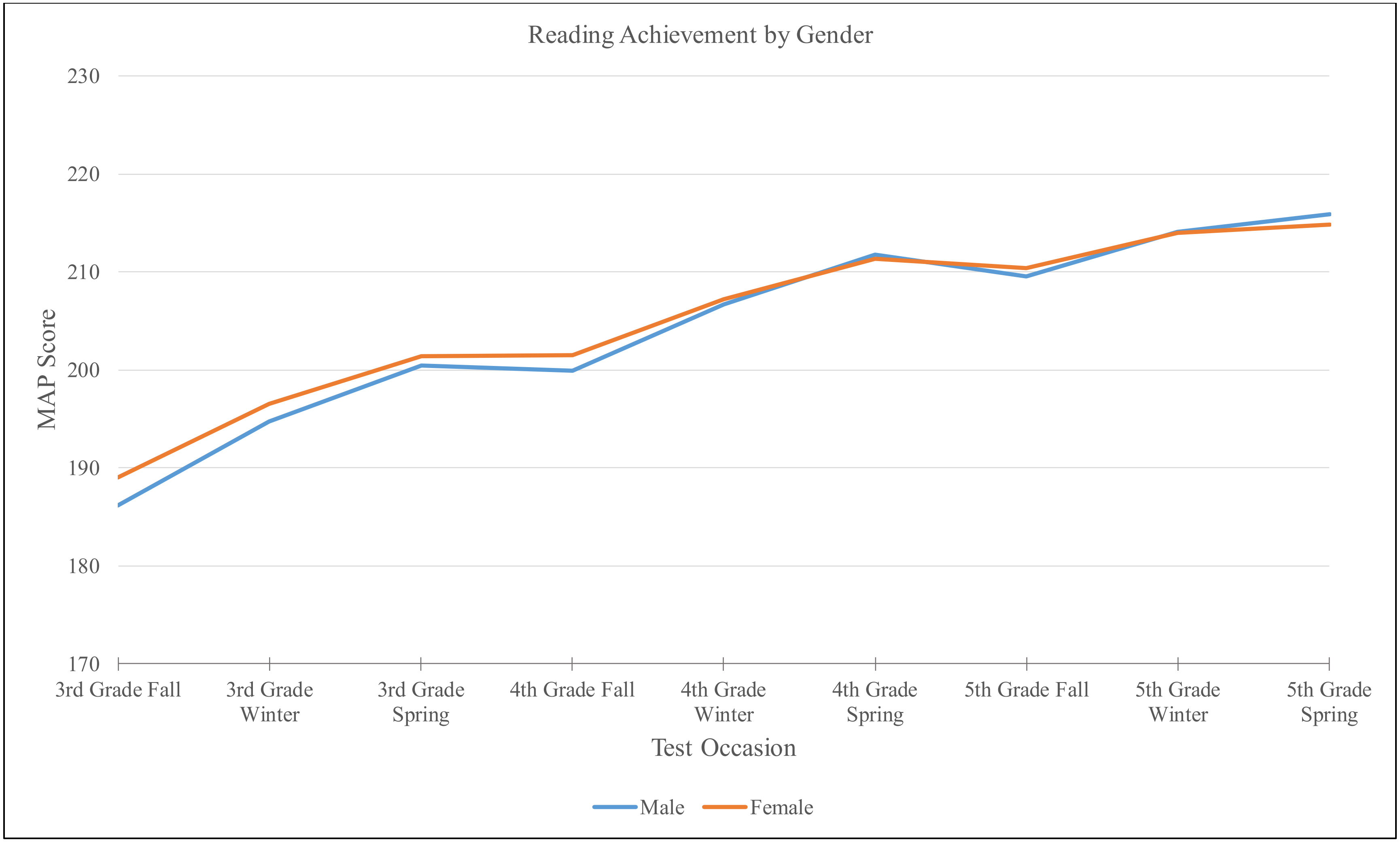

Figure 7. Reading achievement by gender and grade level.

