Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this dissertation. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use in future works (such as articles or books) all or part of this dissertation.

Signature:

Molly Embree | Date

Individual Differences in Academic Trajectories from Elementary to Late Middle School: Influences of Gender, Ethnicity, and Income

By

Molly Embree
Doctor of Philosophy

Psychology

Carolyn F. Furlow
Advisor

Kim Wallen
Advisor

Nancy G. Bliwise
Committee Member

Robyn Fivush
Committee Member

Donna L. Maney
Committee Member

Elaine F. Walker
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Individual Differences in Academic Trajectories from Elementary to Late Middle School: Influences of Gender, Ethnicity, and Income

By

Molly Embree
B.A., Connecticut College, 1993
M.A., University of Mississippi, 1994
M.A., Emory University, 2005

Advisor: Carolyn F. Furlow, Ph.D.
Advisor: Kim Wallen, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2009

Abstract

Individual Differences in Academic Trajectories from Elementary to Late Middle School: Influences of Gender, Ethnicity, and Income

By Molly Embree

Longitudinal academic data for a cohort of elementary to middle school students from a small diverse (50% Caucasian, 44% African American, 5% other ethnicity) Southeastern public school were analyzed with hierarchical modeling (HM). The influences of and interactions among gender, ethnicity, and income on academic growth were examined. The hypotheses tested included: 1) achievement gaps exist at 3rd grade, 2) gaps grow from 3rd to 8th grades, and 3) gender differences are more apparent in analyses of the tail distributions than in averages of the overall distribution. Performance data were scale scores for language arts, math, and science domains, from a widely-used standardized test. Ethnicity and income were strongly correlated ($r(630)= 0.79$, $p<.01$). Performance gaps by ethnicity were found at 3rd grade in language arts ($\beta_{02} = -20.61$, $p<.05$) and science ($\beta_{02} = -17.58$, $p<.05$), and differences in linear and quadratic growth by ethnicity and income opened a performance gap in math by 8th grade ($\beta_{22}=-2.42$, $p < .05$; $\beta_{32}=0.19$, $p < .05$), with African American students averaging lower than Caucasians. Influences of ethnicity and income were large and stable in the tails, with low income and African American children over-represented in the bottom tail and nearly absent from the top tail over time. A significant interaction between gender and ethnicity showed African American boys scoring significantly lower in science than their peers at 3rd grade ($\beta_{04}=-20.90$, $p<.05$). The HM results predicted a significant acceleration in science growth by boys relative to girls after 6th grade ($\beta_{21}=-2.89$, $p<.05$; $\beta_{31}=0.25$, $p<.05$). Analyses of the tail distributions make clear that average differences obscure important information: sex differences differed by academic domain, by tier of performance, and changed with time. These complex results preclude simple generalizations of “male advantage” or “female advantage” by academic domain.

Individual Differences in Academic Trajectories from Elementary to Late Middle School: Influences of Gender, Ethnicity, and Income

By

Molly Embree
B.A., Connecticut College, 1993
M.A., University of Mississippi, 1994
M.A., Emory University, 2005

Advisor: Carolyn F. Furlow, Ph.D.
Advisor: Kim Wallen, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2009

Acknowledgments

I thank the members of my committee, Elaine Walker, Robyn Fivush, Nancy Bliwise, Donna Maney, Carolyn Furlow, and Kim Wallen for their thoughtful and helpful contributions. I thank Carolyn for advising on the hierarchical modeling, which made this project possible. I thank Nancy for many critical discussions at critical moments. I especially thank my advisor Kim for many years of patient and flexible support, for excellent professional advice and feedback, for far more lunches and dinners than any graduate student is entitled to, and for his persistent optimism that, as long as something is theoretically possible, it will probably turn out all right.

I thank Pat Marsteller for her personal and professional advice, determination, and encouragement, and I thank her and Jordan Rose, and all at the Emory Center for Science Education for the professional opportunities to finish my degree. The Emory University PRISM Graduate Fellowship and the GK-12 program of the National Science Foundation (Awards #DGE0536941 and #DGE0231900) made it possible.

I thank the Psychology Department and the NAB faculty for their support, and in particular, I thank Michael Zeiler for his early confidence in me.

I thank Dr. Tjaard van Silvolde on behalf of his school district for the generosity of his collaboration. Logistics between schools and researchers are notoriously difficult; it could not have been easier or more of a pleasure.

I thank two undergraduate assistants Tim Wallen and Steven Sherrin for invaluable help in careful error-checking and speedy graphing, gathering materials and superlative attention to detail.

I thank James LeBaron for his enduring interest in Japanese and learning in general. You have made me a better teacher. I thank Charles LeBaron and The Ben Franklin Academy for the opportunity to work with James while I pursued my degree. It has been an important part of my professional training, and I continue to enjoy the challenge.

I thank lab members and friends for support and good times over the years: Julia Zehr, Tanja Jovanovic, Marlyne Israelian, Michelle Broth, Rebecca Herman, Franklynn Graves, Heather Rupp, Henry Lange, Janice Hassett, Shannon Stephens, and Elena and Michael Monarch. Janice, your help has been steady and expert, and you've tolerated my neatness for years – I'll find some way to repay you! Special thanks to Julia and Tanja, who came back to help me run the last mile of the marathon. I can't tell you how glad I am to have started and finished this project in your company.

I thank the writers of Wagtown for the alternate universe which has balanced the dissertation, for their imaginations, the space to breathe and create, and their love.

I thank my parents for, well, for everything. For your constant examples of life well lived, work well done, joy enjoyed. I thank you for your trust and confidence in me, your love spoken and unspoken, your expectations, and your hopes.

. . . e Jorge, eu te encontrei no início do caminho pelo doutorado, este caminho que me levou muito longe de você . . . ao chegar ao fim, eu acho você aqui me encorajando. Que maravilha, que felicidade. Vamos?

Table of Contents

List of Tables

List of Figures

The past hundred years of research on the development of cognitive abilities has illuminated one simple conclusion: it is a complex process, neither determined dichotomously for males and females nor dichotomously by nature or nurture. Serious assessment of the origins of potential sex differences in cognition must address a wide array of conflicting data and multiple theoretical perspectives, including socio-cultural, environmental, and biological influences (review: Halpern, 2000). Early adolescence has been suggested as a time of emerging or increasing differentiation of sex-typed behavior, including cognition and academic achievement, by academic researchers (e.g. Maccoby & Jacklin, 1974) and educational analysts (e.g. Willingham & Cole, 1997). Amid popular framing of the topic of whether a "boy crisis in education . . . [is] more real than the girl crisis" (Sacks, 2005), sound research of sex differences in cognition accrues more slowly than a concerned public would like to enact educational policy. Yet, neuroscience is advancing towards the goal of linking brain structure and activation to specific cognitive function, and multidisciplinary approaches can integrate these empirical data and theoretical perspectives from traditionally separated fields of biological and social psychology.

Neuroscience studies of the normal development and maturation of the human brain (as opposed to disease-based research) have increased from a marked scarcity to a steady flow in recent years. Emerging information documents the enormous change that the human brain undergoes during adolescence. The ratio of white matter to grey matter increases with age over the whole brain (Paus et al., 2001) through a series of regional increases over the frontal, temporal and parietal cortices, and interior structures, though it is not clear whether the increase in white matter is due primarily to increased intra-cortical myelination or “pruning” of grey matter (Paus, 2005). Furthermore, the pattern of structural maturation differs by sex, with girls often showing earlier maturation.

In parallel to these structural changes, executive function, including selective attention in decision-making, capacity of working memory, efficiency in information processing, and greater control inhibiting inappropriate responses, improves with age from childhood to late adolescence (Anderson, 2001; Blakemore & Choudhury, 2006; Paus, 2005; Steinberg, 2005). In addition to improving executive function, children demonstrate increasing social and emotional cognitive function throughout adolescence: they become more self-aware and able to regulate their behavior ahead of external sanction (Bussey & Bandura, 1999) and more able to infer others’ perspectives related to improved processing of facial expression of emotions (Blakemore & Choudhury, 2006). Importantly, peer social context becomes more salient during this developmental period than it had been in earlier childhood in influencing the real behaviors that result from multidimensional cognitive processing (review: Steinberg, 2005). Steinberg notes that while adolescents’ appraisal of hypothetical risk in a lab setting may not differ from adults’, their behavior is more influenced towards risk-taking by the presence of peers than adults. He speculates that the normal regional maturation of the brain may lead to changes in arousal and motivation prior to the development of integrated executive function, making early and middle adolescents vulnerable to heightened affect without adequate regulatory control. In addition, while structural information of neurodevelopment is rapidly increasing, few studies have addressed cognitive function in adolescents. While acknowledging important technical challenges to establishing relationships between brain structure and function, Paus (2005) and others (e.g. Casey, Giedd, & Thomas, 2000; Segalowitz & Davies, 2004; Stiles, Moses, Passarotti, Dick, & Buxton, 2003) find much promise in the technological advances allowing direct and concurrent examination of brain activity and behavior.

Sex differences in brain maturation rates imply possible functional consequences in cognition or cognitive processing. For example, sex differences in growth rates of the left inferior frontal gyrus (Blanton et al., 2004), may be related to differences in verbal fluency, one of the earliest developmental and best-documented behavioral sex differences (Hyde & Linn, 1988; Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997). Neufang and colleagues (2009) found sex differences in gray matter volumes of the amygdala (implicated in emotional processing) and hippocampus (central to working memory) that varied with pubertal stage and were associated with circulating testosterone levels in both boys and girls. They further infer that the negative association between circulating testosterone and volume of gray matter in parietal cortex may reflect a normal process of neuronal cell death as parietal cortex matures, associated with improving visuospatial function.

Though the action of prenatal hormones in causing anatomical sex differentiation of a developing fetus is not controversial, the action of prenatal hormones on the fetal brain to influence sex differences in cognition and behavior is wildly controversial, a “political minefield” for researchers and the public (Halpern, Wai, & Saw, 2005). Nevertheless, a large body of animal research (e.g. rat, mouse, guinea pig, rhesus monkey) has demonstrated that prenatal hormones influence behavioral sex differentiation (for review, see Wallen & Baum, 2002). In rhesus monkeys, prenatal androgens clearly influence the organization of several infant and juvenile social behaviors (Wallen, 1996; 2005). For humans as well, differential prenatal hormone exposure has been hypothesized to influence sex differences in cognition (reviews: Cohen-Bendahan, van de Beek, & Berenbaum, 2005; Halpern, 2000; Kimura, 1999). And yet, as Wallen (1996) demonstrated, behavioral sex differentiation in rhesus monkeys is shaped by the interaction of biological and social influences, whereby the degree to which sexually dimorphic behaviors are expressed depends on the specific social context in which the monkeys are reared.

Certainly, human society also exerts influence on the development of sex differences in cognition. Change in some academic performance trends by gender in national and international assessments just since data reviewed by Halpern (2000) surely suggests the influence of cultures in educating their children, whether promoting or mitigating gaps in performance. Thorough reviews of theories of sex differences in cognition increasingly acknowledge the interaction rather than counteraction of biological and socio-cultural influences (e.g. Golombok & Fivush, 1994; Ruble, Martin, & Berenbaum, 2006). Further, Bussey & Bandura (1999) and Halpern (2000; Halpern et al., 2007) propose theoretical models of reciprocal causation or interaction, whereby biological and environmental (including socio-cultural) factors influence cognition in a continuous feedback system.

For example, if a child has excelled in math from an early age and has had parental support and reinforcement for that talent, it is likely that he or she finds math rewarding and practices the behaviors that, through repeated practice, increase his or her math skills. At adolescence, that child is likely to become more sensitive to sanctions and rewards from peers for engaging in "appropriate behavior" (Bussey & Bandura, 1999). If a girl or an African American boy receives a different signal from the peer group about the appropriateness of demonstrating skill in math than a Caucasian boy receives, then that child is likely to choose behaviors more in line with social standing than with cognitive ability. Repeated choices to engage or not engage in activities which increase particular cognitive skills modify neural development as certainly as physiological influences on cognition; this is well-known as *practice effect*. While personal choice becomes increasingly influential in the development of adolescents, the importance of the opportunities afforded and promoted for adolescents by the socio-cultural environment (including parents, teachers, communities, and media) must not be overlooked. Interactionist models which recognize the simultaneous influences of physiology, psychology, socio-cultural and physical environments (as opposed to nature *or* nurture models) provide a mechanism for dynamic change throughout the lifespan of an individual, as well as immense variation across individuals, such as we note in cognitive abilities.

Most of the empirical data on human cognitive sex differences (often broadly categorized into verbal, quantitative, and visuo-spatial domains) come from studies of high school students, undergraduate students, and adults, though an early examination (Maccoby & Jacklin, 1974) suggested that they might emerge at around the age of puberty. Cross-sectional data from studies of academic performance in math and science also suggest that sex differences emerge in mid adolescence (Cheng & Seng, 2001; Cole, 1997; Fierros, 1999), while sex differences in verbal ability appear to arise as soon as speech production begins (Huttenlocher, 1991). It is not clear whether sex differences in verbal ability increase or decrease over adolescence, with conflicting results from cross-sectional studies (National Center for Educational Statistics, 2006; Georgia Governor's Office of Student Achievement, 2006).

Review of Educational Achievement Literature

Data relevant to assessing the emergence of sex differences in cognitive abilities must be pulled from a wide array of sources. Meta-analyses by broad cognitive domain, such as verbal and quantitative abilities, provide the most efficient overview. Data from national and international educational assessments also provide information at regular intervals, usually focused on academic performance in language, math and science at fourth, eighth, and twelfth grades. Most large scale studies have focused on high school and older students rather than elementary and middle school students. The number of small studies examining younger students has been few, but is recently increasing.

Examining verbal abilities, Hyde and Linn (1988) conducted a meta-analysis of 165 studies. They found a weighted mean effect size of *d* = 0.11, a small difference they argue indicates no real sex differences in verbal ability exist. They also found no developmental age trend of sex differences in verbal ability. However, Kramer and colleagues (1997) found significant sex differences in verbal ability in five age groups of children from 5-16 years old: at all five age levels, boys showed higher average performance on vocabulary, while girls showed higher average performance in verbal learning (recall and recognition tasks with distractors). Kramer and colleagues (1997) found overall small effect sizes (*d* = .27 to .32) for these tasks. Comparing across age levels, effect sizes increased systematically with age, though the small number of effect sizes to examine precluded statistical significance. Other examinations of language measures in young children have shown significant sex differences showing female advantage in achievement (Fiorentino & Howe, 2004; Huttenlocher, 1991) and growth (Huttenlocher, 1991; McCoach, O'Connell, Reis, & Levitt, 2006). These more recent studies conducted since Hyde and Linn’s 1988 review suggest that verbal sex differences are evident quite early in development, certainly well before puberty, though it is less clear whether the magnitude of these differences changes over adolescence.

Examining quantitative abilities in a meta-analysis of 100 studies, Hyde, Fennema, and Lamon (1990) found an age trend for girls’ advantage in mathematic computation in elementary and middle school, no sex difference in problem-solving in elementary school, and a growing math advantage for males in high school (*d*=0.29) and college (*d*=0.32).

The National Assessment of Educational Progress (NAEP: NCES, 2006) suggests that sex differences favoring girls in reading and writing exist from the beginning of school and remain stable or increase over time, while average male advantages in math already exist at fourth grade or emerge by eighth grade, and grow during high school. International cross-sectional studies such as the Third International Math and Science Study conducted in 1994-95 have suggested previously that a male average advantage in math and science emerges during middle school and increases until the end of high school (Cheng & Seng, 2001; Fierros, 1999). More recent administrations of TIMSS (Trends in International Math and Science Studies) have shown different findings at fourth and eighth grades (Martin, Mullis & Foy, 2008; Mullis, Martin, & Foy, 2008).

Many decades of data on the Scholastic Aptitude Test (SAT), Verbal (SAT-V), and Mathematics (SAT-M), have consistently shown higher average male than female achievement (Halpern, 2000). However, since an additional writing component was added to the SAT-V in 2005, girls are expected to gain the advantage in this domain (Mead, 2006). Analyses of sex differences in large cross-sectional datasets of secondary education “mental tests” (Hedges & Nowell, 1995; Nowell & Hedges, 1998) have examined older students, ages fifteen to early twenties. They have found small and moderate sex differences in language arts, math, and science (and a variety of other domains), and quite large sex differences in mechanical reasoning and electronics information.

Few studies have examined sex differences in primary and middle school achievement. Martin and Hoover (1987) examined sex differences in performance and variability on standardized tests of language arts and math in a sample of students followed from elementary to middle school from 1978-84. Boys' performance was more variable than girls' on all math and language arts subtests except vocabulary. In elementary and middle school, more girls scored higher on language arts subtests than boys across all levels of the performance distribution. Sex different performance on math subtests was more complex. On two of three math subtests, boys' performance advantage was limited to the high end of the distribution, where differences increased with age. On the computation subtest, girls' average advantage increased systematically with age. Beller and Gafni (1996) examined international cross-sectional data of 9 and 13 year olds, showing increased sex differences in math and science performance in the older students. Recently, Strand, Deary, and Smith (2006) found small significant sex differences in a large, nationally-representative U.K. sample of 11 and 12 year-olds on verbal, nonverbal, and quantitative reasoning measures, and greater variance in performance among boys than girls on all measures. None of these studies reported analyses by ethnicity or family income.

More attention has been paid to sex differences in math than to other academic domains, but much of this attention has focused on gifted students. Within the high end of math performance, sex differences appear to emerge earlier and remain robust over time (Casey, Nuttall, & Pezaris, 1997; Robinson, Abbott, Berninger, & Busse, 1996; Shea, Lubinski, & Benbow, 2001; Stumpf & Stanley, 1998). An examination of the whole spectrum of abilities is needed in order to illuminate the emergence of sex differences, as one of the expected patterns is that variance will be greater among boys than girls, with more boys performing at the highest and lowest ends of the spectrum.

In recent years, a few studies have used longitudinal designs to examine academic achievement. Most address only one academic domain (*verbal/language arts*: McCoach et al., 2006; *math*: Ai, 2002; Aunola, Leskinen, Lerkannen, & Nurmi, 2004; Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Ma, 2005; *science*: Dekkers, Bosker, & Driessen, 2000; Ma & Wilkins, 2002; Muller, 2001; Von Secker, 2002), and none of these studies captures the time frame from before to after puberty onset.

Zvoch and Stevens (2003; 2006) used multilevel modeling to examine math and language arts achievement in a large, diverse sample of American public middle school students with regard to sex, ethnicity, socio-economic status, English proficiency, special education status. They found that female, ethnic minority, and special education students had slower growth in math (but not language arts), over middle school. So, while female students began with higher average scores in math than boys at sixth grade, the girls' performance grew more slowly, and boys gained higher relative performance by eighth grade. Ethnic minority and special education students began sixth grade with lower average math scores than non-minority and general education students respectively, and their slower growth indicates an increasing average performance gap between these subgroups. Zvoch and Stevens (2003; 2006) did not have access to data prior to sixth grade, which means that puberty was well underway for most girls in this sample.

Rationale

Numerous calls for true longitudinal analysis, especially beginning in pre-adolescent children, have been made by researchers interested in the early patterns of cognitive development (Halpern et al. 2007; Kramer et al., 1997; Ruble et al., 2006). We know that hormones directly affect brain development and neuronal function (e.g. Birzniece et al., 2006). We know that experience and learning *also* affect brain development and neuronal function (Iaria, Petrides, Dagher, Pike, & Bohbot, 2003; Pych, Chang, Colon-Rivera, & Gold, 2005; Stiles et al., 2003). Hormones also affect motivation and interest (e.g. Wallen, 1990), which affect behaviors (choices), which in turn affect experiences and learning, causing subsequent changes in the brain (Figure 1). Thus, a feedback system is always operating to shape brain development and the cognitive capacities we might try to measure. It is obvious that, if adolescence is a time of massive change in the human brain, and also a time of increasing pursuit of personal choices in experience, there is huge potential to increase cognitive differences among adult individuals. Trends over time in cross-sectional data, especially those across different nations and cultures, are highly suggestive, but inconclusive by nature of being cross-sectional. Furthermore, these measurements come too late (mostly post-puberty) and too wide in interval (usually 4 years apart) to provide much information. A longitudinal approach which begins examining cognitive patterns well before the onset of pubertal hormones is necessary to document the possible emergence of cognitive sex differences at puberty.

In addition to this, it must be acknowledged that sex differences in academic achievement are generally small in comparison to differences by family income and ethnicity (NCES, 2006; Governor's Office of Student Achievement, 2006), which are often correlated in the U.S. (Muller, Stage, & Kinzie, 2001). Gaps in performance by ethnicity and income on the National Assessments of Educational Progress tests are routinely 5-10 times as large as gender gaps (NCES, 2006). Therefore, any account of systematic sex differences of 2-3 scale score points must also honestly examine systematic differences of 20-30 points by ethnicity and income.

The current study fills a need for examining developmental patterns in academic growth in a cohort of students from elementary through middle school within the important context of ethnicity and income. Longitudinal achievement data from a small Southeastern, urban public school district allow true growth trajectories to be modeled using hierarchical analysis, and to test hypotheses about the development of gender differences. The small size of the district offers a rare opportunity to examine developmental patterns using the entire school population spanning the whole range of abilities from special education to gifted programs. The data set captures a period from third grade (average age 8, prepuberty for most girls and virtually all boys) to eighth grade (average age 14, late puberty onset for most students). Individual growth curves for achievement in language arts, math, & science are examined for the relative influence of sex, ethnicity, income and interactions among these variables.

This study contributes solid data with which to evaluate some general claims about the emergence and developmental patterns of gender differences which have been made in the public and research arenas. In addition, these data provide a foundation for targeted research questions to be addressed in future studies. For example, how does clinically assessed puberty onset relate to the timing of changes in sex differences in math and language arts? That is, does puberty onset, above and beyond anatomical sex, influence the trajectories of academic performance of student subgroups by income and ethnicity? Does puberty onset precede or follow changes in academic interest or expectation of success? What educational interventions successfully prevent or overcome performance gaps? Clearly, the ultimate practical purpose for this analysis is to allow instructional interventions early enough in academic development to preserve opportunities for students to pass courses, to achieve well-rounded competence for high school diplomas, and to prepare for college and the profession of their choice. Towards that goal, we must first investigate and understand what factors influence the emergence of gaps, and when.

Research Questions and Hypotheses

Given the paucity of information on preadolescent populations sampled across a broad spectrum of ability, the questions to be addressed were basic: are there gaps by ethnicity, income level or gender at third grade? Does variance differ by gender as expected? Do rates of academic growth differ? Do differences in growth depend on the academic domain? Do they change over time? Are differences perceptible earlier or more prominently over time in the tails of the distributions?

Hypotheses for the Complete Distribution

1. Gaps by sex, ethnicity, and income already exist at 3rd grade, specifically:
    a. Average score is higher for Caucasian than for African American students, across all subjects
    b. Average score is lower for low income students, across all subjects
    c. Average score is higher for girls than boys, across all subjects, particularly in language arts
2. Ethnicity and/or income account for more variance in performance by domain than does sex
3. Ethnicity and income are significantly correlated within the sample
4. Boys' achievement in all three domains across all time points is characterized by greater variance.
5. Initial status (performance level at 3^{rd} grade) is an important predictor of 8^{th} grade performance in that domain.
6. Changes in achievement growth rates are greatest during middle school (grades 6-8)
7. Gaps by sex, ethnicity, and income are maintained or increase from 3^{rd} to 8^{th} grades, specifically:
    a. Ethnicity and/or income gaps increase across all domains by the end of 8^{th} grade

b. Girls maintain average performance advantages across all domains from 3rd to 8th grades

8. An interaction of sex and ethnicity, or sex and income, accounts for significant variance in: a) initial status, and b) growth rate

Hypotheses for the Tail Distributions

9. Significant differences in academic performance by ethnicity, income, and sex exist <u>at 3rd grade</u> in the tail distributions, such that:

   a. Average score of Caucasian students in tails is higher than African-American students' across all domains; African American students are more numerous in bottom tail, Caucasian students are more numerous in the top tail

   b. Average score of low income students in tails is lower than that of more advantaged students across all domains; low income students are more numerous in bottom tail, more advantaged students are more numerous in top tail

   c. Average score of girls in tails <u>at 3rd grade</u> is higher than boys' across all domains, with the greatest advantage in language arts; girls are more numerous in top tails across domains at 3rd (especially in language arts), boys are more numerous in the bottom tails across all (especially in language arts)

10. Gaps present at 3rd grade are maintained or increase by 8th grade, with the exception of a gender gap in math and science, which reverses. In particular:

a. Language Arts, top tail: girls outnumber boys and have a higher mean score from 3rd to 8th grades

b. Math & Science, top tail: boys outnumber girls by 8th grade; boys' mean score is higher than girls' mean score by 8th grade

c. All domains, bottom tail: boys outnumber girls; girls' mean score is higher than boys' mean score

11. At all time points, differences by predictors are larger in the tails than the middle distribution.

Method

Student Sample

Longitudinal student achievement data in language arts, math, and science were obtained from archival records of a small, urban, public school district in the Southeastern United States. The data describe the academic achievement patterns of a cohort of students from third to eighth grades, from academic years 2002-03 to 2008-09. This cohort is projected to graduate from high school in the spring of 2012.

The school district serves about 2400 students across four elementary schools, one middle school, and one high school. The school population is diverse (51% Caucasian, 43% African-American, 1% Asian, 2% Hispanic, and 4% multi-racial), with male and female students at 51% and 49%, respectively. About 35% of the school population qualifies for free or reduced-price lunch, for which eligibility is determined by family income. About 12% of the population are registered as students with disabilities, 23% are enrolled in gifted programs, and 2% qualify for assistance in learning English as an additional language.

In 2004, the district enacted a bold restructuring plan to address disparities among its several elementary schools, which, as neighborhood schools serving different communities separated largely by ethnicity and socio-economic status, had been de facto segregated. The district closed three elementary schools, and merged the entire school population into three k-3 elementaries and one 4-5th grade elementary school. The study cohort was the first cohort of the district to merge as one population at the fifth grade. Thereafter, the cohort has remained together as one population at the middle school (grades 6-8) and currently at the high school.

The school district receives Title 1 funding, and all schools in the district attained “Adequate Yearly Progress” as defined by No Child Left Behind provisions in three of the last four academic years (Governor’s Office of Student Achievement, 2009). Achievement in the district is routinely equal to or better than the state averages for the same measures.

Measures

The Iowa Test of Basic Skills (ITBS) is a norm-referenced, multiple-choice format test of educational achievement and is used widely in public schools for grades k-8. The ITBS has been developed by the research program in educational measurement at the University of Iowa over more than 70 years, and is designed to measure growth in vocabulary, reading, language, mathematics, science, social studies, and inquiry skills. Standardization procedures are designed to make the norming sample “reflect the national population as closely as possible, ensuring proportional representation of minority and socioeconomic groups” (Hoover, Dunbar, & Frisbie, 2003). The ITBS Form A (ITBS-A) was normed in 2000, based on a large, nationally representative probability sample. Internal consistency reliability estimates (K-R 20) for math and language arts domains are typically around .95 (Cross, 1998). Equated forms of the ITBS may be used to compare student scores across years of administration.

For the cohort sampled in this study, the ITBS-A was administered for academic years 2002-03 to 2008-09. The outcome measures analyzed here are the scale scores for overall language arts, overall math, and science achievement. In the normed national sample, scale scores between 100 and 400 represent developmental performance levels from kindergarten to ninth grade. The raw score median for fourth grade students in the national sample defines the scale score of 200 (Hoover et al., 2003).

A second set of achievement test data from the state Criterion-Referenced Curriculum Test (CRCT) were used as an independent measure of a student’s level of achievement for language arts, math, and science in a given year. Though the CRCT provides complementary measures of overall language arts, math, and science achievement, it is based on the state’s curriculum standards and is not a vertically equated instrument. It is important to note that CRCT scores were not analyzed as outcomes, but were used only to categorize a student’s overall performance within a domain as falling in the top or bottom quintile or middle distribution relative to the other members of the sample cohort in the given year.

Categorization of a student's ethnicity and gender were cross-verified from ITBS records to other school records, based on birth certificate and immunization record. For the purposes of this study, low income was defined by eligibility for the federally subsidized lunch program, which provides free or reduced price school lunches to students whose families earn up to 1.85 times the annual income designated as the poverty level. If the student was eligible for subsidized lunch in any year of the data set, that student was categorized as representing a low income family throughout the data set. Measures of socio-economic status, such as levels of parental education or type of parental occupation, were not available in the data set, which was limited to background information collected in routine achievement testing, de-identified, and protected in accordance with the Family Educational Rights and Privacy Act (FERPA, 2000).

Procedures

Data collection and de-identification

Archival student ITBS data were de-identified for secondary data analysis, in accordance with the Family Educational Rights and Privacy Act (FERPA, 2000) regulations and protocol approved by the Emory University Institutional Review Board and the participating school district.

Data selection criteria

The Class of 2012 cohort was selected because it had data for more time points than any other cohort. ITBS test data for language arts, math, and science were available for five time points: third, fifth, sixth, seventh, and eighth grades. The first criterion for inclusion in the data set was that the student have test data for two or more of the five possible time points. Second, data were included as long as testing accommodations followed a student's Individual Educational Plan (IEP) (e.g. large print for students with visual impairment, or a test reader for students with attention disorders) and guidelines strictly defined by the ITBS administration protocols.

The third criterion for inclusion in the data set was that the student have a matching set of achievement test data (by grade and by outcome measure) from the state Criterion-Referenced Curriculum Test (CRCT). In other words, a student's ITBS test outcomes in language arts, math and science for a given grade were matched with the same student's CRCT outcomes in language arts, math, and science for the same grade. Those students who had matched sets of data for two to five time points were included in the dataset. The district's high attendance rate overall and in particular for testing days resulted in about 80% of the ITBS dataset being matched with CRCT scores and remaining in the sample.

As sample sizes for Asian and Hispanic students within the cohort were too small ($n < 10$) to allow valid comparison, these students' test data were dropped from the sample. Therefore, ethnicity categories were limited to Caucasian and African American students, as reported on ITBS test documents and checked against independent student information records from the school district. Roughly 36% of the sample was eligible for subsidized lunch, permitting inclusion of low income as a variable. No student data had to be excluded on the basis of ambiguity of gender.

Design

The longitudinal data were organized using SPSS 15.0 software into a “person-period data set,” as Singer and Willett (2003) describe univariate format data files which contain multiple records for each individual. Each student has between two and five records over the six-year period, one record for each grade tested for both ITBS and CRCT. Thus, the data set exhibits a nested structure of observations across time within individuals. In the final data set, 630 observations were recorded for 164 students. Forty-nine percent of students had complete ITBS and CRCT test data across five time points. Nine percent of students had test data across four time points, and 19% and 23% of students had test data for three and two time points, respectively.

Model Specification

Hierarchical modeling was used to test hypotheses about student outcomes in language arts, math, and science by predictor and control variables over time. Students’ learning over time was assumed to occur, but it was not assumed that growth must be linear. Therefore, linear, quadratic, and cubic growth components were tested at Level 1. The assumption of normal distribution of error variance at Level 1 was tested as a preliminary step in analysis. Between-student variance in achievement level and growth were modeled at Level 2. After testing the unconditional model for linear and nonlinear growth components at Level 1, student-level variables were entered in the conditional model at Level 2 in order to test the hypotheses.

Unconditional Model.

Level 1

$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + \pi_{2i}(TIME)^2_{ti} + \pi_{3i}(TIME)^3_{ti} + \pi_{4i}(TOP)_{ti} + \pi_{5i}(BOTT)_{ti} + e_{ti}$

Level 2

$\pi_{0i} = \beta_{00} + r_{0i}$

$\pi_{1i} = \beta_{10} + r_{1i}$

$\pi_{2i} = \beta_{20} + r_{2i}$

$\pi_{3i} = \beta_{30}$

$\pi_{4i} = \beta_{40}$

$\pi_{5i} = \beta_{50}$

such that, for the outcome variable (language arts, math, or science):

Y_{ti}	is the achievement outcome at time t for student i
$(TIME)_{ti}$	equals zero for student i at grade 3
$(TOP)_{ti}$	equals zero for student i in performance midspread at time t; equals 1 for student i in top quintile at time t
$(BOTT)_{ti}$	equals zero for student i in performance midspread at time t; equals 1 for student i in bottom quintile at time t
π_{0i}	is the achievement initial status for student i
π_{1i}	is the linear growth rate for student i
π_{2i}	is the quadratic growth rate for student i
π_{3i}	is the cubic growth rate for student i

π_{4i} is the average difference of the top quintile contrast with the midspread (across all students and times, because equal to β_{40})

π_{5i} is the average difference of the bottom quintile contrast with the midspread (across all students and times, because equal to β_{50})

e_{ti} is the unique error term for achievement at time t for student i

At level 1, time-varying predictors TOP and BOTT were modeled dichotomously, such that the student received a value of one for TOP if he or she scored in the top quintile for that time point or zero if he or she scored in the middle or bottom quintiles (and conversely for BOTT). A student who scored within the middle three quintiles at a particular time point received a value of zero for both TOP and BOTT. Thus, the two dichotomously coded variables TOP and BOTT provide a means for contrasting top and bottom quintile performance with performance in the middle of the distribution as a control for examining the effect of sex in the conditional model.

Conditional Model.

Level 1

$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME})^2_{ti} + \pi_{3i}(\text{TIME})^3_{ti} + \pi_{4i}(\text{TOP})_{ti} + \pi_{5i}(\text{BOTT})_{ti} + e_{ti}$

(same as unconditional model)

Level 2

$\pi_{0i} = \beta_{00} + \beta_{01}(\text{SEX})_i + \beta_{02}(\text{ETH})_i + \beta_{03}(\text{FRL})_i + \beta_{04}(\text{SEXETH}) + \beta_{05}(\text{SEXFRL}) + r_{0i}$

$\pi_{1i} = \beta_{10} + \beta_{11}(\text{SEX})_i + \beta_{12}(\text{ETH})_i + \beta_{13}(\text{FRL})_i + \beta_{14}(\text{SEXETH}) + \beta_{15}(\text{SEXFRL}) + r_{1i}$

$\pi_{2i} = \beta_{20} + \beta_{21}(\text{SEX})_i + \beta_{22}(\text{ETH})_i + \beta_{23}(\text{FRL})_i + \beta_{24}(\text{SEXETH}) + \beta_{25}(\text{SEXFRL}) + r_{2i}$

$\pi_{3i} = \beta_{30} + \beta_{31}(\text{SEX})_i + \beta_{32}(\text{ETH})_i + \beta_{33}(\text{FRL})_i + \beta_{34}(\text{SEXETH}) + \beta_{35}(\text{SEXFRL})$

$\pi_{4i} = \beta_{40} + \beta_{41}(SEX)_i$

$\pi_{5i} = \beta_{50} + \beta_{51}(SEX)_i$

such that:

$\boldsymbol{\beta_{00}}$ **mean initial status achievement** for female, Caucasian, non-free/reduced lunch students

β_{01} average difference by sex in achievement status

β_{02} average difference by ethnicity in achievement status

β_{03} average difference by low income in achievement status

β_{04} portion of achievement status variance accounted for by the interaction of sex and ethnicity

β_{05} portion of achievement status variance accounted for by the interaction of sex and low income

$\boldsymbol{\beta_{10}}$ **mean linear growth rate** for female, Caucasian, non-low-income students

β_{11} average difference by sex in achievement linear growth rate

β_{12} average difference by ethnicity in achievement linear growth rate

β_{13} average difference by income in achievement linear growth rate

β_{14} portion of linear growth variance accounted for by the interaction of sex and ethnicity

β_{15} portion of linear growth variance accounted for by the interaction of sex and low income status

$\boldsymbol{\beta_{20}}$ **mean quadratic growth rate** for female, Caucasian, non-low-income students

β_{21} average difference by sex in quadratic growth rate of achievement

β_{22} average difference by ethnicity in achievement quadratic growth rate of achievement

β_{23} average difference by low income status in achievement quadratic growth rate

β_{24} portion of quadratic growth variance accounted for by the interaction of sex and ethnicity

β_{25} portion of quadratic growth variance accounted for by the interaction of sex and low income status

$\boldsymbol{\beta_{30}}$ **mean cubic growth rate** for female, Caucasian, non-low income students

β_{31} gender gap in achievement cubic growth rate

β_{32} ethnicity gap in achievement cubic growth rate

β_{33} income gap in achievement cubic growth rate

β_{34} portion of cubic growth variance accounted for by the interaction of sex and ethnicity

β_{35} portion of cubic growth variance accounted for by the interaction of sex and low income status

$\boldsymbol{\beta_{40}}$ the average difference of the top quintile contrast with the midspread (across all students, across all times)

β_{41} the effect of gender in the top quintile contrast with the midspread across times

$\boldsymbol{\beta_{50}}$ the average difference of the bottom quintile contrast with the midspread (across all students, across all times)

β_{51} the effect of gender in the bottom quintile contrast with the midspread across times

r_{0i} residual error term for student-level achievement after controlling for effects of all student-level variables (sex, ethnicity, low income, and two-way interactions)

r_{1i} residual error term for student-level linear growth rates after controlling for effects of all student-level variables (sex, ethnicity, low income, and two-way interactions)

r_{2i} residual error term for student-level quadratic growth rates after controlling for effects of all student-level variables (sex, ethnicity, low income, and two-way interactions)

At level 2, all student-level variables are dichotomous, such that:

SEX (0 = female, 1= male) is the predictor variable

ETH (0 = Caucasian, 1 = African American) is a control variable

FRL (0 = no free/reduced lunch, 1= free/reduced lunch = low income) is a control variable.

Estimation and Inference

Parameters for hypothesis testing were estimated from the model with HLM6 software, using Full Maximum Likelihood (FML) estimation. FML estimation computes estimates of population parameters which maximize the likelihood of fit among the sample data and the hypothesized model and its assumptions, and in contrast to Restricted Maximum Likelihood (RML) estimation, it may be used to test hypotheses about fixed effects as well as variance components (Singer & Willett, 2003). An alpha level of .05 was used as the criterion in all statistical tests.

Results

Descriptive Statistics

Mean scale scores for performance on the ITBS overall language arts, overall math, and science domains are presented by grade level in Figure 2. In the normed national sample, scale scores between 100 and 400 represent developmental performance levels from kindergarten to ninth grade. The raw score median for fourth grade students in the national sample defines the scale score of 200. As may be seen in Figure 2, the mean performance in this study's sample at third grade approximated the national median fourth grade performance in language arts, mathematics, and science. The mean performance at eighth grade in this sample attained or surpassed the national median ninth grade performance in all three domains.

Distribution curves for overall performance in language arts, math, and science are presented by grade level in Figure 3. Learning is apparent in all three domains, as the distribution curves shift upwards in scale from third to eighth grades. Bimodality appears to increase after sixth grade in all three domains. Distribution curves plotted by gender by grade (Figure 4) again exhibit some bimodality, particularly from sixth grade on, and more so for boys than for girls. Distribution curves separating the overall sample into four subgroups by ethnicity and gender indicate that the bimodality evident in the previous graphs is associated with ethnicity (which is associated with income in this sample). Mean performance by the four subgroups over time are presented in Figure 6, illustrating another view of the raw data.

Model Fitting

As a preliminary step to fitting hierarchical models, individual growth plots were inspected for linear and nonlinear shape. Random selections of 20 percent of the sample of individual student plots are shown in Figure 7. More nonlinearity can be observed in the plots for math and science than for language arts.

As shown in the model taxonomy presented in Appendix A, an unconditional means model (Pre) was fitted to the data simply to partition the outcome variance into that occurring within and between individuals. The resulting intraclass correlations (ICC) indicate that in language arts, math, and science, respectively, 54%, 37%, and 48% of the total variance is attributable to differences between students. Thus, 46%, 63%, and 52% of the total outcome variance in language arts, math, and science respectively, occurs within individuals (over time) at level 1 of the hierarchical model.

Next, linear time was added to fit an unconditional growth model (Model A in Appendix A) to the data for language arts, math, and science. This model predicts an outcome score Y_{ti} for a student i at time t as a function of that student's true score at third grade (π_{0i}) plus the unique deviation of error at time t for student i (e_{ti}) from the slope of that student's true change trajectory $\pi_{1i}(TIME)_{ti}$. As significant heterogeneity of variance at level 1 remained after modeling linear time, a quadratic time component was added in Model B

(Appendix A). Subsequently, a cubic time component was supported when slope variance was constrained to zero in Model D for all three academic outcome variables. Thus, Model D, containing linear, quadratic and cubic time components, became the unconditional growth model for comparison with conditional models containing student-level predictors at level two.

Model F in Appendix A is a conditional model which tested for main effects of three variables: SEX, ETHNICITY, and FRL (free/reduced lunch status as a measure of low income). The level 2 submodels test for the significant influence of these three variables across students' performance at third grade (initial status, π_0), linear growth rate (π_1), quadratic growth rate (π_2), and cubic growth rate (π_3). Results for Model F are presented in Appendix B. In language arts, significant heterogeneity of variance remained across students around initial status, and around linear and quadratic growth rates, after modeling these main effects. In math and science, significant heterogeneity of variance remained across students around initial status, after modeling the main effects of gender, ethnicity, and income level.

Model I in Appendix A adds two interaction terms to the conditional model to test the hypotheses that the interaction of gender with ethnicity and gender with income level may account for significant variance in students' third grade performance, or in their linear and nonlinear growth rates. Similar to Model F, Model I indicated significant residual heterogeneity of variance across students' initial status, linear and quadratic growth rates in language arts, but significant residual heterogeneity of variance around initial status only for math and science.

Finally, Model M in Appendix A is the fully conditional model, adding the control variables TOP and BOT at level 1 and the predictor SEX to level 2 to test the hypothesis that the effect of gender in the tails of the performance distribution may contrast with its effect in the overall distribution.