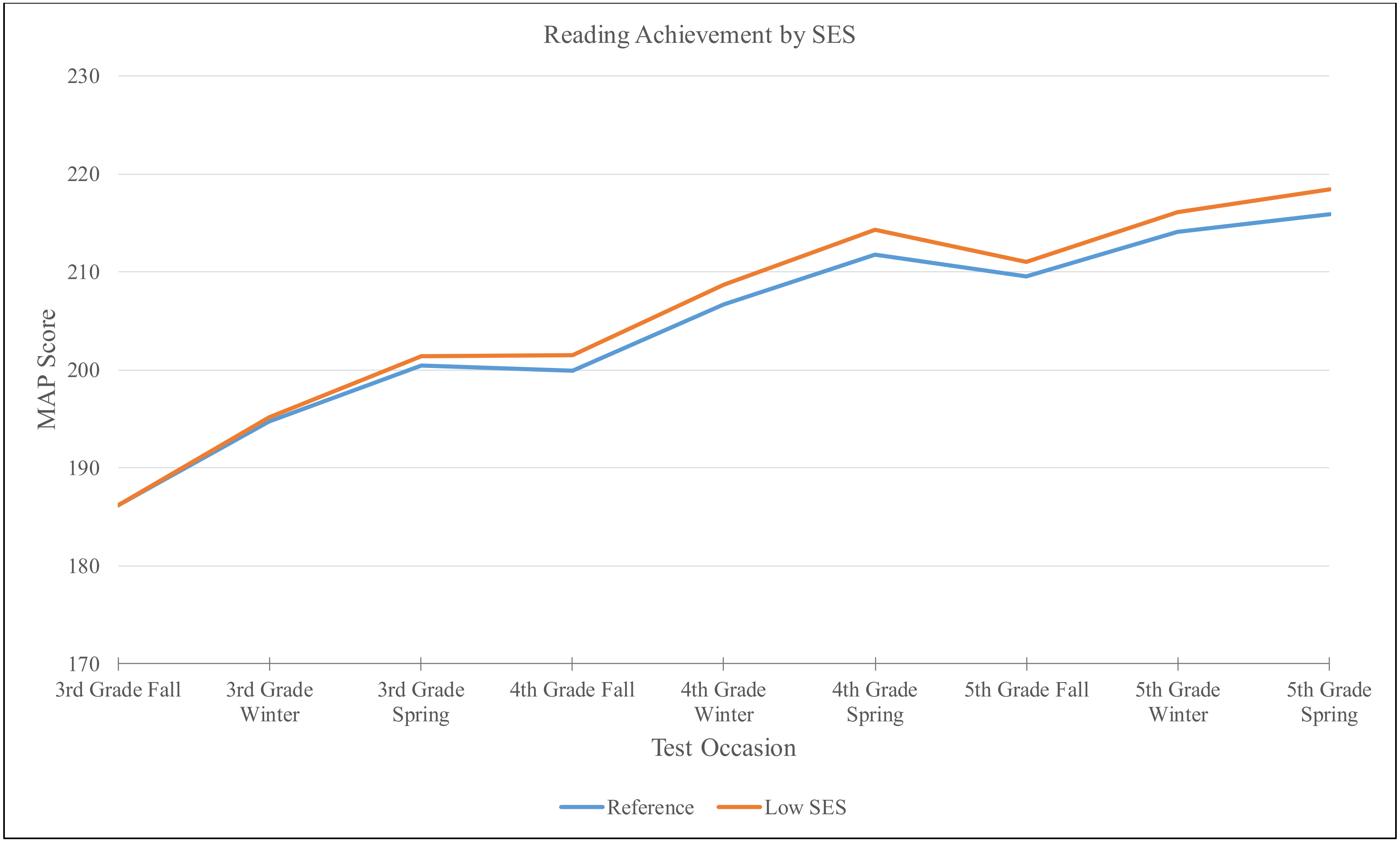

Figure 8. Reading achievement by SES and grade level.