Main Findings

Hypothesis 1: Gaps by sex, ethnicity, and income already exist at 3rd grade.

At third grade, a significant achievement gap by ethnicity (controlling for sex, income, and interactions) already existed in language arts and science. As shown in Table 1 (A and C), the average score for African American students was lower than that for Caucasian students in both language arts (β_{02} = -20.61, p < .05) and science (β_{02} = -17.58, p < .05). In math, as shown in Table 1B, the average African American score was not estimated to be significantly different from the average Caucasian score at third grade (β_{02} = -11.98, *ns*).

Controlling for income and ethnicity at third grade, no significant sex differences were found in overall language arts, math, or science achievement. Controlling for sex and ethnicity, no significant gap was found by income at third grade.

Hypothesis 2: Ethnicity and/or income account for more variance in performance by domain than does sex.

The significant effect of ethnicity in two domains (language arts and science) with the absence of a gender gap supports this hypothesis

Hypothesis 3: Ethnicity and income are significantly correlated within the sample.

Ethnicity and income were strongly correlated within the sample ($r(630) = 0.79$, $p < .01$) over all time points. For example, 39 of 42 students who qualified for subsidized lunch in third grade were African-American, in contrast to eight of 74 students who did not receive subsidized lunch. This correlation appeared steady from third grade ($r(116) = 0.80$, $p < .01$) to eighth grade ($r(121) = 0.78$, $p < .01$).

Hypothesis 4: Boys' achievement in all three domains across all time points is characterized by greater variance.

As shown in Table 2, boys' variance in performance in math and science was greater than girls' at all grades except fifth. However, in language arts, girls' variance was greater than boys' except at third grade.

Hypothesis 5: Initial status (performance at third grade) is an important predictor of eighth grade performance in that domain.

Overall performance level at third grade was strongly related to linear achievement growth rate until eighth grade in language arts, math, and science, (τ_{01} = 0.92, 0.74, and 0.88 expressed as correlations, respectively).

Hypothesis 6: Changes in achievement growth rates are greatest during middle school (grades 6-8).

On average across students, more nonlinear growth was estimated in math and science than for language arts (Figure 8). Where nonlinear growth patterns were estimated, the greatest accelerations and decelerations appeared after sixth grade. Different patterns of nonlinear growth were estimated for subgroups of students.

<u>Hypothesis 7: Gaps by sex, ethnicity, and income are maintained or increase from third to eighth grades.</u>

This hypothesis was supported across all three domains for performance gaps by ethnicity, controlling for other variables, as shown in the model-based graphs in Figure 8. In language arts (Fig. 8A), though growth rates did not differ significantly by ethnicity (β_{12} = -6.61, *ns*; β_{22} = 0.754, *ns*; β_{32} = -0.039, *ns*), the cumulative effect of less growth per semester by African American students from third to eighth grades widened the initial gap. In mathematics (Fig. 8B), no initial ethnicity gap was estimated, controlling for other variables, but significantly different quadratic and cubic growth components by ethnicity predicted the emergence of a gap by eighth grade (β_{22} = -2.42, $p < .05$; β_{32} = 0.19, $p < .05$). In science (Fig. 8C), the initial 18-point gap in performance by ethnicity appeared to be maintained from third to eighth grades, controlling for other variables (β_{12} = -8.07, *ns*; β_{22} = -0.16, *ns*; β_{32} = 0.10, *ns*).

Support was not found for the hypothesis that girls would maintain an average performance advantage across all domains from third to eighth grade. In language arts, controlling for other variables, no significant gender difference was estimated at third grade (β_{01} = -7.52, *ns*) or emerged in linear or nonlinear growth rates by eighth grade (β_{11} = -4.89, *ns*; β_{21} = 0.97, *ns*; β_{31} = -0.06, *ns*). Similarly, in math, no gender gap was estimated at third grade (β_{01} = 3.62, *ns*) or emerged by eighth grade (β_{11} = -0.86, *ns*; β_{21} = 0.14, *ns*; β_{31} = -0.01, *ns*). However, in science, though no significant gender difference in performance was estimated at third grade or in linear growth rate (β_{01} = 1.15, *ns*; and β_{11} = 6.88, *ns*, respectively), significant differences in quadratic and cubic growth rates emerged after sixth grade (β_{21} = -2.89, $p < .05$; β_{31} = 0.25, $p < .05$), predicting greater average gains by boys relative to girls (Fig. 8).

In summary, the hypothesis of maintained gaps by ethnicity was supported across all three domains, but the hypothesis of girls maintaining an average performance advantage across all domains was not supported in the complete distribution, controlling for other variables.

Hypothesis 8: An interaction of gender and ethnicity, or gender and income, accounts for significant variance in initial status and/or growth rate.

In language arts and math, no significant variance in third grade performance or in academic growth was accounted for by interactions of gender with ethnicity or low income status (Table 1 A & B). However, in science (Table 1C), a significant interaction between gender and ethnicity was estimated at third grade, reflecting an average 21-point difference between African American boys and girls (β_{04} = -20.90, $p < .05$). This means that in science, where African American students were predicted to score on average 18 points lower than Caucasian students at third grade, African American *boys* were estimated to score an *additional* 21 points lower, controlling for other variables. However, linear and nonlinear growth rates were not estimated to differ significantly by this interaction (β_{14} = 2.03, *ns*; β_{24} = 2.11, *ns*; β_{34} = -0.27, *ns*).

Hypothesis 9: Within the tails of the distribution, significant differences in academic performance by ethnicity, sex, and/or income already exist **at third grade**.

Given limited power in the HM analysis, only one predictor (sex) could be added to the level 2 submodels for the top and bottom tails of the distribution for each outcome. The top and bottom submodels tested for a main effect of sex across all time points. Thus, the HM analysis does not pinpoint a third grade gender effect in the tail distributions, but takes a broader view. As indicated in Tables 3, 4, and 5, no significant effect of sex was found within the top or bottom tails across all time points for language arts, math, or science (language arts: β_{41} = 5.67, *ns*; β_{51} = -0.27, *ns*; math: β_{41} = 1.48, *ns*; β_{51} = 2.53, *ns*; science: β_{41} = -5.97, *ns*; β_{51} = 0.95, *ns*). Descriptive statistics indicated important differences by ethnicity, sex, and income within the top and bottom tails of language arts, math, and science distributions at third grade and across time (Tables 3-5 and Figures 9-14).

At third grade, the most striking difference by ethnicity was the ratio of students comprising each tail, given that 45% of the sample are African American. African American students were underrepresented in the top tail in language arts, and were nearly absent from the top tails in math and science. Conversely, African American students were overrepresented in the bottom tail of all three academic domains at third grade. African American student averages were lower than Caucasian student averages in the top tail of language arts, and in both tails of science. Averages by ethnicity were similar in the bottom tails of language arts and math.

As expected given the relatedness of income and ethnicity within the sample, disproportionately more students receiving subsidized lunch comprised the bottom tail across domains, while more students not receiving subsidized lunch comprised the top tail across domains. Though third grade language arts averages were similar by income within tails, in math and science, students receiving subsidized lunch showed lower averages than their counterparts in the top and bottom tails.

At third grade, girls' averages were hypothesized to be higher than boys' averages in both tails across all three domains but especially in language arts. Similarly, girls were hypothesized to be more numerous in the top tails and boys more numerous in the bottom tails across domains, particularly for language arts. Results differed by domain and by tail.

In language arts, effect sizes indicate a slightly higher average for boys in the top tail, and slightly higher girls' average in the bottom tail. As expected, girls outnumbered boys in the top tail and boys outnumbered girls in the bottom tail of language arts.

In math, effect sizes show no difference in average at the top tail, and a slightly higher average for girls in the bottom tail. The differences in proportion of girls and boys in the tails were opposite to expectation: more boys comprised the top tail and more girls comprised the bottom tail in math at third grade.

In science, expected differences in average were not supported, as girls showed lower averages than boys in both the top and bottom tails. However, the expected sex ratio was found, with more girls in the top tail and more boys in the bottom tail.

Thus, at third grade, hypotheses about proportions of students in the top and bottom tails by ethnicity, gender, and income were largely supported, while differences in average within a tail were less predictable.

Hypothesis 10: Within the tails of the distribution, gaps present at third grade are maintained or increase by eighth grade, with the exception of a gender gap in math and science which reverses.

Trends in top and bottom tail descriptive statistics support the hypothesis that the ethnicity gap present at third grade was maintained or increased by eighth across all three domains. In language arts, the average for Caucasian students was consistently higher than the average for African American students in both tails, and Caucasian students were more numerous in the top tail while African American students were more numerous in the bottom tail over time. In the top tail of math, there was a remarkable absence of African American students (only one at seventh and eighth grades), and the proportion of African American students in the bottom tail increased with time. Where averages could be compared, Caucasian student averages were higher. In science as well, African American students were nearly absent from the top tail over time, and were overrepresented at the bottom tail, where their averages were generally lower than those of Caucasian students also in the bottom tail.

A similar pattern indicates that students receiving subsidized lunch tended to have lower averages than their counterparts within the top and bottom tails of language arts, math, and science. Students receiving subsidized lunch were underrepresented in the top tails and overrepresented in the bottom tails of all three domains from third to eighth grades as well.

Support was mixed for hypotheses about gender differences over time in the top and bottom tails for each domain. As predicted, girls were increasingly numerous in the top tail of language arts over time. Girls' averages were higher than boys' in the top tail after third grade, though the difference diminished steadily towards negligible at eighth grade. In the top tails of both math and science, hypotheses of an increasing proportion of boys towards eighth grade were not supported. In math, the ratio of boys to girls in the top tail changed steadily from twice as many boys in third grade towards nearly equal numbers by eighth grade, opposite from the expected pattern. In science, the trend was not stable over time, but at eighth grade the proportion of boys to girls in the top tail was nearly equal. The hypothesis that boys' average in the top tail of math would be higher than girls' by eighth grade was supported by a slight trend after fifth grade. No stable trend appeared for differences in average by gender at the top tail of science, though at eighth grade, the average for boys was higher, as predicted.

Hypotheses about gender differences in the bottom tail found less support. In language arts, the hypothesis that boys would be more numerous in the bottom tail over time was confirmed, but no stable trend appeared for gender differences in averages at the bottom tail. In math, more girls were present in the bottom tail at third grade, but the proportion changed steadily towards the expected ratio of more boys than girls by eighth grade. No large differences or stable trend for higher average were observed in the math bottom tail over time. In science, boys were generally more numerous in the bottom tail over time, as expected. However, the trend for gender differences in average at the bottom tail in science was opposite from predicted: the girls' average changed from lower to higher by eighth grade.

Hypothesis 11: At all time points, differences by predictors are larger in the tails than in the middle distribution.

Differences were more robust for student ratios than for differences in average. Certainly by ethnicity and income, the ratio of students was more skewed in the tails than in the middle distributions of language arts, math, and science.

The gender ratio differed by academic domain and by stratum of the distribution. In language arts, the top tail was skewed towards more girls over time, while the middle distribution changed from more girls to more equal over time; the bottom tail maintained a stable skew of more boys. In math, the gender ratio became more equal in all strata over time. In science, the bottom tail consistently included more boys, while the middle distribution skewed towards more girls over time; the gender ratio in the top tail fluctuated roughly equally over time.

Thus, differences in subgroup ratios by ethnicity and income were much more robust in the tails than in the middle distribution in the predicted direction. Differences in gender ratio were more complex and differed by domain. Across all three academic domains, little support was found for greater difference in averages by predictors at the tails than in the mid-distribution.

Discussion

Variance and Variability

The current findings support the hypothesis that boys have more variable test performance than girls in math and science, but not in language arts, echoing similar findings by Maccoby and Jacklin (1974) and Feingold (1992). If one considers ratios of student subgroups and relative performance within tail distributions, one may argue that boys showed more variability across all three domains. Greater variability was initially visible in the increasingly bimodal distributions of boys across domains over time, largely related to differences in income level correlated with ethnicity in this sample. It is also evident at third grade across domains that boys were over-represented in the bottom tail, and, though fewer in number in the top tail, boys scored relatively well compared to girls at the highest performance levels. The most prominent example of boys' divergence into top and bottom scorers in this study was the significant interaction in science of sex by ethnicity reflecting a lower starting point at third grade for African-American boys, compared to Caucasian boys, with African American and Caucasian girls' average performance falling in between.

Machin and Pekkarinen (2008) examined cross-cultural patterns in average gender differences in test data from the 2003 PISA (Programme for International Student Assessment). They found that while average differences between 15 year-old girls and boys may vary widely across countries for reading and math, the pattern of greater male variance (strictly defined) was a robust finding for U.S. and international data, with variance ratios (boys' variance/girls' variance) consistently 1.13-1.20 for both reading and math. In comparison, several observations are striking about the current study's variance ratios (VR, similarly defined, in Table 2). First, where girls showed higher variance in language arts test scores, the ratios are close to equal. Second, the variance ratios in math and science are quite similar to those Machin and Pekkarinen report, with a steady increase from fifth to eighth grades. Third, in science, the steady increase in boys' variance from fifth to seventh was followed by a spike at eighth grade, indicating that some boys were progressing rapidly in science, while others lagged near the bottom; apparently, the significant sex by ethnicity interaction at third grade translated into quite different growth rates by eighth grade.

And yet, it is worth revisiting the finding that doesn't strictly match the "greater male variance" pattern: girls in language arts. Since this pattern has been found before, it may reflect important information. In this sample of third to eighth graders, boys showed higher variance at third, and then variance by gender nearly equalized. Perhaps this reflects a "catching-up" by boys at the low end of the spectrum after fifth grade. Of the three domains examined here, language is the most accessible and practicable at home, regardless of family education or income. Perhaps the slightly larger variance of girls' performance in middle school was influenced by greater numbers of girls accelerating at the top tier. The variance ratios reported by Machin and Pekkarinen are for older students and for the subdomain of reading. This context is important for interpretation, as well as whether variability by gender refers to variance of test scores, student ratios, and/or effect sizes within the tails of the distribution. When variance is examined as a proportion of its mean as with a coefficient of variation ratio (CVR, in Table 2), variability again appears larger for boys across all three domains, including language arts. The view of rapidly increasing variability among boys in science in late middle school does not change as a result of this different view of variability.

Early Differences

Ethnicity and Income

The expected gap by ethnicity (correlated with income in this sample, as is common in the U.S.) was found at third grade for language arts and science. Unfortunately, lower performance across all academic subjects remains consistent with a long-standing body of national testing data for students of African American and Hispanic ethnicity, students whose mothers have less formal education, and students from families of income below the poverty threshold (NSB, 2008). Important context for this sample is that the cohort was separated largely by ethnicity in different neighborhood schools during its early elementary years, and was integrated as one school population in fifth grade. Thus, ethnicity in this sample is not only correlated with income, but related to early school effects, such as the overall percentage of students who qualify for subsidized lunch (Zvoch & Stevens, 2006), and number of school resources, which have also been shown to influence individual outcomes (Ai, 2002).

In this study, the gap by ethnicity (correlated with income) is evident in the absolute averages by gender-ethnicity subgroup (Figure 6) as well as the model-based trajectories which control for other variables. The gaps by ethnicity and income in the tails of the distribution are accentuated as strong stable effects, often to the exclusion of disadvantaged students from the top level of performance. The significant main effects by ethnicity/income provide important context for the interpretation of findings by gender.

Gender

Though national and international data consistently find an average advantage for girls in language arts in absolute terms at fourth grade (PIRLS: Mullis, Martin, Kennedy, & Foy, 2007; NAEP: NCES, 2006), in this study, the effect of gender did not reach significance at third grade in the hierarchical analysis, controlling for other variables. National and international data regarding average gender differences in math and science in elementary grades are less clear, with some showing average male advantages (e.g. NAEP), some showing female average advantages (e.g. Hyde, Fenema, & Lamon, 1990; Martin & Hoover, 1987), and some no significant differences (TIMSS 1994-95: Fierros, 1999). Based on recent state testing data in language arts, math, and science for this school district, an average advantage for girls was expected across all three domains.

Whereas the hierarchical model view of average differences at third grade did not indicate any effect of gender controlling for other variables, examination of the tails of the distribution showed interesting differences by sex were already occurring:

In science at third grade, examination of the tails of the distributions reveals a strong male performance advantage (*d*=.75) in the top tier. A recently begun longitudinal study of U.S. elementary students found an average male advantage in science a third grade, the first time the students were assessed (Princiotta & Hausken, 2006; Rathbun, West, & Hausken, 2004)

In math at third grade, the striking gender difference was that boys were more numerous in the top tier, and though girls were more numerous in the mid-distribution, boys in the mid-distribution scored higher (*d*=1.10). Beller and Gafni (1996) and Martin and Hoover (1987) as well as much literature devoted to gifted students indicates that male advantages in gifted populations are visible much earlier than for the average tier.

However, a male performance advantage (*d*=.38) in the top tier of language arts at third grade was unexpected. Though boys have shown performance advantages for vocabulary as early as five years old (Kramer et al., 1997), and in verbal analogies in high school students taking the SAT-V (reviewed in Halpern, 2000) national and international data consistently show higher performance by girls in the global language arts domain, in overall average and performance at the top tail.

Differences by Ethnicity/Income Widen Over Time

The latest published information on trends in national samples (NSB, 2008) indicates, disturbingly, that gaps present in early education usually widen rather than narrow. Analysis of the Early Childhood Longitudinal Study data showed that performance gaps as early as kindergarten (in math) and third grade (in science) were still widening at fifth grade (Princiotta & Hauskin, 2006; Rathbun et al., 2004) and the Educational Longitudinal Study data showed that performance gaps at tenth grade were wider at twelfth (NSB, 2008).

In this sample, initial gaps in language arts and science widened from third to eighth grades, and where no significant gap existed in math at third, one opened up as the students entered middle school. Examination of the tails of the distribution over time makes clear that the effects of ethnicity and/or low income were large and rather stable at all tiers of the performance spectrum from third to eighth grades across all three domains. Clearly, differences in performance related to ethnicity and income precede puberty in this and national samples.

Gender Effects Differ Over Time by Domain and Performance Level

In contrast to rather stable effects of ethnicity and income over time in this study, effects by gender were more complex. Gender effects in performance differed by academic domain and by tier of performance examined, and these patterns changed over time. Some gender effects were perceptible at third grade within the tails of the distribution, though not perceptible in the HM overall averages, such as boys' greater variance and greater presence in the bottom performance tier. Other gender effects were evident only across the time frame from third to eighth grades, again more prominently in the extremes of the performance distribution than in average differences. As longitudinal data for this age group are lacking, results of the current study are interpreted in the light of findings from cross-sectional studies of fourth and eighth graders in national and international samples.

Examining Differences in Growth by Curved Averages and by Average Curves

The academic growth results presented in Figures 6 and 8 reflect two quite different views of the current study's data. Figure 6 presents subgroup's mean performance by grade level plotted from the raw data, without controlling for other variables; these are the curved averages, or the "data-based plots" according to multilevel analysis terminology. Figure 8 presents the average growth curve for each subgroup, based on the hierarchical model's estimation of fixed effects and variance components; these curves are "model-based" and reflect average trajectories controlling for other variables. This distinction is important, as the two views are not necessarily expected to look similar. In addition, there are no published data directly comparable to the current study's sample, with regard to multilevel analysis of one cohort's growth in three academic domains, examining gender effects while controlling for effects of ethnicity and income. National and international cross-sectional studies most often test for significant average differences in absolute terms, without controlling for other variables; thus, comparison of the data-based curved averages of the current findings with these studies may be most appropriate.

Looking across the raw data-based plots (Fig. 6) for language arts, math, and science, the most notable performance divergences by gender appear limited to African American students (correlated with lower income in this sample). The data-based plots show that African American girls scored higher on average than African American boys by eighth grade across all three domains and most dramatically in science. This uncontrolled view of the raw data suggests a widening gap between girls and boys for both Caucasian and African American students. Otherwise, the raw data show strikingly similar averages for Caucasian boys and girls in math and science over time. The model-based graphs largely agree that average performance by Caucasian girls and boys was similar, especially in math and science. The model-based trajectories estimated a significant effect of sex, controlling for other variables, only in science, predicting an acceleration in growth for both Caucasian and African American boys towards the end of middle school. Examining the model-based trajectories in Figure 8 (C), it appears that if the gap by ethnicity/income was narrowing towards eighth grade, it was due more to gains by African American boys than by African American girls, which differs drastically with the (uncontrolled) view by data-based curved averages (Fig. 6C).

The single significant model-based prediction of a gender difference in science growth contrasts with long-term national data, which consistently find that U. S. girls' average performance in language arts is higher than that of boys, and that boys' average performance is significantly higher than girls' at fourth and eighth grades (NAEP, TIMSS US data, PIRLS US data). International data indicate that girls' average test advantage in language arts at fourth grade and among 15 year-olds is a robust phenomenon within most countries as well as cross-culturally (PIRLS, PISA). However, it is also important to note that the long-term U.S. pattern of average male advantage in math and science at grades

4, 8, and 12, is *not* the typical pattern observed cross-culturally. The Third International Math and Science Study, conducted in 1994-95, showed few significant differences within countries at eighth grade but a significantly higher male average across the group of countries, and a strong male average advantage in math at twelfth grade (Fierros, 1999).