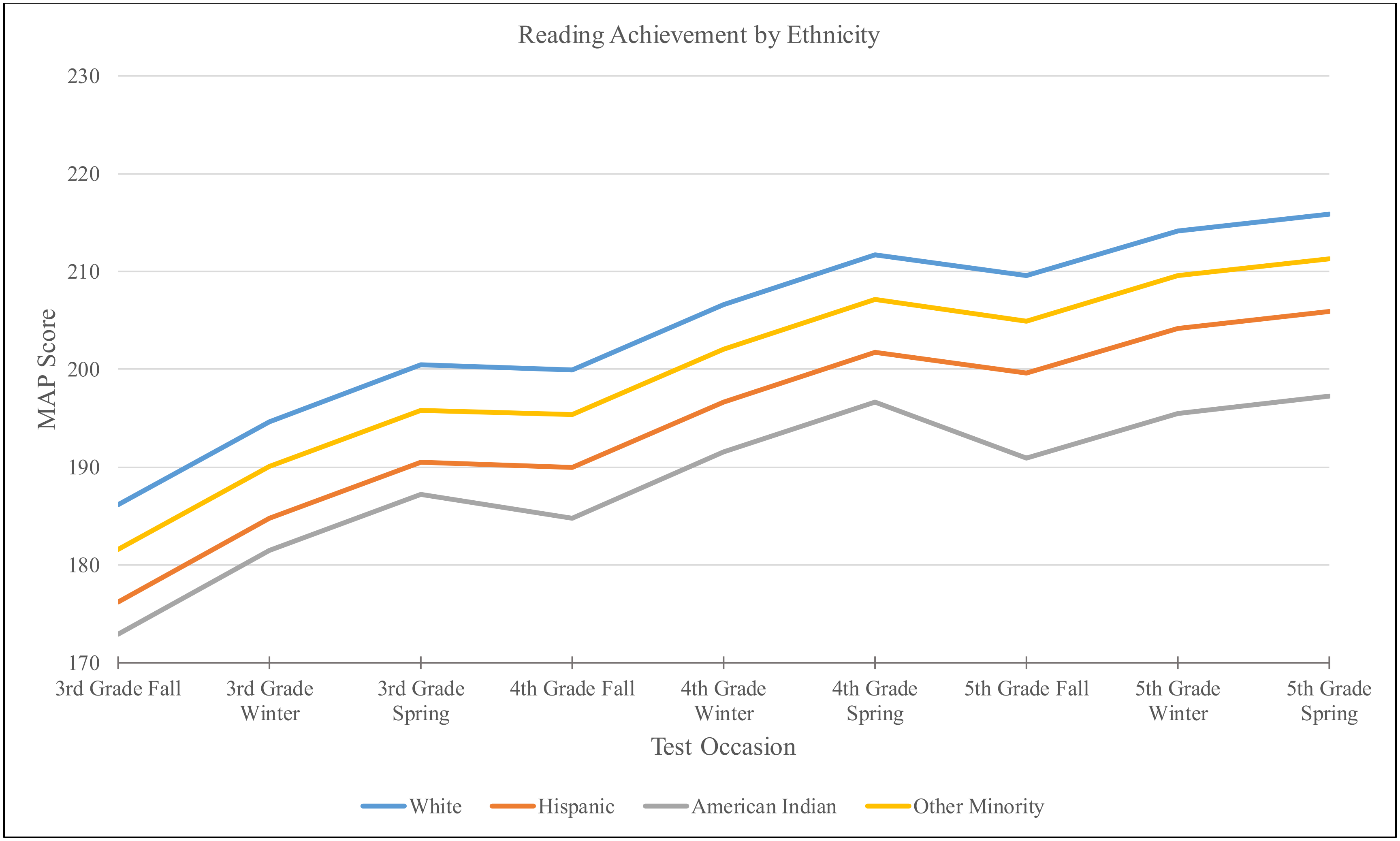

Figure 9. Reading achievement by ethnicity and grade level.

contained in Table 11. Annual scores show only 2 to 3 points of growth each year in any of the three subjects. As suggested by literature and illustrated in the above model, this model grossly underestimates the amount of growth that occurs on average each year. The quadratic trends of ELA and reading show reiterate the decline of growth each year. In previous models, the slope estimate decreases each year. In a model with only one score each year, the overall negative quadratic term is negative. Unlike the previous model, math growth in the current model is linear and increases approximately the same amount each year.

Demographic variables of socioeconomic status, gender, and ethnicity were added to the earlier model to test the variables' significance in predicting initial third-grade scores of students (intercept), students' subsequent growth in third, fourth, and fifth grades (grade slopes) and the summer loss between each grade (summer offsets). Models were estimated separately for ELA, reading and math and compared.

Table 11

Annual Spring Scores Model Parameters

Parameter	Math	ELA	Reading
Intercept	207.283*	199.465*	194.143*
Overall Slope	2.678*	2.136*	3.371*
Overall Quadratic	0.022	-0.030*	-0.125*
Residual Variance	35.306*	19.855*	28.494*
Intercept Variance	192.380*	132.224*	143.405*
-2 Log Likelihood	64483.378	60484.701	62330.484

*$p < .05$.

Teacher-Level Modeling

In adding a third level (teachers) to our unconditional two-level model, we ran into computational limitations. Simply adding the student and teacher-level intercepts as random effects significantly increased the analysis run time from a few minutes to several hours when using SPSS. Similar results were found when using Mplus to add third- (teacher) and fourth-level (school) random effects for intercepts and slopes for each grade level and for each subject separately. The results are shown in Table 12.

The majority of variance is accounted for in student differences. For both the intercepts and slopes for ELA and reading achievement, intercept variance (75%-84% of the total variance)