Further, the subsequent three cycles of TIMSS (changed to *Trends* in International Math and Sciences Study) over 1999, 2003, and 2007, have presented a cross-cultural pattern of gender differences from which the United States often differs. Because cross-sectional views of average performance by gender at fourth and eighth grades form the bulk of the data for comparison with the current longitudinal trajectories, it is worth examining in a little detail how patterns of change are similar or different across nations. A brief summary follows of gender differences in math and science from the current TIMSS data.

For math at fourth grade, no significant difference was found across 36 countries in average math performance by girls and boys. About half the countries individually showed no significant difference between boys and girls; eight countries showed girls averaging higher than boys, and twelve (including the U.S.) showed boys averaging higher than girls. At eighth grade, girls' average math performance was significantly higher than boys' across 49 countries; individually, 16 countries showed girls' averages higher than boys and eight countries showed boys' averages higher than girls', and in this case, no significant differences were found between U.S. girls' and boys' averages.

In science, girls scored significantly higher on average than boys at fourth and eighth grades across 49 countries. About half of the nations individually showed no gender differences at eighth grade, 14 showed an average female advantage and 11 (including the U.S.) showed a male advantage.

What is particularly interesting is that the top performing nations in math and science (Singapore, Chinese Taipei, Japan, Korea, and Hong Kong SAR) and the United States often differed from the cross-cultural patterns in different ways. As seen in the summary above, it is more common to find an average advantage for girls cross-culturally at grades 4 and 8 (whereas a male average advantage continues to be more common at the end of high school). In the most current three cycles of TIMSS, the top-performing nations most often show no significant gender differences at fourth or eighth grades in math and science, while the U. S. usually shows average male advantages in these domains. (One important exception is very top performer Singapore, where no significant differences are found frequently, but when they are, they systematically show female average advantage).

Thus, while the current study's data are interpreted with reference to U.S. cross-sectional and historical data, a more international perspective on gender differences across domains at different ages is important context for considering theoretical perspectives on how the physiology of puberty and socio-cultural factors may influence the patterns we observe. Certainly, differences across nations may point to important roles of curriculum standards, classroom differences, extra-curricular opportunities, and freedom of personal choice in study on the range of student differences.

Examining Differences in Growth by Performance Tiers

A shift from discussing average differences in the complete distribution of the current study's sample to the comparison of girls' and boys' representation and scores in different tiers of performance allows a more dynamic view of the variance question we started with. Looking at plots of effect sizes (Figures 9, 11, and 13) and student ratios (Figures 10, 12, and 14), some familiar patterns are visible:

1. more boys were present in the bottom tail across all three domains, either consistently or increasingly over time
2. girls were more numerous in the top tail of language arts and averaged higher than top tail boys, but boys at the top tier were closing that score difference over time, even though they numbered fewer and fewer
3. boys in the top tails of math and science scored higher than their female top tail peers.

Other patterns are opposite to the most widely reported findings, including the numbers of girls and boys equalizing with time in the top tiers of math and science, and boys in the middle and bottom tiers of language arts averaging higher than their respective female peers.

Examination of the trends in the ratio of boys to girls at different tiers suggests that boys and girls migrated between performance levels differently over middle school. In math and science, the main migration appeared to be between the bottom and middle tiers, with more girls initially at the bottom moving into the middle, and more boys migrating from the middle to the bottom tier over time. In language arts, the major migration appeared to occur between the middle and top tiers, with more boys migrating out of the top into the middle (but those few remaining in the top scoring better over time), and more girls migrating from the middle to the top tier with time.

In summary, whereas ethnicity and/or income relate to academic performance in this sample in a large and steady way over time across domains, the relationship of gender to outcomes differs over time by domain and performance tier. This study finds support for boys consistently outnumbering girls in the bottom performance levels across domains, and for fewer, higher performing boys at the top tiers of the distribution, which equally suggests that girls are more numerous in the middle tier and the lower end of the top of the distribution. Unexpected findings included equalizing numbers of girls and boys in the top tiers of math and science and greater female variance in language arts.

Limitations and Implications for Theory, Method, and Practice

In overview, analyses of these longitudinal data from a single cohort of children entering puberty do answer the proposed basic questions about patterns of sex difference emergence in cognitive performance in three broad domains. Limitations of the data set, comprised of correlational data under statistical control and lacking accompanying measures of puberty onset and self-efficacy by domain, for example, preclude specific tests of theory. However, the current study's findings answer several general claims and draw attention to measures of variability as an issue to re-examine in the study of cognitive gender differences. Perhaps most importantly, an integrated theoretical view of the current study's observations offers multidisciplinary *and* targeted questions for future experimental work. The notion of multidisciplinary *and* targeted questions is not contradictory, but inherent to a reciprocal causation model recognizing socio-cultural, environmental, and biological influences on cognitive development.

As expected based on regional empirical data, African American ethnicity and low income are thoroughly confounded in this data set. Referring again to an interactionist model (Figure 1), the expected findings of lower average academic achievement by African American students in elementary school are consistent with theories of biological processes as a consequence of low income (e.g. poor nutrition, higher exposure to environmental hazards associated with substandard housing) as well as socio-cultural processes (e.g. different access to resources and exposure to stereotyping). Both types of influences are likely real rather than theoretical in this cohort's development. Though the dataset did not include a measure of socio-economic status (SES), local demographics indicate that Caucasian students disproportionately come from higher SES families associated with universities and research institutions, while African American students disproportionately come from poor and lower SES families. High SES children are more likely to develop cognitive abilities with continuous access to good nutrition. With regard to influences of stereotyping within peer social context, it is important to note that the Caucasian and African American students in this cohort had been largely isolated from each other until the school population merged at fifth grade. Stereotype threat is the phenomenon of lower performance than demonstrated ability in a situation where one has been reminded that one's failure will confirm a negative stereotype (Dar-Nimrod & Heine, 2006; Steele, 1997). It is quite conceivable that African American students in this cohort experienced a sudden increase in stereotype threat within the new context of a mixed school population, particularly given the average discrepancy in income associated with ethnicity. Now that national longitudinal studies have begun for early childhood academic performance, large sample data sets will allow hierarchical analyses with better power to differentiate effects of ethnicity and income in academic growth patterns. Additional types of data are required to address questions of stereotype threat.

The variable "sex" in this study equally well represents physiological sex as it represents experience of gender. The socio-cultural environment has expectations of a child based on anatomical sex and can impose experiences which influence brain development (e.g. parental choice to send a child to dance or computer camp or the difficulty of a teacher's question), as well as enhance or suppress a child's choices and behaviors. It is interesting to note, however, that the variable sex in this study picks up change over time in more subtle and striking ways than ethnicity/income. It is likely that the variable "sex" in this study also represents the *interactions* of 1) physiological sex with puberty and 2) experience of gender with puberty. As such, patterns observed (e.g. the far lower achievement in science of African American boys with respect to African American girls compared to Caucasian boys and girls) provide some questions for experimental work (within the bounds of stringent ethics) to examine developmental aspects of stereotype threat, and expectancies-values theory.

For example, a well-designed experimental study might subtly manipulate exposure to a stereotype prior to a test on the variable of interest (e.g. algebra), while taking advantage of the natural physiological development of children. If a reliable, noninvasive indicator of pubertal status (e.g. salivary hormones) could be collected periodically as well as measures of the children's interests in the subject matter, value attributed to it, and expectations of success, one could begin to address the dynamic influences of expectations/values attributed, performance outcome, puberty onset, and stereotype threat. It is plausible that puberty onset heightens one's susceptibility to stereotype threat, and this may predict different outcomes at different times for African American girls and Caucasian girls and African American and Caucasian boys. Alternatively, if expectations of success predict performance outcomes regardless of puberty or stereotyping, this should also be demonstrable from such an experiment. Only experimental work designed from such multidisciplinary perspective can show how particular elements of theories (biological, socio-cultural, cognitive) relate to each other in explaining differences in cognitive performance at different levels of analysis.

From the current study's observations, we can answer the specific prediction that nonlinear growth in academic performance would be greatest in middle school by saying that it was, but it differed by academic domain. This may point back to the regional maturation of the cortex during adolescence, and we may hypothesize that nonlinear patterns of academic performance relate to the nonlinear brain maturation we are beginning to document. Beyond that, information is limited for specific predictions, especially looking at such broad cognitive domains. It is likely that more specific cognitive tasks are necessary to illuminate processing mechanisms.

Confirmation that differences are more readily perceptible in the tails of a performance distribution points to the importance of the shapes of distributions and not merely their dispersion. Further, the dynamics in the tails of odds-ratios, effect size, variance ratios, and student ratios illustrate different dimensions of trends in individual variation that are obscured by average differences. This observation has particular implication for the methods used to investigate sex differences in cognition.

The call by researchers for better assessment and reporting of variability is not new (Feingold, 1992; Hedges & Friedman, 1993; Rosenthal & Rubin, 1982), as variability itself may be an important dimension to sex differences in cognition. However, these appeals and even recent ones (Halpern et al., 2007) refer to measures and tests which depend on normal distributions of performance, overall and within each sex. The challenge, as shown in the current sample, is that particularly for populations starting with significant heterogeneity in performance, bi- or multi-modality may increase over time, as small but repeated effects of the external and socio-cultural environments plus individual choices and experiences develop individuals with disparate skills. As subsets of distributions are not necessarily normal, and not necessarily the same pattern of non-normal, it is appropriate to investigate other measures, such as standardized moments, which examine skewness and kurtosis in addition to dispersion around a mean.

What about practical applications? Clearly seen in the current sample and in other longitudinal studies of elementary students, a child's early development has already set some foundations that continue to influence his or her academic growth. Nevertheless, schools and society have an impact as evidenced by the full standard deviation gap between the top-performing nations' average TIMSS math scores and those of the United States. What can parents and teachers take from this study to best educate their children? An exciting genesis is underway in neuroscience, perhaps analogous to genetics research in the 1960's and 70's, where new tools and techniques are being put to use and information is rapidly accruing. However, the interpretation of findings takes time. We may be a long way from understanding how particular information is processed in the brain. What we do know does *not* suggest that different techniques should be used to teach boys and girls (differently). Rather, a variety of experiences and learning strategies should benefit all students, who, male and female, learn in multiple ways.

In the course of working out the mechanisms that translate structure and activation of the brain into cognitive function, it is certain that we will find differences in processing. Indeed, the differences provide the signal over the noise which helps to define the processing mechanisms. Most likely, we will see much individual variation in processing just as we see in cognitive performance, which defies simple explanation by sex. Our recognition of the multitude of influences generating individual variability does not mean abandoning a search for underlying principles, but should impel us to test hypotheses with attention to timing and specific, dynamic context.

References

Ai, X. (2002). Gender Differences in Growth in Mathematics Achievement: Three-Level Longitudinal and Multilevel Analyses of Individual, Home, and School Influences. *Mathematical Thinking and Learning, 4*(1), 1-22.

Anderson, V. (2001). Assessing executive functions in children: biological, psychological, and developmental considerations. *Pediatric Rehabilitation, 4*(3), 119-136.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. *Journal of Educational Psychology, 96*(4), 699-713.

Beller, M., & Gafni, N. (1996). The 1991 International Assessment of Educational Progress in Mathematics and Sciences: The Gender Differences Perspective. *Journal of Educational Psychology, 88*(2), 365-377.

Birzniece, V., Backstrom, T., Johansson, I. M., Lindblad, C., Lundgren, P., Lofgren, M., et al. (2006). Neuroactive steroid effects on cognitive functions with a focus on the serotonin and GABA systems. *Brain Research Reviews, 51*(2), 212-239.

Blakemore, S. J., Choudhury, S., Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. [Research Support, Non-U.S. Gov't Review]. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 47*(3-4), 296-312.

Blanton, R. E., Levitt, J. G., Peterson, J. R., Fadale, D., Sporty, M. L., Lee, M., et al. (2004). Gender differences in the left inferior frontal gyrus in normal children. *Neuroimage, 22*(2), 626-636.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. [Research Support, Non-U.S. Gov't Review]. *Psychological Review, 106*(4), 676-713.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology, 54*(1-3), 241-257.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: a comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*(4), 669-680.

Cheng, S. K., & Seng, Q. K. (2001). Gender differences in TIMSS mathematics achievement of four Asian nations: a secondary analysis. *Studies in Educational Evaluation, 27*, 331-340.

Cohen-Bendahan, C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neuroscience and Biobehavioral Reviews, 29*(2), 353-384.

Cole, N. S. (1997). *The ETS Gender Study: How Females and Males Perform in Educational Settings*: Educational Testing Service, Princeton, NJ.

Cross, L. H. (1998). Review of the Iowa Test of Basic Skills, Forms K, L, and M. *Mental Measurements Yearbook*, 13, [Electronic Version] from http://search.ebscohost.com/login.aspx?direct=true&db=loh&AN=13012057&site=ehost-live

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Science, 314*(5798), 435.

Dekkers, H. P. J. M., Bosker, R. J., & Driessen, G. W. J. M. (2000). Complex Inequalities of Educational Opportunities: A Large-Scale Longitudinal Study on the Relation between Gender, Social Class, Ethnicity, and School Success. *Educational Research and Evaluation (An International Journal on Theory and Practice), 6*(1), 59-82.

Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g (2000).

Feingold, A. (1992). Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy. *Review of Educational Research, 62*(1), 61-84.

Fierros, E. G. (1999, April). *Examining Gender Differences in Mathematics Achievement on the Third International Mathematics and Science Study (TIMSS)*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Georgia Governor's Office of Student Achievement. (2006). 2005-2006 K-12 Public Schools Annual Report Card. [On-line]. Retrieved January 24, 2006 from http://reportcard2006.gaosa.org/k12/content.aspx

Georgia Governor's Office of Student Achievement. (2009). K-12 Public Schools Annual Report Card. [On-line]. Retrieved March 30, 2009 from http://www.doe.k12.ga.us

Golombok, S., & Fivush, R. (1994). *Gender Development*. New York, NY: Cambridge University Press.

Halpern, D. F. (2000). *Sex Differences in Cognitive Abilities (3rd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbache, M. A. (2007). The science of sex differences in science and mathematics: *Psychological Science in the Public Interest* Vol 8(1) Aug 2007, 1-51.

Halpern, D., Wai, J., & Saw, A. (2005). Why Females Are Sometimes Greater Than and Sometimes Less Than Males in Math Achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender Differences in Mathematics: An Integrative Psychological Approach*. Cambridge, UK: Cambridge University Press.

Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63*(1), 94-105.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2003). Norms and Score Conversions: Student Norms and School Average Norms. The University of Iowa: Riverside Publishing.

Huttenlocher, J. (1991). Early Vocabulary Growth: Relation to Language Input and Gender. *Developmental Psychology, 27*(2), 236-248.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

Hyde, J. S., & Linn, M. C. (1988). *Gender Differences in Verbal Ability: A Meta-Analysis*: National Science Foundation, Washington, DC.

Iaria, G., Petrides, M., Dagher, A., Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *Journal of Neuroscience, 23*(13), 5945-5952.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*(1), 11-26.

Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: The MIT Press.

Kramer, J. H., Delis, D. C., Kaplan, E., O'Donnell, L., & Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology, 11*(4), 577-584.

Ma, X. (2005). A Longitudinal Assessment of Early Acceleration of Students in Mathematics on Growth in Mathematics Achievement. *Developmental Review, 25*(1), 104-131.

Ma, X., & Wilkins, J. L. M. (2002). The Development of Science Achievement in Middle and High School: Individual Differences and School Effects. *Evaluation Review, 26*(4), 395-417.

Maccoby, E. E., & Jacklin, C. N. (1974). *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.

Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*(5906), 1331-1332.

Martin, D. J., & Hoover, H. D. (1987). Sex differences in educational achievement: A longitudinal study. *Journal of Early Adolescence* Vol 7(1) Spr 1987, 65-83 Sage Publications.

Martin, M. O., Mullis, V. S., & Foy, P. (2008). TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

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*(1), 14-28.

Mead, S. (2006). The evidence suggests otherwise: the truth about boys and girls. [Electronic Version]. *The Education Sector, Biweekly Digest*.

Retrieved August 9, 2006 from http://www.educationsector.org/enewsletter/enewsletter_list_more.htm?issue_id=2686.

Muller, P. A., Stage, F. K., & Kinzie, J. (2001). Science Achievement Growth Trajectories: Understanding Factors Related to Gender and Racial-Ethnic Differences in Precollege Science Achievement. *American Educational Research Journal, 38*(4), 981-1012.

Mullis, V. S., Martin, M. O., & Foy, P. (2008). TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). PIRLS 2006 International Report. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

National Center for Educational Statistics. (2006). The Nation's Report Card, National Assessments of Educational Progress (NAEP), 1992, 1994, 1998, 2002, 2003, and 2005 Reading, Mathematics, and Science Assessments. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

National Science Board. (2008). *Science and Engineering Indicators 2008.* Two volumes. Arlington, VA: National Science Foundation (volume 1, NSB 08-01; volume 2, NSB 08-01A).

Neufang, S., Specht, K., Hausmann, M., Gunturkun, O., Herpertz-Dahlmann, B., Fink, G. R., et al. (2009). Sex differences and the impact of steroid hormones on the developing human brain. [Research Support, Non-U.S. Gov't]. *Cerebral Cortex, 19*(2), 464-473.

Nowell, A., & Hedges, L. V. (1998). Trends in Gender Differences in Academic Achievement from 1960 to 1994: An Analysis of Differences in Mean, Variance, and Extreme Scores. *Sex Roles: A Journal of Research, 39*(1-2), 21-43.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences, 9*(2), 60-68.

Paus, T., Collins, D. L., Evans, A. C., Leonard, G., Pike, B., & Zijdenbos, A. (2001). Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin, 54*(3), 255-266.

Princiotta, D., & Hausken, G. E. (2006). Fifth Grade: Findings From the Fifth Grade Follow-up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). E.D. TAB. NCES 2006-038. *SO - National Center for Education Statistics.*

Pych, J. C., Chang, Q., Colon-Rivera, C., & Gold, P. E. (2005). Acetylcholine release in hippocampus and striatum during testing on a rewarded spontaneous alternation task. *Neurobiology of Learning and Memory, 84*(2), 93-101.

Rathbun, A., West, J., & Hausken, E. G. (2004). From Kindergarten Through Third Grade Children's Beginning School Experiences. NCES 2004-2007. *SO - US Department of Education.*

Robinson, N. M., Abbott, R. D., Berninger, V. W., & Busse, J. (1996). Structure of abilities in math-precocious young children: Gender similarities and differences. *Journal of Educational Psychology,* 88(2), 341-352.

Rosenthal, R., & Rubin, D. B. (1982). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology, 74*(5), 708-712.

Ruble, D. N., Martin, C.L., & Berenbaum, S. A. (2006). Gender Development. In: Damon W, Lerner RM, (Eds.), *Handbook of Child Psychology*, (pp. 858-918). Hoboken, New Jersey: John Wiley & Sons.

Sacks, G. (2005, March 20). Boys or girls – pick your victim. *Los Angeles Times*, retrieved March 14, 2009 from http://www.latimes.com.

Segalowitz, S. J., & Davies, P. L. (2004). Charting the maturation of the frontal lobe: an electrophysiological strategy. *Brain and Cognition, 55*(1), 116-133.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of Assessing Spatial Ability in Intellectually Talented Young Adolescents: A 20-Year Longitudinal Study. *Journal of Educational Psychology, 93*(3), 604-614.

Singer, J. D., & Willet, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *American Psychologist, 52*(6), 613-629.

Steinberg, L. (2005). Cognitive and affective development in adolescence. [Review]. *Trends in Cognitive Sciences, 9*(2), 69-74.

Stiles, J., Moses, P., Passarotti, A., Dick, F. K., & Buxton, R. (2003). Exploring developmental change in the neural bases of higher cognitive functions: the promise of functional magnetic resonance imaging. *Developmental Neuropsychology, 24*(2-3), 641-668.

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: a UK national picture. *The British Journal of Educational Psychology, 76*(Pt 3), 463-480.

Stumpf, H., & Stanley, J. C. (1998). Stability and change in gender-related differences on the College Board Advanced Placement and Achievement Tests. *Current Directions in Psychological Science,* 7(6), Dec 1998, 192-196 Blackwell Publishing.

Von Secker, C. (2002). Effects of inquiry-based teacher practices on science excellence and equity. *Journal of Educational Research,* 95(3) Jan-Feb 2002, 151-160 Heldref Publications.

Wallen, K. (1990). Desire and ability: hormones and the regulation of female sexual behavior. *Neuroscience and Biobehavioral Reviews, 14*(2), 233-241.

Wallen, K. (1996). Nature needs nurture: the interaction of hormonal and social influences on the development of behavioral sex differences in rhesus monkeys. *Hormones and Behavior, 30*(4), 364-378.

Wallen, K. (2005). Hormonal influences on sexually differentiated behavior in nonhuman primates. *Frontiers in Neuroendocrinology, 26*(1), 7-26.

Wallen, K., & Baum, M. J. (2002). Masculinization & Defeminization in Altricial and Precocial Mammals: Comparative Aspects of Steroid Hormone Action. In *Hormones, Brain, and Behavior*: Elsevier Science.

Willingham, W. W., & Cole, N. S. (1997). *Gender and Fair Assessment*: Educational Testing Service, Princeton, NJ.

Zvoch, K., & Stevens, J. J. (2003). A Multilevel, Longitudinal Analysis of Middle School Math and Language Achievement. *Education Policy Analysis Archives. v11 n20 2003*.

Zvoch, K., & Stevens, J. J. (2006). Longitudinal Effects of School Context and Practice on Middle School Mathematics Achievement. *Journal of Educational Research, 99*(6), 347-356.

Appendix A. Taxonomy of multilevel models for change fitted to academic growth data

Model	Level 1 Model	Level 2 Model
Pre	$Y_{ti} = \pi_{0i} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$
A	$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$
B	$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + \pi_{2i}(TIME)^2_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$ $\pi_{2i} = \beta_{20} + r_{2i}$
D	$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + \pi_{2i}(TIME)^2_{ti}$ $+ \pi_{3i}(TIME)^3_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$ $\pi_{2i} = \beta_{20} + r_{2i}$ $\pi_{3i} = \beta_{30}$
F	$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME)_{ti} + \pi_{2i}(TIME)^2_{ti}$ $+ \pi_{3i}(TIME)^3_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + \beta_{01}(SEX)_i + \beta_{02}(ETHN)_i + \beta_{03}(FRL)_i + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(SEX)_i + \beta_{12}(ETHN)_i + \beta_{13}(FRL)_i + r_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(SEX)_i + \beta_{22}(ETHN)_i + \beta_{23}(FRL)_i + r_{2i}$ $\pi_{3i} = \beta_{30} + \beta_{31}(SEX)_i + \beta_{32}(ETHN)_i + \beta_{33}(FRL)_i$

Appendix A cont. Taxonomy of multilevel models for change fitted to academic growth data

Model	Level 1 Model	Level 2 Model
I	$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME})^2_{ti}$ $+ \pi_{3i}(\text{TIME})^3_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{SEX})_i + \beta_{02}(\text{ETHN})_i + \beta_{03}(\text{FRL})_i + \beta_{04}(\text{SEX x ETHN})$ $+ \beta_{05}(\text{SEX x FRL}) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(\text{SEX})_i + \beta_{12}(\text{ETHN})_i + \beta_{13}(\text{FRL})_i + \beta_{14}(\text{SEX x ETHN})$ $+ \beta_{15}(\text{SEX x FRL}) + r_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(\text{SEX})_i + \beta_{22}(\text{ETHN})_i + \beta_{23}(\text{FRL})_i + \beta_{24}(\text{SEX x ETHN})$ $+ \beta_{25}(\text{SEX x FRL}) + r_{2i}$ $\pi_{3i} = \beta_{30} + \beta_{31}(\text{SEX})_i + \beta_{32}(\text{ETHN})_i + \beta_{33}(\text{FRL})_i + \beta_{34}(\text{SEX x ETHN})$ $+ \beta_{35}(\text{SEX x FRL})$
Mu	$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME})^2_{ti}$ $+ \pi_{3i}(\text{TIME})^3_{ti} + \pi_{4i}(\text{TOP})_{ti}$ $+ \pi_{5i}(\text{BOTT})_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$ $\pi_{2i} = \beta_{20} + r_{2i}$ $\pi_{3i} = \beta_{30}$ $\pi_{4i} = \beta_{40}$ $\pi_{5i} = \beta_{50}$
M	$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME})^2_{ti}$ $+ \pi_{3i}(\text{TIME})^3_{ti} + \pi_{4i}(\text{TOP})_{ti}$ $+ \pi_{5i}(\text{BOTT})_{ti} + e_{ti}$	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{SEX})_i + \beta_{02}(\text{ETHN})_i + \beta_{03}(\text{FRL})_i + \beta_{04}(\text{SEX x ETHN})$ $+ \beta_{05}(\text{SEX x FRL}) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(\text{SEX})_i + \beta_{12}(\text{ETHN})_i + \beta_{13}(\text{FRL})_i + \beta_{14}(\text{SEX x ETHN})$ $+ \beta_{15}(\text{SEX x FRL}) + r_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(\text{SEX})_i + \beta_{22}(\text{ETHN})_i + \beta_{23}(\text{FRL})_i + \beta_{24}(\text{SEX x ETHN})$ $+ \beta_{25}(\text{SEX x FRL}) + r_{2i}$ $\pi_{3i} = \beta_{30} + \beta_{31}(\text{SEX})_i + \beta_{32}(\text{ETHN})_i + \beta_{33}(\text{FRL})_i + \beta_{34}(\text{SEX x ETHN})$ $+ \beta_{35}(\text{SEX x FRL})$ $\pi_{4i} = \beta_{40} + \beta_{41}(\text{SEX})_i$ $\pi_{5i} = \beta_{50} + \beta_{51}(\text{SEX})_l$

Appendix B. Model Taxonomy Results.

Table B1.

Model Taxonomy Results for Language Arts

		Model	**Pre**	**A**	**B**	**D**	**F**	**I**	**Mu**	**M**
Fixed Effects,		Parameter								
Initial status,	**Intercept**	β_{00}	**230.73**	**190.00**	**192.99**	**192.81**	**207.20**	**206.88**	**193.99**	**207.42**
π_{0i}			2.70	2.03	1.94	1.94	2.79	3.21	1.86	3.03
	SEX	β_{01}					-6.90	-6.61		-7.52
							3.34	4.81		4.43
	ETHN	β_{02}					**-20.48**	**-22.57**		**-20.61**
							5.63	6.80		6.93
	FRL	β_{03}					-4.40	-1.89		-2.00
							5.76	7.14		7.20
	SEXETH	β_{04}						2.31		2.21
								10.07		9.62
	SEXFRL	β_{05}						-3.39		-2.22
								10.31		9.92
Linear rate of										
change, π_{1i}	**Intercept**	β_{10}		**7.62**	**8.58**	**9.58**	**13.32**	**14.23**	**9.44**	**13.97**
				0.24	0.55	1.06	1.96	2.29	1.10	2.36
	SEX	β_{11}					-3.13	-4.88		-4.89
							2.09	2.89		2.95
	ETHN	β_{12}					-4.49	-6.70		-6.61
							3.60	5.90		5.71
	FRL	β_{13}					-0.56	-0.54		-0.85
							3.68	6.07		5.94
	SEXETH	β_{14}						2.25		2.76
								7.06		6.79
	SEXFRL	β_{15}						2.15		2.57
								7.21		6.96
Quadratic rate										
of change, π_{2i}	Intercept	β_{20}			-0.11	-0.42	-0.80	-1.09	-0.38	-1.04
					0.06	0.31	0.58	0.67	0.32	0.69
	SEX	β_{21}					0.42	0.98		0.97
							0.62	0.83		0.84
	ETHN	β_{22}					0.78	0.80		0.75
							1.27	2.01		1.96
	FRL	β_{23}					-0.39	0.36		0.50
							1.29	2.07		2.03
	SEXETH	β_{24}						0.67		0.63
								2.39		2.31
	SEXFRL	β_{25}						-2.21		-2.32
								2.44		2.36
Cubic rate of										
change, π_{3i}	Intercept	β_{30}				0.02	0.04	0.06	0.02	0.06
						0.02	0.04	0.05	0.02	0.05
	SEX	β_{31}					-0.02	-0.06		-0.06
							0.05	0.06		0.06
	ETHN	β_{32}					-0.06	-0.04		-0.04
							0.10	0.15		0.15
	FRL	β_{33}					0.04	-0.03		-0.04
							0.10	0.16		0.15
	SEXETH	β_{34}						-0.08		-0.09
								0.19		0.18
	SEXFRL	β_{35}						0.20		0.21
								0.19		0.18

Table B1 (continued). *Language Arts Results*.

		Model Parameter	**Pre**	**A**	**B**	**D**	**F**	**I**	**Mu**	**M**
Effect of										
TOP status,	Intercept	β_{40}							2.62	-0.15
π_{4i}									1.60	2.15
	SEX	β_{41}								5.67
										3.10
Effect of										
BOTTOM	**Intercept**	β_{50}							**-7.36**	**-6.58**
status, π_{5i}									1.64	2.38
	SEX	β_{51}								-0.27
										3.55
Variance Components										
Level 1	Within	σ^2	806.90	117.71	99.72	98.84	99.45	98.57	102.63	99.51
	person		28.41	10.85	9.99	9.94	9.97	9.93	10.13	9.98
Level 2	In initial	τ_{00}	**962.94**	**574.06**	**468.01**	**468.56**	**314.76**	**309.34**	**378.83**	**252.15**
	status		31.03	23.96	21.63	21.65	17.74	17.59	19.46	15.88
	In linear	τ_{11}		**4.95**	**16.50**	**16.89**	**11.60**	**11.43**	**15.94**	**12.71**
	change			2.22	4.06	4.11	3.41	3.38	3.99	3.56
	In quadratic	τ_{22}			**0.13**	**0.14**	**0.12**	**0.13**	**0.14**	**0.13**
	change				0.36	0.37	0.35	0.36	0.37	0.37
	covariances	τ_{01}		0.77	0.94	0.92	0.94	0.93	0.96	0.92
	(as correl)	τ_{02}			-0.64	-0.62	-0.68	-0.65	-0.64	-0.67
		τ_{12}			-0.82	-0.83	-0.87	-0.87	-0.83	-0.89
			ρ = ICC = 0.544 54%							
Pseudo R2 Statistics and Goodness-of-Fit										
		$R^2_{y,\hat{y}}$		0.31	0.31	0.31	0.56	0.56		**0.57**
		Deviance	6260.34	5393.37	5361.77	5360.71	5274.48	5267.70	5346.08	**5251.71**
		AIC	6266.34	5405.37	5381.77	5382.71	**5320.48**	5329.70	5372.08	**5321.71**
		BIC	6275.64	5423.97	5412.77	5416.81	**5391.78**	5425.80	5412.38	5430.21

Table B2.

Model Taxonomy Results for Math

		Model	**Pre**	**A**	**B**	**D**	**F**	**I**	**Mu**	**M**
Fixed Effects,		Parameter								
Initial status, π_{0i}	**Intercept**	β_{00}	**230.64**	**191.26**	**192.08**	**192.75**	**202.02**	**200.72**	**193.11**	**200.01**
			2.13	1.58	1.62	1.63	2.23	2.53	1.48	2.32
	SEX	β_{01}					2.04	4.37		3.62
							2.72	3.58		3.31
	ETHN	β_{02}					**-16.87**	-13.87		-11.98
							5.17	7.07		6.25
	FRL	β_{03}					-7.42	-7.61		-5.64
							5.34	7.41		6.44
	SEXETH	β_{04}						-7.01		-4.25
								9.71		8.61
	SEXFRL	β_{05}						1.19		-2.52
								10.03		8.75
Linear rate of change, π_{1i}	**Intercept**	β_{10}		**8.14**	**7.47**	**4.01**	**5.18**	**5.14**	**3.83**	**4.78**
				0.21	0.43	0.98	1.58	1.78	1.02	1.93
	SEX	β_{11}					-0.34	-0.30		-0.86
							1.95	2.65		2.78
	ETHN	β_{12}					-1.09	1.59		3.87
							3.07	3.04		3.08
	FRL	β_{13}					-0.93	-3.70		-5.60
							3.07	3.12		3.24
	SEXETH	β_{14}						-6.82		-10.10
								6.30		5.81
	SEXFRL	β_{15}						7.00		11.60
								6.32		5.88
Quadratic rate of change, π_{2i}	**Intercept**	β_{20}			0.08	**1.16**	**1.28**	**1.34**	**1.19**	**1.37**
					0.04	0.29	0.47	0.53	0.30	0.56
	SEX	β_{21}					0.08	-0.02		0.14
							0.58	0.78		0.81
	ETHN	β_{22}					-0.86	**-1.73**		**-2.42**
							0.95	0.85		0.84
	FRL	β_{23}					0.46	1.22		**1.83**
							0.96	0.90		0.89
	SEXETH	β_{24}						2.29		3.29
								1.94		1.79
	SEXFRL	β_{25}						-2.07		-3.33
								1.95		1.80
Cubic rate of change, π_{3i}	**Intercept**	β_{30}				**-0.08**	**-0.09**	**-0.10**	**-0.08**	**-0.10**
						0.02	0.03	0.04	0.02	0.04
	SEX	β_{31}					-0.01	0.00		-0.01
							0.04	0.06		0.06
	ETHN	β_{32}					0.08	**0.14**		**0.19**
							0.07	0.07		0.06
	FRL	β_{33}					-0.05	-0.09		**-0.14**
							0.07	0.07		0.07
	SEXETH	β_{34}						-0.15		-0.23
								0.14		0.13
	SEXFRL	β_{35}						0.12		0.21
								0.15		0.14

Table B2. (continued) *Model Taxonomy Results for Math*

		Model Parameter	**Pre**	**A**	**B**	**D**	**F**	**I**	**Mu**	**M**
Effect of TOP status, π_{4i}	**Intercept**	β_{40}							**7.00**	**6.37**
									1.59	2.34
	SEX	β_{41}								1.48
										3.09
Effect of BOTTOM status, π_{5i}	**Intercept**	β_{50}							**-5.64**	**-6.60**
									1.57	2.31
	SEX	β_{51}								2.53
										3.15
Variance Components										
Level 1	Within person	σ^2	865.06	87.20	85.13	81.11	79.93	78.86	86.30	82.63
			29.41	9.34	9.23	9.01	8.94	8.88	9.29	9.09
Level 2	In initial status	τ_{00}	**497.97**	**312.64**	**307.79**	**310.05**	**183.73**	**182.86**	**209.77**	**116.84**
			22.32	17.68	17.54	17.61	13.55	13.52	14.48	10.81
	In linear change	τ_{11}		3.92	3.37	3.73	1.13	1.23	3.40	1.65
				1.98	1.84	1.93	1.06	1.11	1.84	1.28
	In quadratic change	τ_{22}			0.01	0.01	0.02	0.02	0.03	0.03
					0.10	0.11	0.14	0.14	0.16	0.17
	covariances	τ_{01}			0.92	0.90	0.73	0.74	0.91	0.74
	(as correl)	τ_{02}			-0.13	-0.20	0.00	-0.02	0.01	0.02
		τ_{12}			-0.05	-0.17	-0.38	-0.44	-0.35	-0.65
			ρ = ICC = 0.365 37%							
Pseudo R2 Statistics and Goodness-of-Fit										
	$R^2_{y,\hat{y}}$			0.43	0.43	0.43	0.71	**0.71**		0.70
	Deviance		6204.01	5164.11	5160.19	5144.93	5016.12	5008.40	5117.92	**4980.42**
	AIC		6210.01	5176.11	5180.19	5166.93	5062.12	5070.40	5143.92	**5050.42**
	BIC		6219.31	5194.71	5211.19	5201.03	**5133.42**	5166.49	5184.22	5158.92

Table B3.

Model Taxonomy Results for Science

Fixed Effects,		Model Parameter	**Pre**	**A**	**B**	**D**	**F**	**I**	**Mu**	**M**
Initial status, π_{0i}	**Intercept**	β_{00}	**239.52**	**199.31**	**197.65**	**197.51**	**216.64**	**214.05**	**200.46**	**213.44**
			2.80	2.48	2.40	2.40	3.36	3.90	2.44	3.90
	SEX	β_{01}					-4.78	0.02		1.15
							3.84	5.64		5.75
	ETHN	β_{02}					**-32.51**	**-21.96**		**-17.58**
							5.08	6.17		6.53
	FRL	β_{03}					-5.94	-11.62		-10.81
							5.18	6.34		6.57
	SEXETH	β_{04}						**-20.49**		**-20.90**
								8.18		8.11
	SEXFRL	β_{05}						10.52		11.51
								8.14		7.89
Linear rate of change, π_{1i}	**Intercept**	β_{10}		**8.33**	**9.79**	**10.56**	**9.57**	**9.04**	**10.43**	**8.92**
				0.28	0.69	1.57	2.81	3.30	1.64	3.44
	SEX	β_{11}					5.99	7.11		6.88
							3.11	4.45		4.55
	ETHN	β_{12}					-5.95	-8.69		-8.07
							5.23	7.24		7.10
	FRL	β_{13}					1.59	6.05		6.44
							5.31	7.14		7.03
	SEXETH	β_{14}						3.25		2.03
								10.88		10.64
	SEXFRL	β_{15}						-6.74		-6.89
								11.07		10.89
Quadratic rate of change, π_{2i}	Intercept	β_{20}			-0.16	-0.41	0.71	1.07	-0.44	0.98
					0.07	0.47	0.80	0.93	0.49	0.95
	SEX	β_{21}					**-2.29**	**-3.04**		**-2.89**
							0.93	1.34		1.35
	ETHN	β_{22}					0.54	-0.05		-0.16
							1.64	2.01		1.95
	FRL	β_{23}					-0.55	-0.86		-0.90
							1.67	1.97		1.92
	SEXETH	β_{24}						2.00		2.11
								3.59		3.50
	SEXFRL	β_{25}						-0.13		-0.03
								3.67		3.60
Cubic rate of change, π_{3i}	Intercept	β_{30}				0.02	-0.08	-0.12	0.02	-0.11
						0.04	0.06	0.07	0.04	0.07
	SEX	β_{31}					**0.18**	**0.26**		**0.25**
							0.07	0.10		0.10
	ETHN	β_{32}					-0.01	0.10		0.10
							0.12	0.14		0.14
	FRL	β_{33}					0.03	0.03		0.03
							0.13	0.14		0.13
	SEXETH	β_{34}						-0.27		-0.27
								0.27		0.26
	SEXFRL	β_{35}						0.07		0.06
								0.28		0.27

Table B3. (continued) *Model Taxonomy Results for Science*

		Model Parameter	Pre	A	B	D	F	I	Mu	M
Effect of TOP status, π_{4i}	**Intercept**	β_{40}							**4.74**	**7.45**
									2.47	2.96
	SEX	β_{41}								-5.97
										4.51
Effect of BOTTOM status, π_{5i}	**Intercept**	β_{50}							**-13.92**	**-11.36**
									2.65	3.83
	SEX	β_{51}								0.95
										4.83
Variance Components										
Level 1	Within person	σ^2	1059.78	262.06	241.20	241.03	234.70	226.97	251.85	233.75
			32.55	16.19	15.53	15.53	15.32	15.07	15.87	15.29
Level 2	**In initial status**	τ_{00}	**975.88**	**730.15**	**628.74**	**628.46**	**297.41**	**292.25**	**415.20**	**209.77**
			31.24	27.02	25.07	25.07	17.25	17.10	20.38	14.48
	In linear change	τ_{11}		3.30	10.30	10.27	2.92	2.86	10.42	3.15
				1.82	3.21	3.20	1.71	1.69	3.23	1.77
	In quadratic change	τ_{22}			0.11	0.11	0.08	0.06	0.13	0.08
					0.33	0.33	0.29	0.25	0.35	0.28
	covariances (as correl)	τ_{01}		0.67	0.96	0.95	0.77	0.92	0.94	0.88
		τ_{02}			-0.62	-0.62	-0.24	-0.46	-0.58	-0.40
		τ_{12}			-0.79	-0.79	-0.76	-0.67	-0.80	-0.74
			ρ = ICC = 0.479 48%							
Pseudo R2 Statistics and Goodness-of-Fit										
	$R^2_{y,\hat{y}}$			0.30	0.30	0.30	0.63	**0.64**		0.63
	Deviance		6386.00	5747.99	5733.44	5733.18	5593.73	5573.32	5709.84	**5554.95**
	AIC		6392.00	5759.99	5753.44	5755.18	5639.73	5635.32	5735.84	**5624.95**
	BIC		6401.30	5778.59	5784.44	5789.28	**5711.03**	5731.42	5776.14	5733.45

Appendix C.

Model Evaluation

The fully conditional Model M was evaluated according to the protocols suggested by Singer and Willett (2003). Assumptions of normally distributed residuals and homoscedasticity of residual variance were evaluated and Model M's fit was compared to prior models. Figures exhibiting the results of assumption checks are presented in Appendix C.

Normality

Normal probability plots were created by plotting residual values against their normal scores for residuals at level 1 and level 2. In language arts, level 1 residuals depart slightly from the normalized scores at the upper end of the distribution. In math, level 1 residuals depart slightly at the upper and lower extremes. In science, level 1 residuals depart from normal distribution at the extreme low end. In all three domains, the level 1 residuals exhibit an approximately normal distribution. Similarly, normal probability plots of level 2 residuals for initial status and linear and quadratic growth exhibit approximately normal distributions, with departures limited to upper and lower extremes.

Further checks of normal distribution of residuals were made by plotting the standardized residuals against the student-level variable ID. The raw residuals at levels 1 and 2 fall largely within ±2 standard deviations of their center with no apparent systematic pattern of deviation, supporting the assumption of normality for all three domains.