Table 12

Variance Components for the Random Effects by Subject and Grade Level

	Intercept Variance			Slope Variance		
Subject and Grade Level	Student Variance	Teacher Variance	School Variance	Student Variance	Teacher Variance	School Variance
Math						
Third	123.90 (83.7%)	1.43 (1.0%)	22.66 (15.3%)	0.40 (20.7%)	0.80 (41.6%)	0.72 (37.7%)
Fourth	134.33 (83.7%)	1.51 (0.9%)	24.59 (15.3%)	0.64 (16.7%)	1.78 (46.5%)	1.41 (36.8%)
Fifth	168.42 (80.0%)	1.12 (0.5%)	40.88 (19.4%)	1.13 (22.8%)	2.33 (47.0%)	1.50 (30.1%)
Reading						
Third	233.61 (84.0%)	2.30 (0.8%)	42.35 (15.2%)	15.32 (93.0%)	0.32 (2.0%)	0.83 (5.0%)
Fourth	223.91 (77.8%)	1.36 (0.5%)	62.66 (21.8%)	15.41 (91.3%)	0.76 (4.5%)	0.71 (4.2%)
Fifth	223.80 (79.8%)	1.76 (0.6%)	54.84 (19.6%)	8.61 (83.1%)	0.47 (4.5%)	1.29 (12.4%)
ELA						
Third	221.03 (83.8%)	1.59 (0.6%)	41.07 (15.6%)	12.30 (89.8%)	0.49 (3.6%)	0.91 (6.6%)
Fourth	194.17 (75.1%)	1.00 (0.4%)	63.24 (24.5%)	7.87 (86.5%)	0.26 (2.8%)	0.98 (10.7%)
Fifth	186.81 (78.2%)	1.45 (0.6%)	50.51 (21.2%)	8.78 (83.2%)	0.10 (0.9%)	1.67 (15.8%)

is lower than the slope variance (83%-93%). Interestingly, student-level variance accounts for much less of the total variance in math slopes (16%-22%) than in the ELA or reading slopes.

The SPSS analysis included only the linear grade slopes and summer offsets as fixed effects, and only included the intercept as a random effect. The analysis included the teacher level as cross-classified and required several hours to run. A fourth level (school) was not included. Resulting variance estimates were similar with larger teacher-level variance estimates than shown in the SEM results. The variance between students was smaller for ELA and reading models and larger for the math model.

Multivariate Model

A multivariate analysis was done to better facilitate comparisons of growth trajectories between subject areas. The analysis was done in Mplus using an SEM approach. To adjust for school-level variance, school was added as a clustering variable. Table 13 shows the estimated parameters for each grade level slope and each summer offset for all three subject areas. In addition, Table 14 shows the associated variance components. Estimated slopes for ELA and reading seem to decrease over time, while math slopes decrease the first year and plateau in fifth grade. For all three, the summer loss decreases over two summers.

Table 13

Multivariate Model Parameter Estimates

	Math			Reading			ELA		
Parameter	Estimate	*t*	*p*	Estimate	*t*	*p*	Estimate	*t*	*p*
Intercept	190.161	223.730	< .001	185.919	160.080	< .001	188.651	174.064	< .001
Third-Grade Slope	10.839	88.539	< .001	6.983	38.794	< .001	7.187	30.541	< .001
Summer 1 Offset	-8.961	-35.879	< .001	-3.447	-13.137	< .001	-4.428	-13.353	< .001
Fourth-Grade Slope	8.882	30.995	< .001	5.569	24.238	< .001	5.254	21.429	< .001
Summer 2 Offset	-7.777	-26.023	< .001	-2.300	-6.713	< .001	-3.206	-7.943	< .001
Fifth-Grade Slope	8.483	21.505	< .001	3.805	17.775	< .001	4.309	16.858	< .001

We also compared the variances between subject areas as is shown in greater detail in Table 14. The intercept variance for math is smaller than reading or ELA with reading intercept variance being the largest. Generally the variances for each summer offset in math are both smaller than the offset variances for reading or ELA, again with reading having the larger variances. The slope variances for reading are larger than math and ELA slope variances.

Table 14

Multivariate Model Fixed Effect Variances by Parameter and Subject

Parameter	Estimate	*p*
Math		
Intercept	137.804	< .001
Third-Grade Slope	5.960	< .001
Summer 1 Offset	8.446	.071
Fourth-Grade Slope	5.612	< .001
Summer 2 Offset	9.780	.001
Fifth-Grade Slope	8.191	< .001
Reading		
Intercept	240.421	< .001
Third-Grade Slope	8.506	< .001
Summer 1 Offset	21.104	< .001
Fourth-Grade Slope	6.614	< .001
Summer 2 Offset	20.101	< .001
Fifth-Grade Slope	6.440	< .001
ELA		
Intercept	213.636	< .001
Third-Grade Slope	5.015	< .001
Summer 1 Offset	19.174	< .001
Fourth-Grade Slope	4.106	.001
Summer 2 Offset	14.104	< .001
Fifth-Grade Slope	4.319	< .001

We continued to explore the relationships between the trajectories of each subject area through the correlations between the fixed effects estimates for each subject as are shown in Tables 15, 16, and 17. As might be expected, ELA and reading effects had more significant correlations. Interestingly, the diagonal correlations between ELA and reading estimates correlated above 1.0 (out of range correlations). Other correlations were primarily around 0.70. Correlations were low between ELA third-grade slope and reading slopes for fourth and fifth grades as well as the summer offset between fourth and fifth grade. Additionally, the correlation between slope for reading in third grade and the ELA summer offset before fifth grade as well as the fifth grade ELA slope were not significantly different from zero.