Homoscedasticity

The assumption of homoscedasticity of residual variance may be checked by plotting raw residuals against predictors at the same level of the model. For all three domains, residual variability appears equal by gender, supporting the assumption of homoscedasticity. Residual variability by ethnicity and by income status indicates some restriction of variability for African American ethnicity and low income status, in comparison to Caucasian ethnicity and non-low income status, most notably in science.

In summary for checks of normality and homoscedasticity, though minor deviations from normality and homoscedasticity occur across the three domains, the overall assumptions are met.

Table 1 A

Results of Fitting Multilevel Model M for Change to the Language Arts Data (n=630)

		Parameter	Estimate	ase	z
Fixed Effects					
Initial status, π_0	Intercept	β_{00}	207.42 ***	3.03	68.39
	SEX	β_{01}	-7.52	4.43	-1.70
	ETHN	$\boldsymbol{\beta_{02}}$	**-20.61 ****	**6.93**	**-2.97**
	FRL	β_{03}	-2.00	7.20	-0.28
	SEXETH	β_{04}	2.21	9.62	0.23
	SEXFRL	β_{05}	-2.22	9.92	-0.22
Linear rate of change, π_{1i}	Intercept	β_{10}	13.97 ***	2.36	5.93
	SEX	β_{11}	-4.89	2.95	-1.65
	ETHN	β_{12}	-6.61	5.71	-1.16
	FRL	β_{13}	-0.85	5.94	-0.14
	SEXETH	β_{14}	2.76	6.79	0.41
	SEXFRL	β_{15}	2.57	6.96	0.37
Quadratic rate of change, π_{2i}	Intercept	β_{20}	-1.04	0.69	-1.50
	SEX	β_{21}	0.97	0.84	1.15
	ETHN	β_{22}	0.75	1.96	0.38
	FRL	β_{23}	0.50	2.03	0.25
	SEXETH	β_{24}	0.63	2.31	0.27
	SEXFRL	β_{25}	-2.32	2.36	-0.98

Table 1A, cont.

		Parameter	Estimate		ase	z
Cubic rate of change, π_{3i}	Intercept	β_{30}	0.06		0.05	1.18
	SEX	β_{31}	-0.06		0.06	-1.00
	ETHN	β_{32}	-0.04		0.15	-0.26
	FRL	β_{33}	-0.04		0.15	-0.29
	SEXETH	β_{34}	-0.09		0.18	-0.48
	SEXFRL	β_{35}	0.21		0.18	1.14
Effect of TOP status, π_{4i}	Intercept	β_{40}	-0.15		2.15	-0.07
	SEX	β_{41}	5.67		3.10	1.83
Effect of BOTTOM status, π_{5i}	Intercept	β_{50}	-6.58	**	2.38	-2.76
	SEX	β_{51}	-0.27		3.55	-0.08
Variance Components						
Level 1	Within person	σ^2	99.51		9.98	
Level 2	In initial status	τ_{00}	252.15	***	15.88	
	In linear change	τ_{11}	12.71	**	3.56	
	In quadratic change	τ_{22}	0.13	**	0.37	
	Covariances	τ_{01}	0.92			
		τ_{02}	-0.67			
		τ_{12}	-0.89			

**p* < .05; **p < .01; ***p < .001.

Table 1B

Results of Fitting Multilevel Model M for Change to the Mathematics Data (n=630)

		Parameter	Estimate		ase	z
Fixed Effects						
Initial status, π_{0i}	Intercept	β_{00}	200.01	***	2.32	86.33
	SEX	β_{01}	3.62		3.31	1.09
	ETHN	β_{02}	-11.98		6.25	-1.92
	FRL	β_{03}	-5.64		6.44	-0.88
	SEXETH	β_{04}	-4.25		8.61	-0.49
	SEXFRL	β_{05}	-2.52		8.75	-0.29
Linear rate of change, π_{1i}	Intercept	β_{10}	4.78	**	1.93	2.48
	SEX	β_{11}	-0.86		2.78	-0.31
	ETHN	β_{12}	3.87		3.08	1.26
	FRL	β_{13}	-5.60		3.24	-1.73
	SEXETH	β_{14}	-10.10		5.81	-1.74
	SEXFRL	β_{15}	11.60		5.88	1.97
Quadratic rate of change, π_{2i}	Intercept	β_{20}	1.37	**	0.56	2.45
	SEX	β_{21}	0.14		0.81	0.17
	ETHN	$\boldsymbol{\beta_{22}}$	**-2.42**	**	**0.84**	**-2.89**
	FRL	$\boldsymbol{\beta_{23}}$	**1.83**	**	**0.89**	**2.04**
	SEXETH	β_{24}	3.29		1.79	1.83
	SEXFRL	β_{25}	-3.33		1.80	-1.85

Table 1B, cont.

		Parameter	Estimate		ase	z
Cubic rate of change, π_{3i}	Intercept	β_{30}	-0.10	**	0.04	-2.45
	SEX	β_{31}	-0.01		0.06	-0.12
	ETHN	$\boldsymbol{\beta_{32}}$	**0.19**	**	**0.06**	**2.98**
	FRL	$\boldsymbol{\beta_{33}}$	**-0.14**	**	**0.07**	**-1.99**
	SEXETH	β_{34}	-0.23		0.13	-1.72
	SEXFRL	β_{35}	0.21		0.14	1.55
Effect of TOP status, π_{4i}	Intercept	β_{40}	6.37	**	2.34	2.72
	SEX	β_{41}	1.48		3.09	0.48
Effect of BOTTOM status, π_{5i}	Intercept	β_{50}	-6.60	**	2.31	-2.86
	SEX	β_{51}	2.53		3.15	0.80
Variance Components						
Level 1	Within person	σ^2	82.63		9.09	
Level 2	In initial status	τ_{00}	116.84	***	10.81	
	In linear change	τ_{11}	1.65		1.28	
	In quadratic change	τ_{22}	0.03		0.17	
	Covariances	τ_{01}	0.74			
		τ_{02}	0.02			
		τ_{12}	-0.65			

*p < .05; **p < .01; ***p < .001.

Table 1C

Results of Fitting Multilevel Model M for Change to the Science Data (n=630)

		Parameter	Estimate		ase	z
Fixed Effects						
Initial status, π_0	Intercept	β_{00}	213.44	***	3.90	54.78
	SEX	β_{01}	1.15		5.75	0.20
	ETHN	$\boldsymbol{\beta_{02}}$	**-17.58**	**	**6.53**	**-2.69**
	FRL	β_{03}	-10.81		6.57	-1.65
	SEXETH	$\boldsymbol{\beta_{04}}$	**-20.90**	**	**8.11**	**-2.58**
	SEXFRL	β_{05}	11.51		7.89	1.46
Linear rate of change, π_{1i}	Intercept	β_{10}	8.92	**	3.44	2.59
	SEX	β_{11}	6.88		4.55	1.51
	ETHN	β_{12}	-8.07		7.10	-1.14
	FRL	β_{13}	6.44		7.03	0.92
	SEXETH	β_{14}	2.03		10.64	0.19
	SEXFRL	β_{15}	-6.89		10.89	-0.63
Quadratic rate of change, π_{2i}	Intercept	β_{20}	0.98		0.95	1.03
	SEX	$\boldsymbol{\beta_{21}}$	**-2.89**	**	**1.35**	**-2.14**
	ETHN	β_{22}	-0.16		1.95	-0.08
	FRL	β_{23}	-0.90		1.92	-0.47
	SEXETH	β_{24}	2.11		3.50	0.60
	SEXFRL	β_{25}	-0.03		3.60	-0.01

Table 1C, cont.

	Parameter				ase	z
Cubic rate of change, π_{3i}	Intercept	β_{30}	-0.11		0.07	-1.63
	SEX	$\boldsymbol{\beta_{31}}$	**0.25**	**	**0.10**	**2.49**
	ETHN	β_{32}	0.10		0.14	0.76
	FRL	β_{33}	0.03		0.13	0.21
	SEXETH	β_{34}	-0.27		0.26	-1.03
	SEXFRL	β_{35}	0.06		0.27	0.24
Effect of TOP status, π_{4i}	Intercept	β_{40}	7.45	**	2.96	2.52
	SEX	β_{41}	-5.97		4.51	-1.32
Effect of BOTTOM status, π_{5i}	Intercept	β_{50}	-11.36	**	3.83	-2.96
	SEX	β_{51}	0.95		4.83	0.20
Variance Components						
Level 1	Within person	σ^2	233.75		15.29	
Level 2	In initial status	τ_{00}	209.77	***	14.48	
	In linear change	τ_{11}	3.15		1.77	
	In quadratic change	τ_{22}	0.08		0.28	
	Covariances	τ_{01}	0.88			
		τ_{02}	-0.40			
		τ_{12}	-0.74			

*p < .05; **p < .01; ***p < .001.

Table 2

Effect Sizes, Variance Ratios, Coefficient of Variance Ratios, and Odds Ratios by Gender and Grade for Three Academic Domains

Language Arts

Grade	Girls					Boys					Comparison				
	σ^2	*SD*	*M*	*CV*	*n*	σ^2	*SD*	*M*	*CV*	*n*	***d***	***VR***	**CV ratio**	**OR_{top}**	**OR_{bot}**
3	508.19	22.54	195.95	11.50	58	603.43	24.56	192.96	12.73	57	**-0.13**	**1.19**	**1.22**	**0.75**	**1.61**
5	1224.96	35.00	225.43	15.53	63	951.70	30.85	215.76	14.30	67	**-0.29**	**0.78**	**0.85**	**0.53**	**2.24**
6	1208.89	34.77	242.72	14.32	65	1146.13	33.85	227.53	14.88	64	**-0.43**	**0.95**	**1.08**	**0.38**	**3.34**
7	1555.53	39.44	256.60	15.37	65	1414.38	37.61	239.47	15.70	62	**-0.43**	**0.91**	**1.04**	**0.43**	**2.20**
8	1563.39	39.54	273.19	14.47	64	1474.12	38.39	253.81	15.13	57	**-0.48**	**0.94**	**1.09**	**0.23**	**3.46**
3 vs 8				1.58					1.41						

Mathematics

Grade	Girls					Boys					Comparison				
	σ^2	*SD*	*M*	*CV*	*n*	σ^2	*SD*	*M*	*CV*	*n*	***d***	***VR***	**CV ratio**	**OR_{top}**	**OR_{bot}**
3	351.83	18.76	192.78	9.73	58	396.54	19.91	198.44	10.03	57	**0.29**	**1.13**	**1.06**	**2.54**	**0.82**
5	583.61	24.16	213.67	11.31	63	535.39	23.14	218.37	10.60	67	**0.20**	**0.92**	**0.88**	**2.12**	**1.81**
6	743.40	27.27	233.14	11.69	65	755.56	27.49	234.67	11.71	64	**0.06**	**1.02**	**1.00**	**1.06**	**0.95**
7	934.75	30.57	250.86	12.19	65	1053.27	32.45	249.47	13.01	62	**-0.04**	**1.13**	**1.14**	**1.06**	**0.94**
8	1049.48	32.40	267.63	12.10	64	1302.46	36.09	262.84	13.73	57	**-0.14**	**1.24**	**1.29**	**0.94**	**0.63**
3 vs 8				1.55					1.87						

Science

Grade	Girls					Boys					Comparison				
	σ^2	*SD*	*M*	*CV*	*n*	σ^2	*SD*	*M*	*CV*	*n*	***d***	***VR***	**CV ratio**	**OR_{top}**	**OR_{bot}**
3	783.97	28.00	201.22	13.91	58	943.75	30.72	200.82	15.30	57	**-0.01**	**1.20**	**1.21**	**0.82**	**1.90**
5	1196.03	34.58	228.94	15.11	63	1156.74	34.01	232.49	14.63	67	**0.10**	**0.97**	**0.94**	**1.36**	**0.84**
6	1722.13	41.50	249.11	16.66	65	1888.76	43.46	235.00	18.49	64	**-0.33**	**1.10**	**1.23**	**0.60**	**2.03**
7	1318.44	36.31	265.51	13.68	65	1588.16	39.85	253.66	15.71	62	**-0.31**	**1.20**	**1.32**	**0.58**	**1.63**
8	1308.25	36.17	279.00	12.96	64	2351.09	48.49	268.95	18.03	57	**-0.24**	**1.80**	**1.93**	**1.03**	**1.76**
3 vs 8				0.87					1.39						

Note. *d* =(boys'mean - girls' mean)/pooled SD; positive *d* reflects higher mean for boys. VR = boys' variance/ girls' variance.
CV= SD/M *100; CV ratio = CV2 boys/ CV2 girls; Consider CV ratio as F ratio with df (1, 2*(HM-1))
OR_t = ratio of boys' odds/girls' odds of scoring in the top quintile; OR_b = ratio of boys' odds/girls' odds of scoring in the bottom quintile

Table 3

Language Arts Effect Sizes and Student Ratios by Quintile

Top Quintile

	Grade	3	5	6	7	8
	SD	11.09	13.53	14.07	12.11	13.29
Ethnicity						
Cauc	*M*	227.83	270.36	283.75	301.43	316.81
AfAm	*M*	216.00	256.00	275.00	297.00	308.00
Cauc	*n*	24	25	24	23	21
AfAm	*n*	1	1	4	2	3
	d	**-1.07**	**-1.06**	**-0.62**	**-0.37**	**-0.66**
	NR	**-0.92**	**-0.92**	**-0.71**	**-0.84**	**-0.75**

Mid Quintile

	Grade	3	5	6	7	8
	SD	4.07	5.32	6.77	7.78	7.33
Cauc	*M*	193.18	218.00	233.94	249.75	270.53
AfAm	*M*	191.00	216.50	233.33	251.40	268.38
Cauc	*n*	17	14	17	16	17
AfAm	*n*	9	12	9	10	8
	d	**-0.54**	**-0.28**	**-0.09**	**0.21**	**-0.29**
	NR	**-0.31**	**-0.08**	**-0.31**	**-0.23**	**-0.36**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	5.63	9.10	12.82	11.63	13.45
Cauc	*M*	162.20	181.43	189.40	197.50	211.50
AfAm	*M*	162.35	176.00	184.52	191.14	204.90
Cauc	*n*	5	7	5	4	4
AfAm	*n*	17	20	21	21	20
	d	**0.03**	**-0.60**	**-0.38**	**-0.55**	**-0.49**
	NR	**0.55**	**0.48**	**0.62**	**0.68**	**0.67**

Table 3, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	11.09	13.53	14.07	12.11	13.29
Income						
not low	*M*	227.70	270.65	283.44	301.33	316.81
low	*M*	223.50	263.33	274.67	299.75	308.00
not low	*n*	23	23	25	21	21
low	*n*	2	3	3	4	3
	d	**-0.38**	**-0.54**	**-0.62**	**-0.13**	**-0.66**
	NR	**-0.84**	**-0.77**	**-0.79**	**-0.68**	**-0.75**

Mid Quintile

	Grade	3	5	6	7	8
	SD	4.07	5.32	6.77	7.78	7.33
not low	*M*	192.90	217.69	234.52	250.79	270.50
low	*M*	190.83	216.70	230.40	249.29	267.71
not low	*n*	20	16	21	19	18
low	*n*	6	10	5	7	7
	d	**-0.51**	**-0.19**	**-0.61**	**-0.19**	**-0.38**
	NR	**-0.54**	**-0.23**	**-0.62**	**-0.46**	**-0.44**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	5.63	9.10	12.82	11.63	13.45
not low	*M*	160.50	180.83	189.60	193.83	213.17
low	*M*	162.72	176.43	184.48	191.63	203.61
not low	*n*	4	6	5	6	6
low	*n*	18	21	21	19	18
	d	**0.39**	**-0.48**	**-0.40**	**-0.19**	**-0.71**
	NR	**0.64**	**0.56**	**0.62**	**0.52**	**0.50**

Table 3, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	11.09	13.53	14.07	12.11	13.29
Gender						
Girls	*M*	225.50	272.31	283.89	302.35	316.11
Boys	*M*	229.73	265.80	279.56	298.38	314.20
Girls	*n*	14	16	19	17	19
Boys	*n*	11	10	9	8	5
	d	**0.38**	**-0.48**	**-0.31**	**-0.33**	**-0.14**
	NR	**-0.12**	**-0.23**	**-0.36**	**-0.36**	**-0.58**

Mid Quintile

	Grade	3	5	6	7	8
	SD	4.07	5.32	6.77	7.78	7.33
Girls	*M*	192.23	215.71	234.17	249.33	268.62
Boys	*M*	192.62	219.17	233.36	250.94	270.92
Girls	*n*	13	14	12	9	13
Boys	*n*	13	12	14	17	12
	d	**0.10**	**0.65**	**-0.12**	**0.21**	**0.31**
	NR	**0.00**	**-0.08**	**0.08**	**0.31**	**-0.04**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	5.63	9.10	12.82	11.63	13.45
Girls	*M*	163.11	173.11	184.57	192.22	203.29
Boys	*M*	161.77	179.56	185.79	192.13	207.12
Girls	*n*	9	9	7	9	7
Boys	*n*	13	18	19	16	17
	d	**-0.24**	**0.71**	**0.10**	**-0.01**	**0.29**
	NR	**0.18**	**0.33**	**0.46**	**0.28**	**0.42**

Table 4

Mathematics Effect Sizes and Student Ratios by Quintile

Top Quintile

	Grade	3	5	6	7	8
	SD	5.86	12.69	10.84	9.31	7.94
Ethnicity						
Cauc	*M*	222.55	250.21	272.33	293.40	307.52
AfAm	*M*				284.00	314.00
Cauc	*n*	22	24	27	25	23
AfAm	*n*				1	1
	d				**-1.01**	**0.82**
	NR				**-0.92**	**-0.92**

Mid Quintile

	Grade	3	5	6	7	8
	SD	3.04	5.03	6.28	7.21	8.02
Cauc	*M*	197.25	217.78	233.43	251.44	270.53
AfAm	*M*	195.60	213.11	229.29	245.29	266.67
Cauc	*n*	12	18	21	18	17
AfAm	*n*	10	9	7	7	6
	d	**-0.54**	**-0.93**	**-0.66**	**-0.85**	**-0.48**
	NR	**-0.09**	**-0.33**	**-0.50**	**-0.44**	**-0.48**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	8.44	9.28	10.90	7.81	10.96
Cauc	*M*	167.40	188.50	201.00	212.00	201.00
AfAm	*M*	166.78	182.26	196.67	207.76	213.82
Cauc	*n*	5	2	3	1	1
AfAm	*n*	18	23	24	25	22
	d	**-0.07**	**-0.67**	**-0.40**	**-0.54**	**1.17**
	NR	**0.57**	**0.84**	**0.78**	**0.92**	**0.91**

Table 4, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	5.86	12.69	10.84	9.31	7.94
Income						
not low	*M*	222.80	251.43	273.04	293.67	307.52
low	*M*	220.00	241.67	266.67	285.50	314.00
not low	*n*	20	21	24	24	23
low	*n*	2	3	3	2	1
	d	**-0.48**	**-0.77**	**-0.59**	**-0.88**	**0.82**
	NR	**-0.82**	**-0.75**	**-0.78**	**-0.85**	**-0.92**

Mid Quintile

	Grade	3	5	6	7	8
	SD	3.04	5.03	6.28	7.21	8.02
not low	*M*	197.23	218.28	232.79	250.84	270.35
low	*M*	195.44	212.11	230.00	246.17	267.17
not low	*n*	13	18	24	19	17
low	*n*	9	9	4	6	6
	d	**-0.59**	**-1.23**	**-0.44**	**-0.65**	**-0.40**
	NR	**-0.18**	**-0.33**	**-0.71**	**-0.52**	**-0.48**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	8.44	9.28	10.90	7.81	10.96
not low	*M*	171.00	192.00	201.50	208.40	215.33
low	*M*	165.78	181.96	196.80	207.81	212.95
not low	*n*	5	2	2	5	3
low	*n*	18	23	25	21	20
	d	**-0.62**	**-1.08**	**-0.43**	**-0.08**	**-0.22**
	NR	**0.57**	**0.84**	**0.85**	**0.62**	**0.74**

Table 4, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	5.86	12.69	10.84	9.31	7.94
Gender						
Girls	*M*	222.00	250.75	271.62	291.38	306.38
Boys	*M*	222.80	249.94	273.00	294.69	309.45
Girls	*n*	7	8	13	13	13
Boys	*n*	15	16	14	13	11
	d	**0.14**	**-0.06**	**0.13**	**0.36**	**0.39**
	NR	**0.36**	**0.33**	**0.04**	**0.00**	**-0.08**

Mid Quintile

	Grade	3	5	6	7	8
	SD	3.04	5.03	6.28	7.21	8.02
Girls	*M*	195.29	216.69	231.27	250.50	270.92
Boys	*M*	198.63	215.55	233.69	248.73	268.00
Girls	*n*	14	16	15	14	12
Boys	*n*	8	11	13	11	11
	d	**1.10**	**-0.23**	**0.39**	**-0.25**	**-0.36**
	NR	**-0.27**	**-0.19**	**-0.07**	**-0.12**	**-0.04**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	8.44	9.28	10.90	7.81	10.96
Girls	*M*	167.86	180.67	196.00	208.15	214.60
Boys	*M*	165.44	185.90	198.21	207.69	212.23
Girls	*n*	14	15	13	13	10
Boys	*n*	9	10	14	13	13
	d	**-0.29**	**0.56**	**0.20**	**-0.06**	**-0.22**
	NR	**-0.22**	**-0.20**	**0.04**	**0.00**	**0.13**