Correlations between math and ELA and math and reading are smaller than the correlations between ELA and reading. The significant correlations are typically between concurrent (shown in the diagonals) or consecutive estimates. For example, the fourth-grade slope for reading correlates highly with fourth-grade math growth and well as math achievement loss in the summers before and after fourth grade. Each of the concurrent correlations are significant; however, the correlations between fifth-grade slope in math and the fifth-grade slopes of reading and ELA suggest there may be a diminishing relationship, meaning that the relationship between math, reading and ELA achievement may lessen in later grades.

As mentioned previously, the intercept for math achievement correlates highly with all ELA and reading estimates. Interestingly, the third-grade math slope has small but negative correlations with both the ELA and reading intercepts, suggesting that students with higher ELA or reading scores when entering third grade are likely to experience less growth in math during third grade.

Table 15

Multivariate Fixed Effects Correlations with Mathematics

Parameter	Intercept	Third-Grade Slope	Summer 1 Offset	Fourth-Grade Slope	Summer 2 Offset	Fifth-Grade Slope
Math						
Intercept		.056	.016	.119	.112	.187
Third-Grade Slope			-.757	.460	-.334	.141
Summer 1 Offset				-.262	.161	-.317
Fourth-Grade Slope					-.617	.366
Summer 2 Offset						-.216
Fifth-Grade Slope						
Reading						
Intercept	.928	-.134	.179	.046	.105	.093
Third-Grade Slope	-.591	.725	-.606	.036	-.038	.073
Summer 1 Offset	.692	-.212	1.013	-.218	.100	-.091
Fourth-Grade Slope	-.675	.127	-.613	.777	-.584	.203
Summer 2 Offset	.709	-.028	.346	-.057	.659	-.081
Fifth-Grade Slope	-.702	.080	-.481	-.131	.000	.348
ELA						
Intercept	.941	-.096	.117	.094	.070	.139
Third-Grade Slope	-.721	.761	-.384	-.110	.130	-.069
Summer 1 Offset	.623	-.308	.881	-.037	-.061	.065
Fourth-Grade Slope	-.792	.184	-.638	.627	-.373	.008
Summer 2 Offset	.672	-.089	.273	-.083	.738	.113
Fifth-Grade Slope	-.759	.156	-.421	-.080	-.281	.237

Note. All correlations were significant at the $p < .05$ level.

Table 16

Multivariate Fixed Effects Correlations with Reading

Parameter	Intercept	Third-Grade Slope	Summer 1 Offset	Fourth-Grade Slope	Summer 2 Offset	Fifth-Grade Slope
Reading						
Intercept		-.657	.814	-.781	.765	-.785
Third-Grade Slope			-.834	.704	-.646	.645
Summer 1 Offset				-.700	.934	-1.036
Fourth-Grade Slope					-.758	.807
Summer 2 Offset						-.789
Fifth-Grade Slope						
ELA						
Intercept	1.014	-.670	.747	-.713	.731	-.759
Third-Grade Slope	-.781	1.182	-.659	.293	-.434	.538
Summer 1 Offset	.626	-.670	1.302	-.714	.651	-.772
Fourth-Grade Slope	-.805	.530	-1.183	1.407	-.734	.641
Summer 2 Offset	.638	-.296	.645	-.608	1.138	-.604
Fifth-Grade Slope	-.694	.277	-.742	.552	-.966	1.026

Note. All correlations were significant at the $p < .05$ level.

Table 17

Multivariate Fixed Effects Correlations with ELA

Parameter	Intercept	Third-Grade Slope	Summer 1 Offset	Fourth-Grade Slope	Summer 2 Offset	Fifth-Grade Slope
ELA						
Intercept		-0.659	0.598	-0.822	0.712	-0.764
Third-Grade Slope			-0.688	0.920	-0.873	0.887
Summer 1 Offset				-0.876	0.736	-0.971
Fourth-Grade Slope					-0.584	1.030
Summer 2 Offset						-1.027
Fifth-Grade Slope						

Note. All correlations were significant at the $p < .05$ level.

Chapter 5: Discussion

The purposes of this study were to explore the growth patterns of math, reading and ELA achievement in students in third, fourth, and fifth grades and relevant methodological questions involved in estimating growth trajectories. Results were informative in both substantive and methodological ways.

The dataset and methods used in this study provide a unique contribution to the current literature. Current studies on growth primarily track growth using one annual measure across several years with one cohort of students. The current study utilized three measures within each of three years across five cohorts of students. In addition, the inclusion of a multivariate model with three outcome variables is novel, and allows for better comparisons of achievement between subjects. Furthermore, the sample included in this study represented rural communities with a large Native American population, whereas most of the studies done in student growth center around urban schools with large proportions of Black minority students. The following conclusions are therefore submitted as a significant expansion on the current literature and provide additional evidence both in agreement with and in contrast to popular conclusions in recent research.

Growth Trends

Achievement growth is complex and varies across grade levels, summer months, subjects, and students. Trends within one subject rarely hold for other subjects. For example, math and reading scores differ between grade levels. However ELA growth is similar across grades. Such complexity makes it difficult to make valid generalizations or to extract general principles. This in itself is a key finding and has practical implications that are discussed later.

Student achievement typically increases across grades, but it increases at a declining rate. This finding is more evident in ELA and reading than in math; however, even math growth decreases after third grade. Fourth- and fifth-grade growth in math is similar. Since causal claims are beyond the scope of this study, further research should investigate possible physiological, psychological, sociological or environmental factors that may influence this trend. For example, perhaps the neural plasticity of younger minds lends them to greater growth in earlier years of schooling. Perhaps there are instructional changes needed to respond to changes in other physiological, psychological, sociological or environmental factors after third grade.

Student growth over the school year can be nonlinear. Results showed that math achievement grew more in the second half of the school year while reading achievement typically increased more in the first half of the school year than during the second half. ELA growth was linear. These nonlinear trends have possible implications on expectations of student progress through the year on relevant formative assessments such as the NWEA MAP test. Additionally, teachers may be advised to focus their instructional efforts during steeper growth periods to maximize these trends. More research should be done to examine how generalizable these findings are and how adapting learning at the beginning or end of the year in accordance with higher growth periods could increase overall understanding.

Summer loss. Our study confirms the findings of other studies suggesting that students lose knowledge and understanding over summer months, although the loss is higher in the second summer. Drops in achievement in both summers support our hypothesis of summer achievement loss across all three subjects. Interestingly, math achievement actually shows a small amount of growth after the spring of third grade; however, there is still a drop from where math achievement would have been in the fall of fourth grade if the third-trajectory had continued.

Patterns of summer loss are different for students of different genders, socioeconomic status, and ethnicity. In this study, these differences vary across grade levels, subject areas and demographic subgroups and are often small but statistically significant. One major limitation of the current study is the lack of data on students' involvement in summer activities or supplemental instruction. Without knowing what students were involved in, the patterns detected in this study could reflect intervention effects. More information on summer instruction could provide stronger and clearer patterns across grade levels, subjects and student subgroups.

Gender differences. Females start slightly ahead of their male peers in ELA and reading at the beginning of third grade, but their male peers eventually catch up in fourth or fifth grade. Conversely, males and females start third grade at approximately the same level in math, but male students pull slightly ahead in third grade and the resulting gap is maintained in fourth and fifth grades. Based on recent research and policy involvement in STEM education for women and girls, these trends would most likely be seen in science related fields as well. Our study confirms the need for additional resources for female students in math. Based on the additional finding that female math students lose less over summer months, remedial programs for STEM education may work well for female students during the summer rather than during the school year.

SES differences. The greatest SES differences were seen in reading; however, in contrast with previous research findings, low SES students were predicted to grow faster than their peers in both reading and ELA. As mentioned above, this finding likely reflects a limitation of this study as data collection did not include potential remedial programs for low SES students in any subject. Higher scores for low SES students may be an intervention effect rather than a reflection of natural trends.

Clear trends for SES differences in ELA and reading over summer months are difficult to interpret clearly; however, it seems that SES has a greater negative effect on summer loss in the second summer. Further research should investigate how summer loss might be aggravated later in students' schooling, particularly for low SES students. Interestingly, math growth was not significantly different between SES groups until fifth grade. SES differences may continue to increase beyond fifth grade as well but would require further investigation.

Ethnic differences. This study found noteworthy trends in Native American students in comparison with other minority students. The low initial achievement and the declining growth of Native American students is sobering. Across subjects, Native American students start third grade well below their peers, and the gap continues to widen across grade levels. The widening happens predominantly during the school year in math and predominantly during summer months in ELA and reading. Greater resources and investigation should be invested in understanding these trends and intervening to help Native American students succeed.

Hispanic students start significantly lower than their white peers but grow in parallel until fifth grade in math. In reading and ELA, Hispanic students show steeper gains during the school year but greater loss than their white peers in summer months. Summer intervention, then, may be more beneficial for Hispanic students than other students based on these patterns.