Table 5

Science Effect Sizes and Student Ratios by Quintile

Top Quintile

	Grade	3	5	6	7	8
	SD	7.81	5.47	7.14	6.92	6.37
Ethnicity						
Cauc	*M*	242.00	277.19	297.25	312.87	328.95
AfAm	*M*	234.00			322.00	328.00
Cauc	*n*	25	26	28	23	19
AfAm	*n*	1			1	2
	d	**-1.02**			**1.32**	**-0.15**
	NR	**-0.92**			**-0.92**	**-0.81**

Mid Quintile

	Grade	3	5	6	7	8
	SD	5.63	5.20	8.70	7.76	7.49
Cauc	*M*	198.74	228.94	243.36	264.58	277.00
AfAm	*M*	197.38	230.25	240.55	259.22	277.78
Cauc	*n*	19	17	14	19	13
AfAm	*n*	8	8	11	9	9
	d	**-0.24**	**0.25**	**-0.32**	**-0.69**	**0.10**
	NR	**-0.41**	**-0.36**	**-0.12**	**-0.36**	**-0.18**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	7.84	9.87	11.68	11.23	17.72
Cauc	*M*	167.00	183.33	178.00	204.50	216.67
AfAm	*M*	161.70	181.77	183.25	203.68	209.00
Cauc	*n*	4	3	3	2	3
AfAm	*n*	20	22	24	22	21
	d	**-0.68**	**-0.16**	**0.45**	**-0.07**	**-0.43**
	NR	**0.67**	**0.76**	**0.78**	**0.83**	**0.75**

Table 5, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	7.81	5.47	7.14	6.92	6.37
Income						
not low	*M*	242.00	277.44	297.69	313.00	328.95
low	*M*	234.00	271.00	291.50	315.00	328.00
not low	*n*	25	25	26	21	19
low	*n*	1	1	2	3	2
	d	**-1.02**	**-1.18**	**-0.87**	**0.29**	**-0.15**
	NR	**-0.92**	**-0.92**	**-0.86**	**-0.75**	**-0.81**

Mid Quintile

	Grade	3	5	6	7	8
	SD	5.63	5.20	8.70	7.76	7.49
not low	*M*	198.33	228.57	241.63	263.79	277.63
low	*M*	198.33	233.50	243.00	260.89	276.50
not low	*n*	21	21	16	19	16
low	*n*	6	4	9	9	6
	d	**0.00**	**0.95**	**0.16**	**-0.37**	**-0.15**
	NR	**-0.56**	**-0.68**	**-0.28**	**-0.36**	**-0.45**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	7.84	9.87	11.68	11.23	17.72
not low	*M*	167.00	183.33	178.00	204.50	218.00
low	*M*	161.42	181.77	183.48	203.60	206.65
not low	*n*	5	3	4	4	7
low	*n*	19	22	23	20	17
	d	**-0.71**	**-0.16**	**0.47**	**-0.08**	**-0.64**
	NR	**0.58**	**0.76**	**0.70**	**0.67**	**0.42**

Table 5, cont.

Top Quintile

	Grade	3	5	6	7	8
	SD	7.81	5.47	7.14	6.92	6.37
Gender						
Girls	*M*	239.00	278.36	297.06	313.13	326.91
Boys	*M*	244.83	276.33	297.55	313.44	331.00
Girls	*n*	14	11	17	15	11
Boys	*n*	12	15	11	9	10
	d	**0.75**	**-0.37**	**0.07**	**0.04**	**0.64**
	NR	**-0.08**	**0.15**	**-0.21**	**-0.25**	**-0.05**

Mid Quintile

	Grade	3	5	6	7	8
	SD	5.63	5.20	8.70	7.76	7.49
Girls	*M*	199.50	231.00	242.77	261.71	276.88
Boys	*M*	197.08	226.90	241.42	264.64	278.80
Girls	*n*	14	15	13	17	17
Boys	*n*	13	10	12	11	5
	d	**-0.43**	**-0.79**	**-0.16**	**0.38**	**0.26**
	NR	**-0.04**	**-0.20**	**-0.04**	**-0.21**	**-0.55**

Bottom Quintile

	Grade	3	5	6	7	8
	SD	7.84	9.87	11.68	11.23	17.72
Girls	*M*	160.11	179.62	182.20	209.00	217.90
Boys	*M*	164.07	184.50	182.94	200.00	204.29
Girls	*n*	9	13	10	10	10
Boys	*n*	15	12	17	14	14
	d	**0.50**	**0.50**	**0.06**	**-0.80**	**-0.77**
	NR	**0.25**	**-0.04**	**0.26**	**0.17**	**0.17**

Figure 1. Conceptual model of reciprocal interaction of environmental, socio-cultural, and biological influences on brain development.

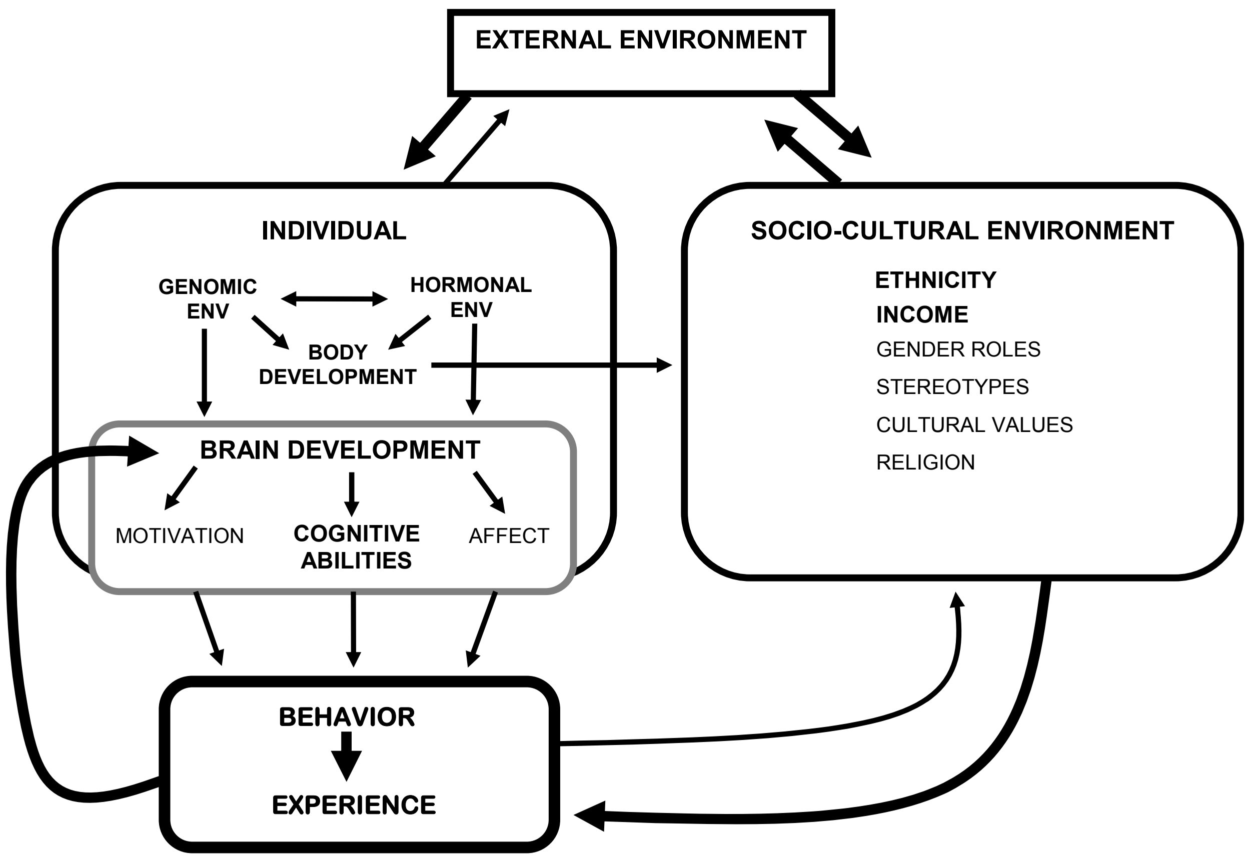


Figure 2. Mean scale scores by grade for ITBS overall language arts, mathematics, and science domains.

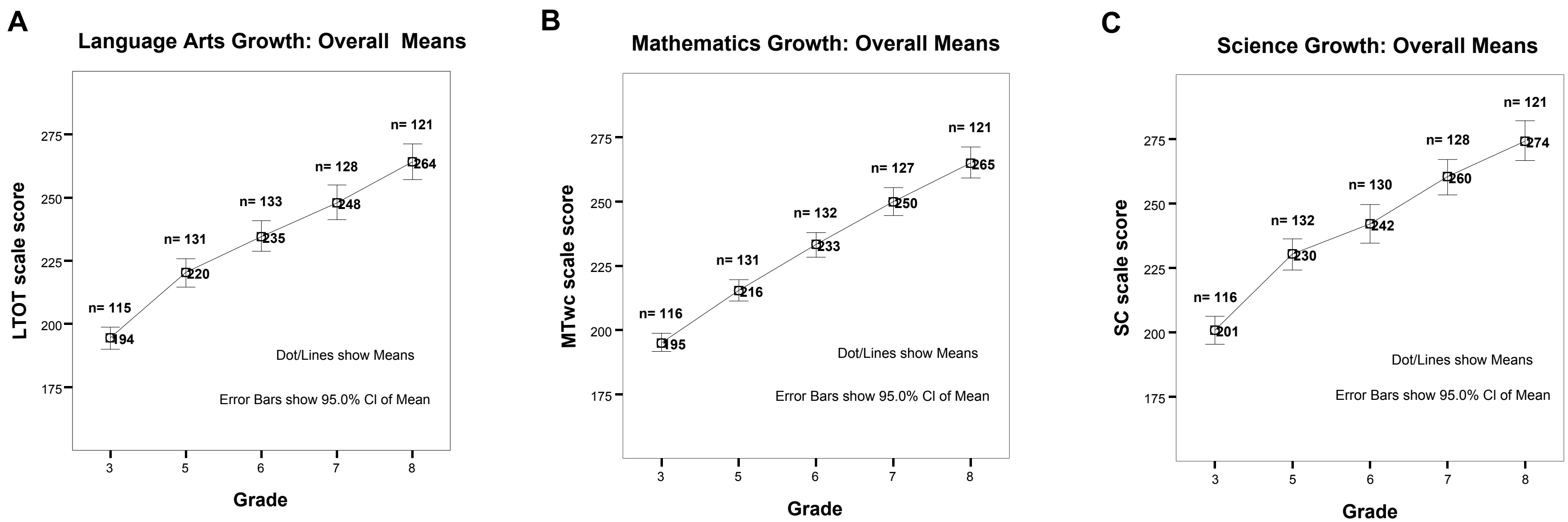


Figure 3. Overall distribution curves by grade level for language arts (A), mathematics (B), and science (C)

A

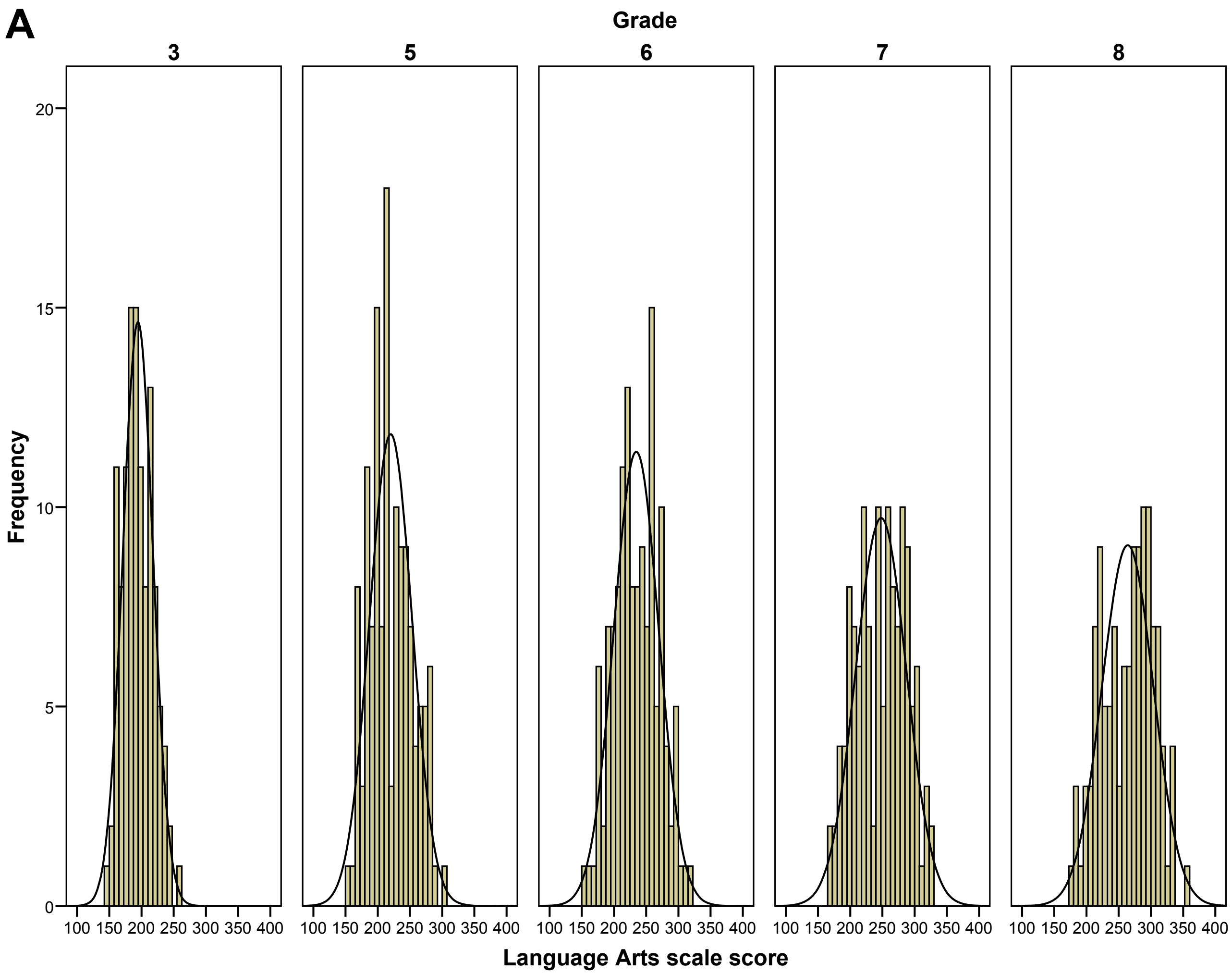


Figure 3, cont. Overall distribution curves by grade level for language arts (A), mathematics (B), and science (C)

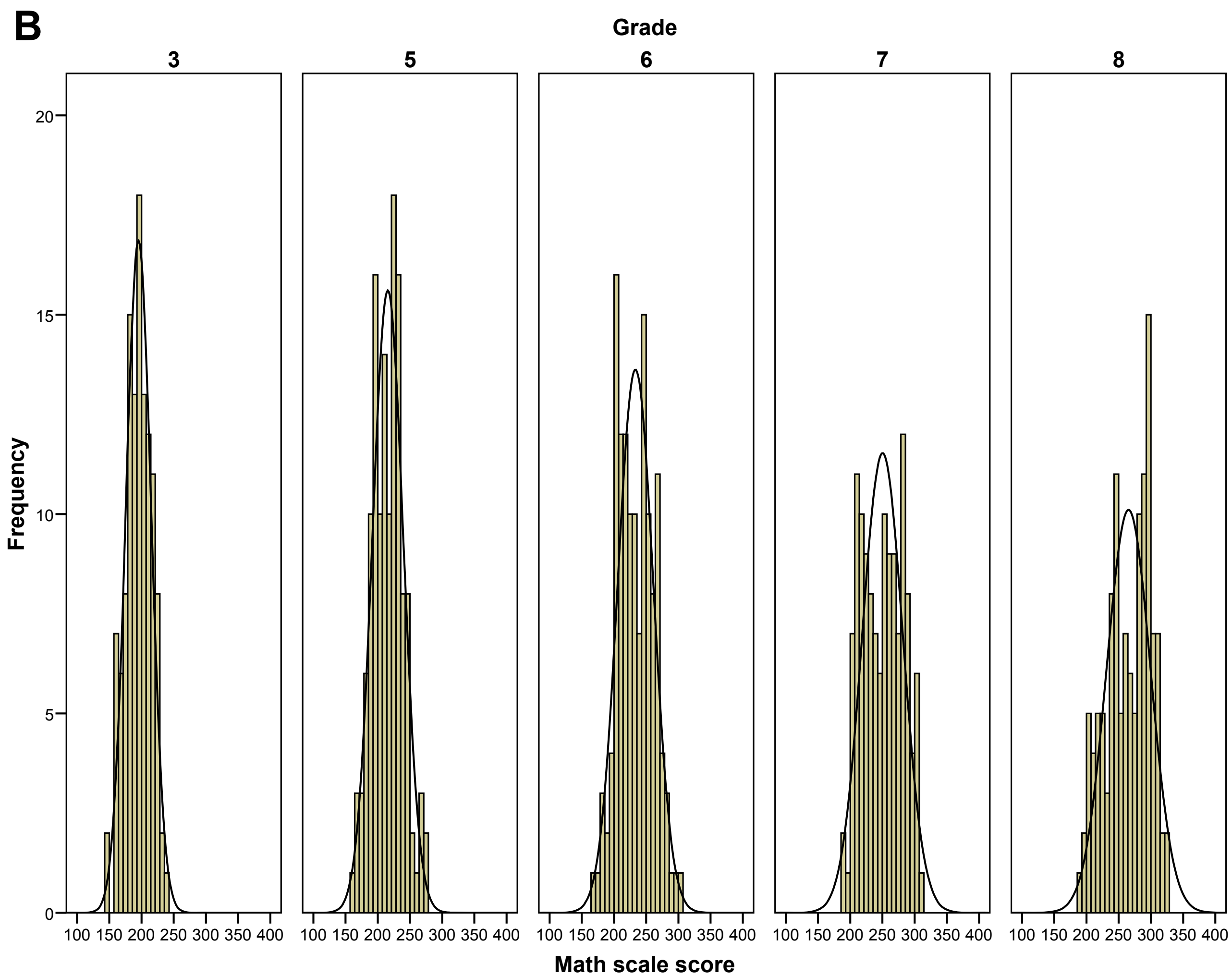


***Figure 3, cont*.** Overall distribution curves by grade level for language arts (A), mathematics (B), and science (C)

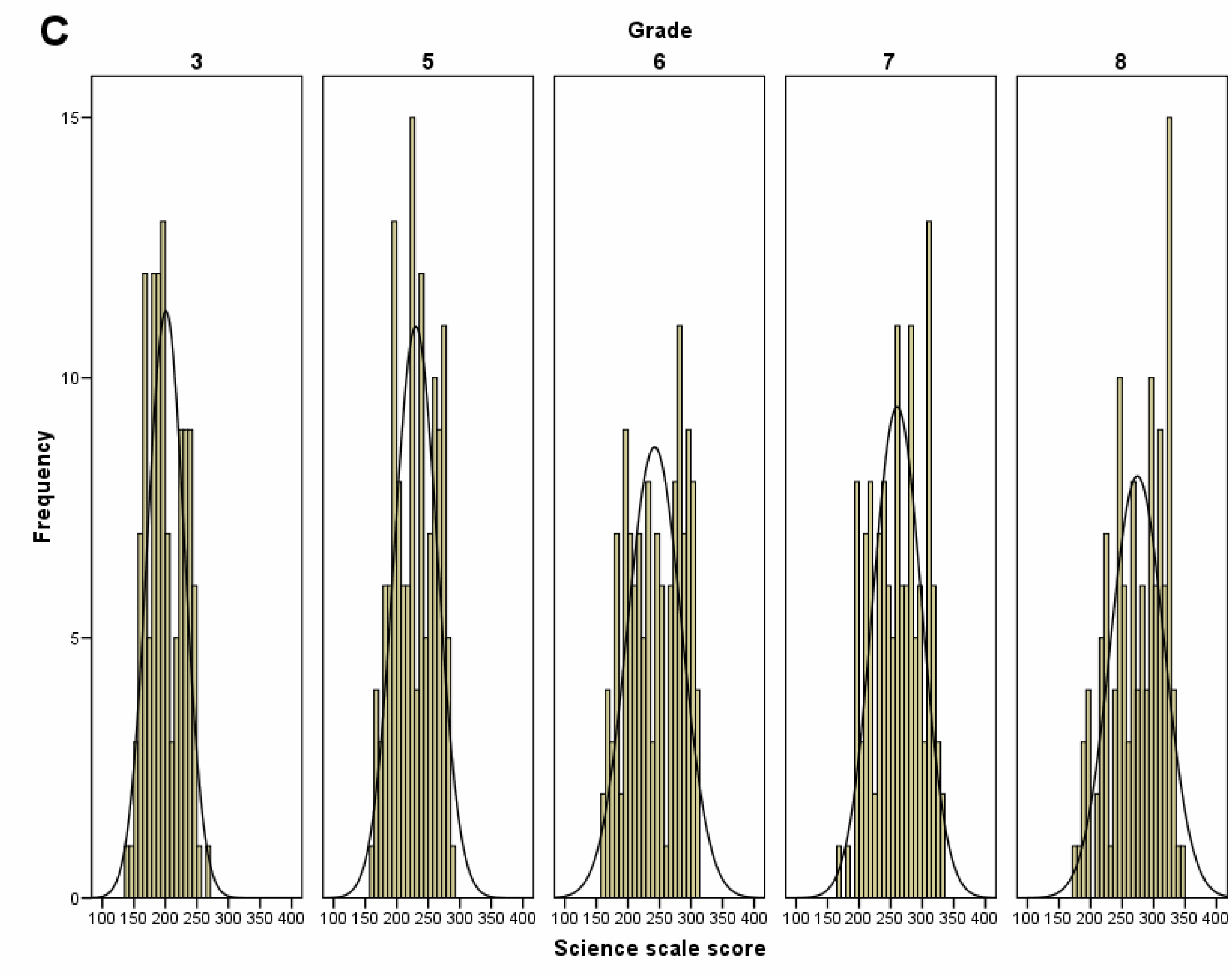


Figure 4. Distribution curves by gender by grade level for language arts (A), mathematics (B), and science (C)