Black, Asian and other minorities were classified in the same category based on the low representation in these two districts in comparison with Hispanic and Native American minorities. This limits the interpretation of the results of this parameter since the patterns between these ethnicities can vary greatly. For example, this group was not significantly different from white students in math. If Asian students excel in math above their white peers, while Black students score lower than their white peers, then grouping these two categories together may cancel any detectable trends.

Comparisons between subjects. Math, ELA and reading achievement are related, with ELA and reading being more closely related to each other than with math. Correlations also suggest a diminishing relationship between subject areas across distal grade levels. While ELA and reading growth may be highly correlated in the same grade level or consecutive grade levels, a third-grade ELA trajectory has only a small correlation with reading growth in fifth grade, and we would hypothesize the diminishing relationship continues with higher grade levels.

This pattern is found with math as well. Third-grade math growth is less correlated with fourth- and fifth-grade reading growth, and the third-grade reading slope has no relationship with fourth- or fifth-grade math growth. More data for trends in subsequent grade-levels (e.g., sixth and seventh grades) would need to be analyzed to explore if this pattern is replicated in subsequent years.

Interestingly, the correlations of initial intercepts for ELA or reading with other estimates are maintained across grade levels. For example, the ELA intercept correlates highly with each of the other reading estimates (i.e., third-grade slope, summer 1 offset, fourth-grade slope, summer 2 offset, and fifth-grade slope) at .67 or above. Again, additional data collection would be required to examine if any decay occurs in these relationships over time. Current data suggests that students' entrance scores in reading and ELA could have more predictive power of subsequent ELA and reading growth than previous grade-level growth or summer loss.

This pattern of preserved correlations with the intercept is also seen between the math intercept and subsequent ELA and reading growth. Reading and ELA intercept scores have a diminishing relationship with math growth. The correlations between ELA intercept and math growth are not statistically significant after the third grade. The correlations between the reading intercept and math growth are not significant starting in the fourth grade.

As part of the multivariate analysis, the variance at each level was computed for each grade and subject separately. Interestingly, the school-level intercept variance is much higher than any other variance estimated. Additionally, the school-level slope variance is typically larger than the teacher-level slope in reading and ELA, suggesting that a larger proportion of variance in student growth was due to school-level differences than teacher-level differences. These patterns persist across grade-levels and subjects and may be more tied to neighborhood characteristics rather than school related variables such as principal effectiveness. It follows reason that the variance between teachers or classrooms would be small at the beginning of the school year when students have only been assigned to their teachers for a short time; however, we would expect the growth of students during the school year to be more related to teacher differences than school differences.

Interestingly, the classroom or teacher-level variance of math slopes is larger than ELA or reading teacher-level variance, suggesting that the teacher or classroom may play a larger role in student growth in math than in ELA or reading.

Measurement Considerations

In building the level 1 and level 2 models, the greatest improvement in data-model fit occurred when adding the summer loss variables, allowing for a more accurate estimation of initial achievement in fourth and fifth grades. In contexts such as teacher or principal accountability studies where the purpose is to measure a students' growth over a school year, having accurate estimates of summer loss and initial achievement is pivotal to correctly measuring student growth. In a research context, studies that exclude a measure of summer loss greatly reduce their measurement quality and their potential to find desired intervention effects.

The use of a multivariate model allowed us to confidently compare and correlate effects across subjects. However, using multiple outcome variables complicates the model estimation, increases computing requirements, and limits other model variations that can be estimated simultaneously. For example, it would be completely impractical to attempt to estimate a cross-classified model in a multivariate analysis.

Incorporating a cross-classified model also introduced limitations in current computing power. Cross-classified models typically only consider one cross, high schools with colleges, for example, whereas this study required two crosses – one for each grade transition. Given the finding that the teacher-level variance within a school year is low there may be a practical advantage to excluding the teacher level as is commonly done in the growth literature, and instead include schools as a third level in longitudinal models. More empirical work should be done to explore the cost of excluding the teacher level in hierarchical models.

Limitations

Test occasions were assumed to be equidistant within school years for the analysis, meaning the elapsed time in months between the first and second test occasions was assumed to be the same as the time between the second and third test occasions within each school year. Based on this assumption, the interpretation of the results is based on conceptual time rather than growth over weeks or months. This limits the interpretation of the resulting estimates, but allows the results to make important contributions to the literature on student growth.

The use of free or reduced lunch status as a proxy for socioeconomic status in this study follows common research practices but can be problematic. A more comprehensive measure of

SES would include the parent's educational attainment and occupational status. Since the dataset contained no other information beyond free or reduced lunch status, our findings related to growth trends for students from low SES backgrounds may not be generalizable.

As mentioned above, a major limitation in data collection was the exclusion of potential moderating variables such as interventions for low SES students or other specific student subgroups. Incorporating these data would strengthen the claims and patterns found in this study.

In incorporating school-level clustering, our estimates are limited by having only 13 schools within the two districts. A larger number of schools would result in more stable estimates of school-level variance.

Practical Implications

The sample included in this study looked specifically at district-level data in two rural Utah districts. The success of finding significant results encourages the use of multilevel modeling in district level analyses, when available, to examine growth trends among students and the potential effects of educational programs implemented within a district.

As mentioned earlier, modeling and distilling patterns across grade levels and subjects is difficult. Practically, creating policies around minimum requirements for growth for teacher and principal accountability would be incredibly complicated and taxing on resources. Additionally it would likely not be transparent to the many stakeholders involved in public educations such as parents, teachers, and students due to the very technical requirements of doing such an analysis.

Rather than using growth as an accountability measure, growth modeling would greatly enhance policies and data-driven decisions at the district level. Similar data analysis could greatly inform target populations for interventions or other resources and would be able to detect effects of planned interventions.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167–180. doi:10.1177/000312240707200202

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J. … Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology*, *31*, 411–427. doi:10.1080/02702711.2010.505165

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*, 448–452.

Bast, J., & Reitsma, P. (1997). Mathew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, *32*, 135-167.

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools. *Child Development 83*, 1347-1367.

Betebenner, D. W. (2011). *Student growth percentiles and percentile growth projections/trajectories introduction: Why student growth?* Dover, NH: The National Center for the Improvement of Educational Assessment.

Briggs, D. C. (2011). *Making inferences about growth and value-added: Design issues for the PARCC Consortium [White Paper]*. Retrieved November 19, 2013.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, *97*, 65-108.

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*, 1–31. doi:10.1177/003804070407700101

Catts, H. W., Bridges, M. S., & Little, T. D. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research, 51*, 1569-1579.

Chaplin, D., & Capizzano, J. (2006). *Impacts of a summer learning program: A random assignment study of Building Educated Leaders for Life (BELL)*. The Smith Richardson Foundation & William T. Grant Foundation.

Cooper, H. (2004). Is the school calendar dated? Education, economics and the politics of time. In G. D. Borman & M. Boulay (Eds.), *Summer learning: Research, policies and programs* (pp. 3–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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*, 227–268. doi:10.3102/00346543066003227

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*, 613–635. doi:10.1177/000312240406900501

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

Han, W. J. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology*, *44*, 1572-1590.

Harris, D. N. (2011). *Value-added measures in education.* Cambridge, MA: Harvard Education Press.

Hipp, J. & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11,* 36-53.

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY: Routledge.

Holt, J. K. (2006). How do we measure up? Capturing the complexities of educational growth. *Mid-western Educational Researcher, 19*, 2-7.

Holt, J. K. (2008). Modeling growth using multilevel and alternative approaches. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 111-159). Charlotte, NC: Information Age Publishing.

Goldschmidt, P., Choi, K., & Beaudoin, J. P. (2012). *Growth model comparison study: Practical implications of alternative models for evaluating school performance*. Washington, DC: The Council of Chief State School Officers.

Grady, M. W. & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research, 45,* 393-419.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development, 82*, 1357-1371.

Guglielmi, R. S. (2012). Math and science achievement in English language learners: Multivariate latent growth modeling of predictors, mediators, and moderators. *Journal of Educational Psychology, 104*, 580-602.

Johnson, W., McGue, M., & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology, 42*, 514-532.

Judge, S. & Watson, S. M. R. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *Journal of Educational Research, 104*, 147-15.

Kieffer, M. J. (2012). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing, 25*, 1725-1746.

Kim, Y. S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading frequency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*, 652-667.

Kline, R. B. (2011). *Principals and practice of structural equation modeling* (3rd Ed.). New York, NY: The Guilford Press.

Mancilla-Martinez, J. & Leaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology, 102*, 701-711.

Marks, A. K. & Garcia Coll, C. (2007). Psychological and developmental correlates of early academic skill development among American Indian and Alaska Native youth: A growth modeling study. *Developmental Psychology, 43*, 663-674.

McCoach, D. B., O’Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children’s reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*, 14-28.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. New York, NY: Routledge.

Roberts, G. R. & Bryant, D. (2011). Early mathematics achievement trajectories: English-language learner and native English-speaker estimates, using the Early Childhood Longitudinal Survey. *Development Psychology, 47*, 916-930.

Roberts, G., Mohammed, S. S., & Vaughn, S. (2010). Reading achievement across three language groups: Growth estimates of overall reading and readings subskills obtained with the early childhood longitudinal survey. *Journal of Educational Psychology, 102*, 668-686.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

Suarez-Orozco, C. & Gaytán, F. X. (2010). Academic trajectories of newcomer immigrant youth. *Developmental Psychology, 46*, 602-618.

Von Drehle, D. (2010). The case against summer vacation. *Time*, *2*, 1–7.