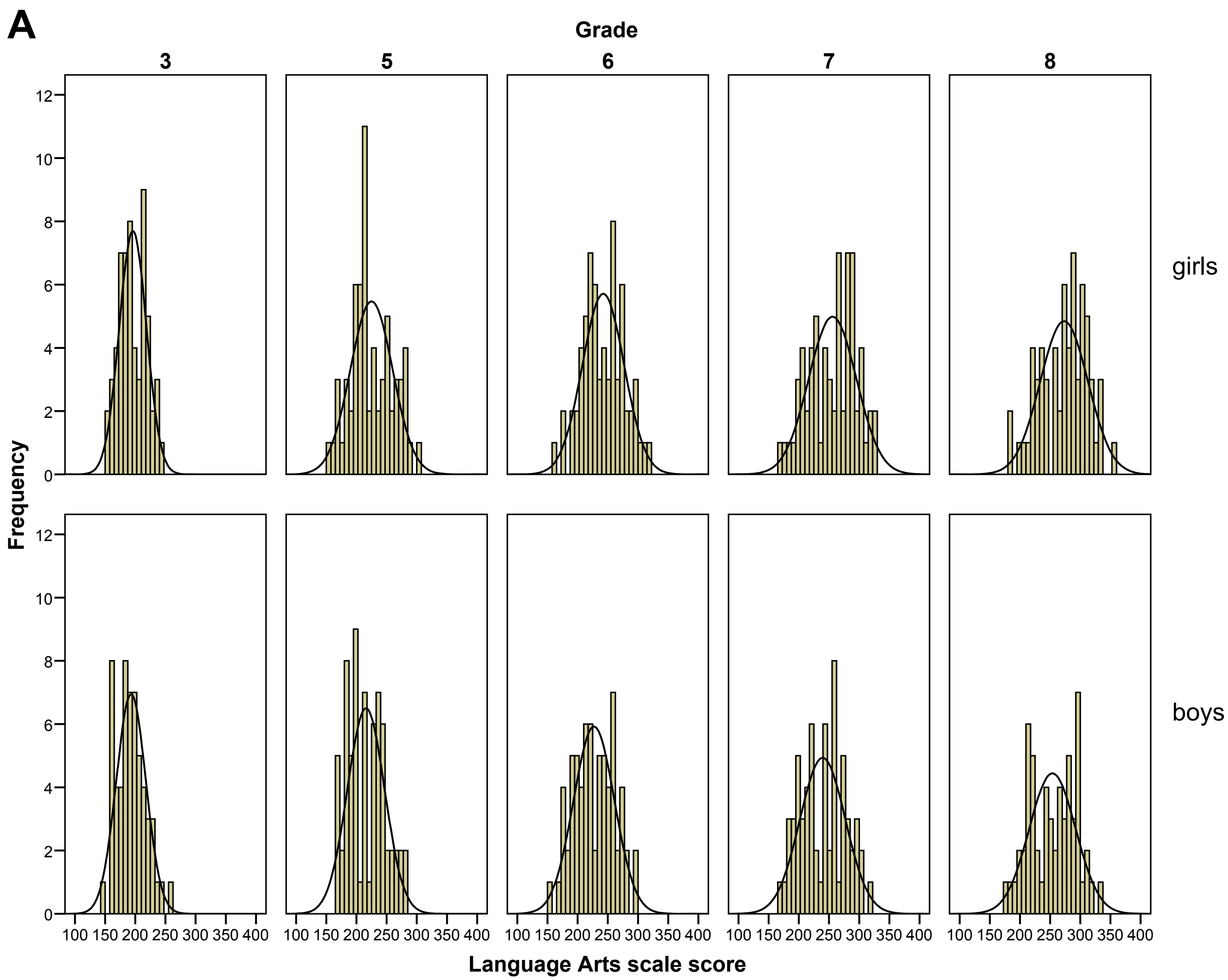


Figure 4, cont. Distribution curves by gender by grade level for language arts (A), mathematics (B), and science (C)

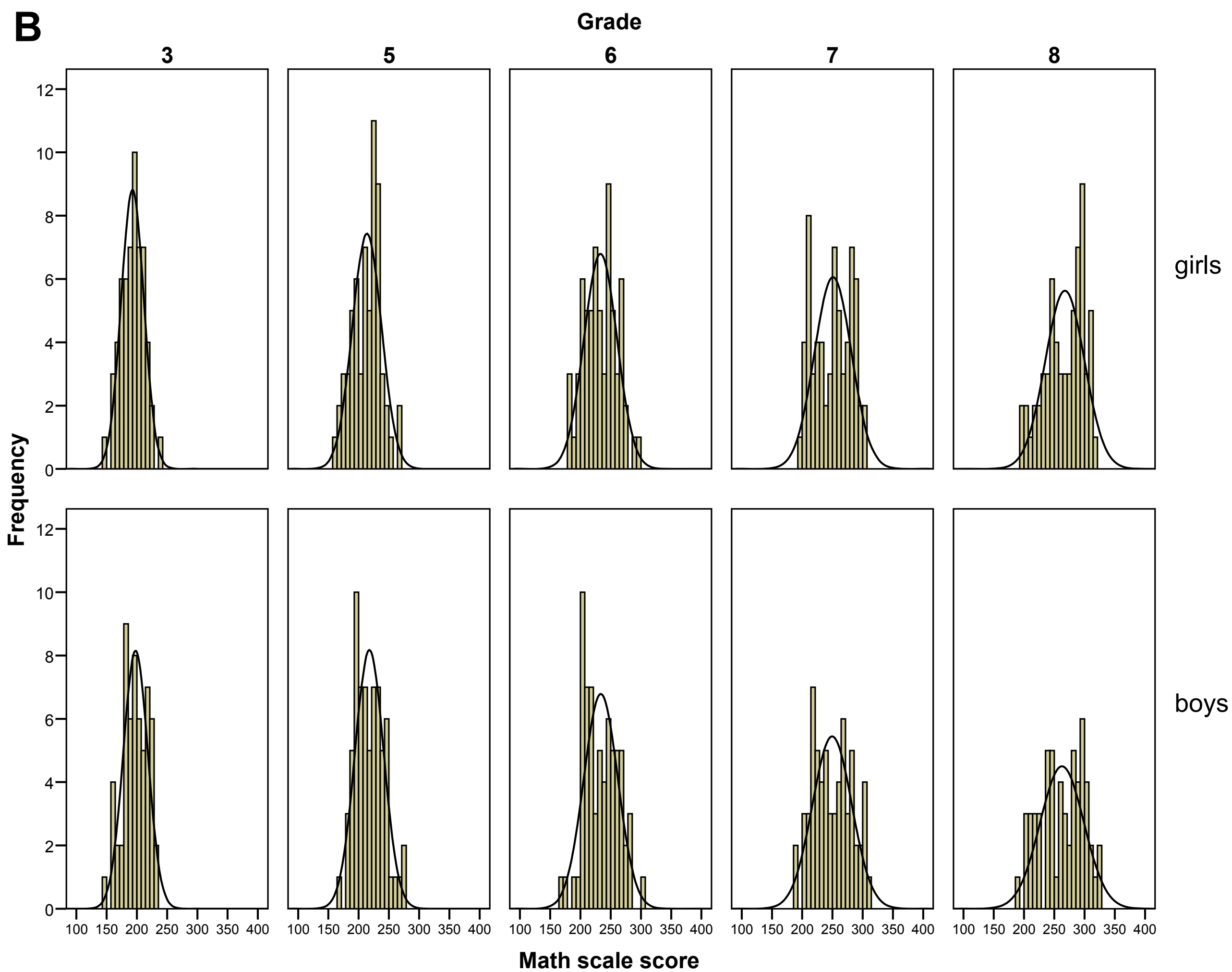


***Figure 4, cont*.** Distribution curves by gender by grade level for language arts (A), mathematics (B), and science (C)

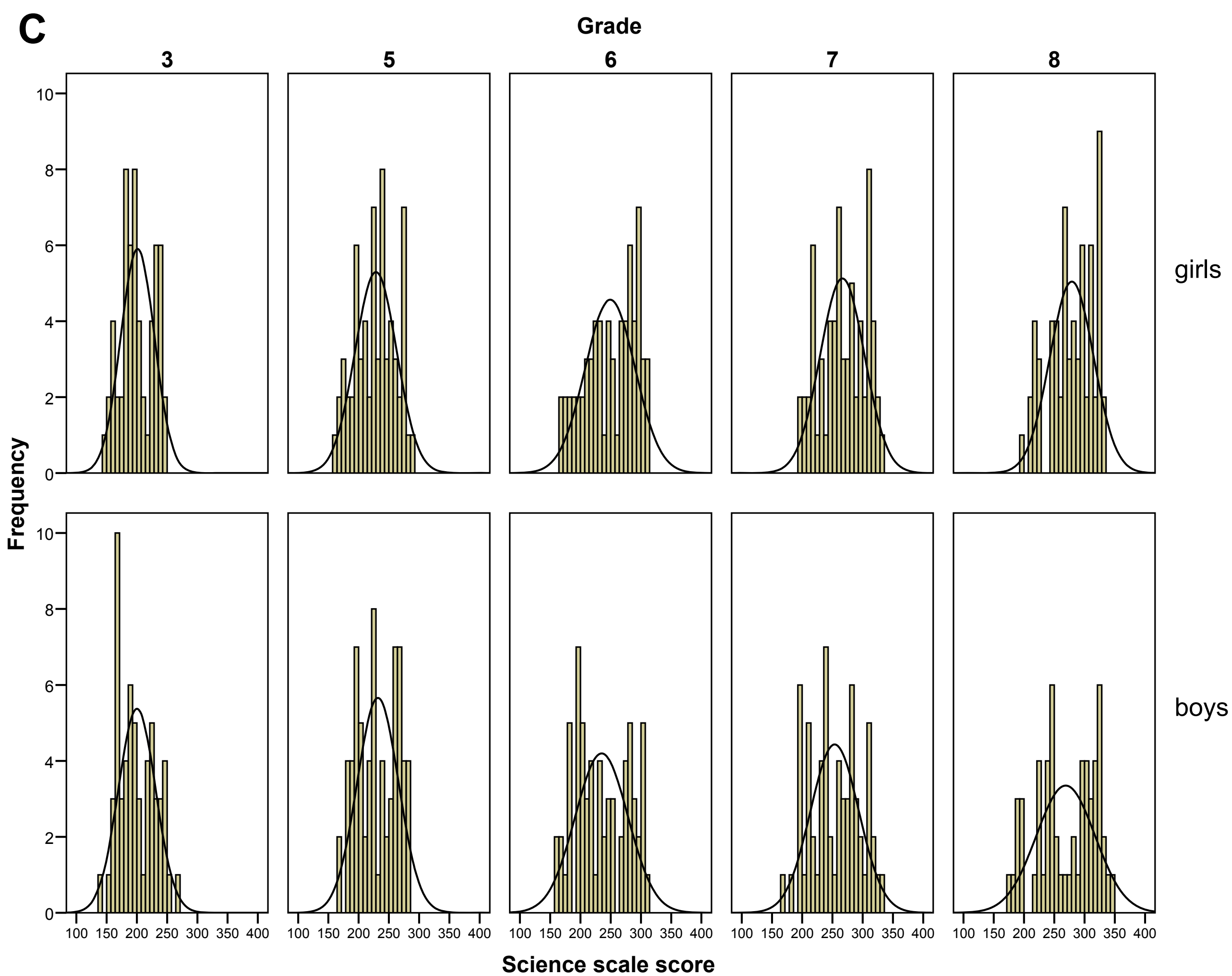


Figure 5. Distribution curves by gender and ethnicity subgroups by grade level for language arts (A), mathematics (B), and science (C)

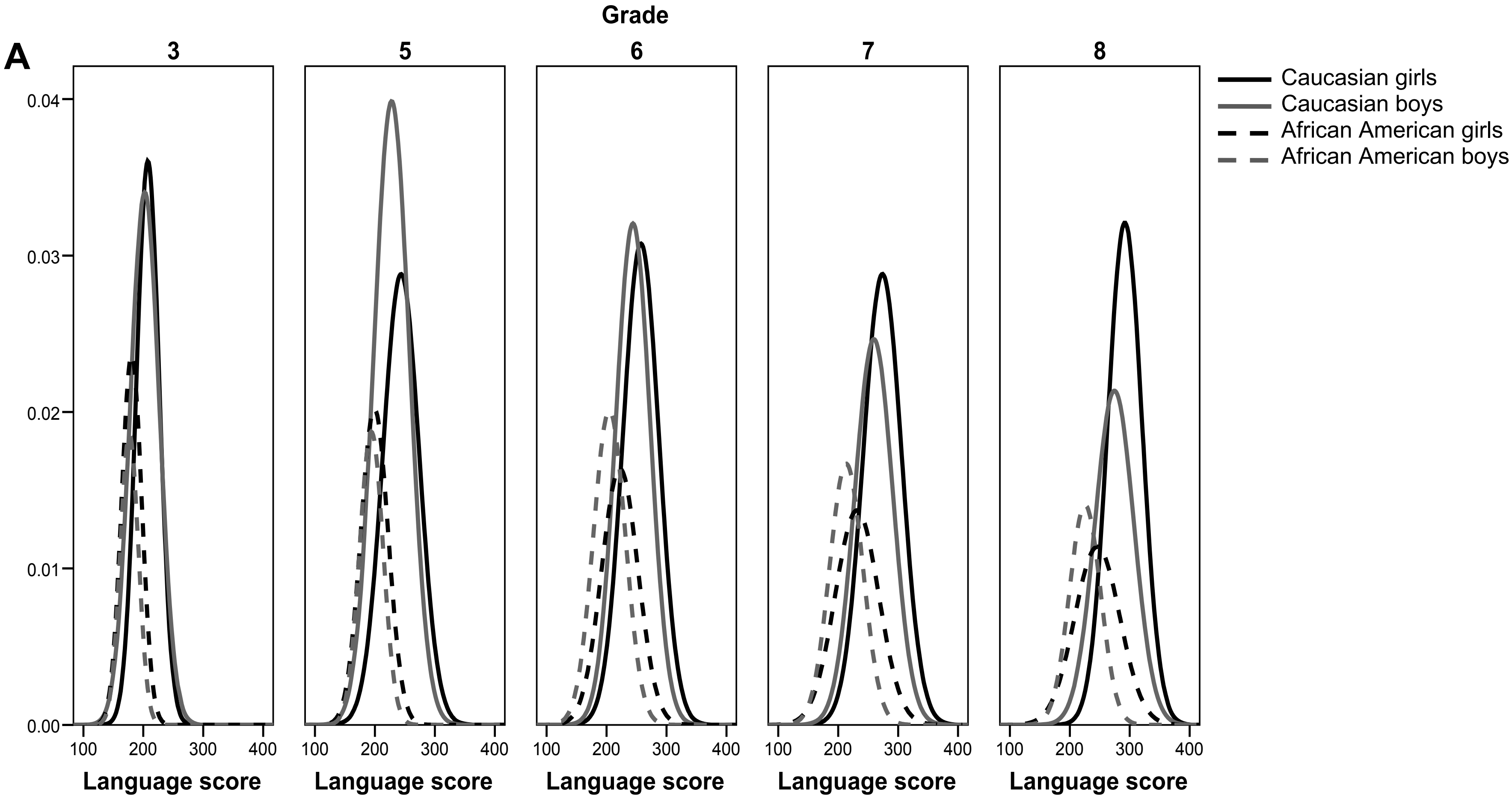


***Figure 5, cont*.** Distribution curves by gender and ethnicity subgroups by grade level for language arts (A), mathematics (B), and science (C)

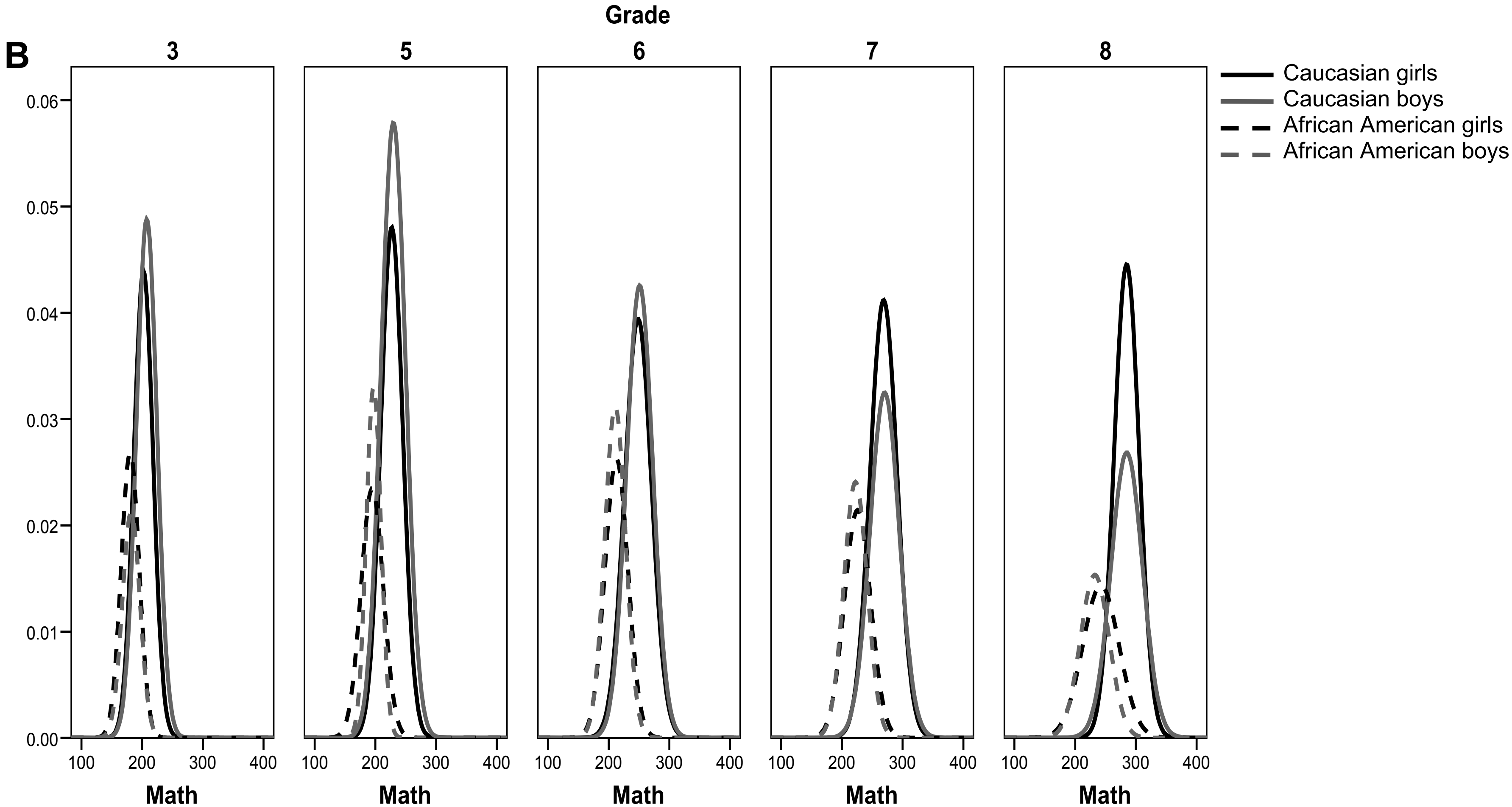
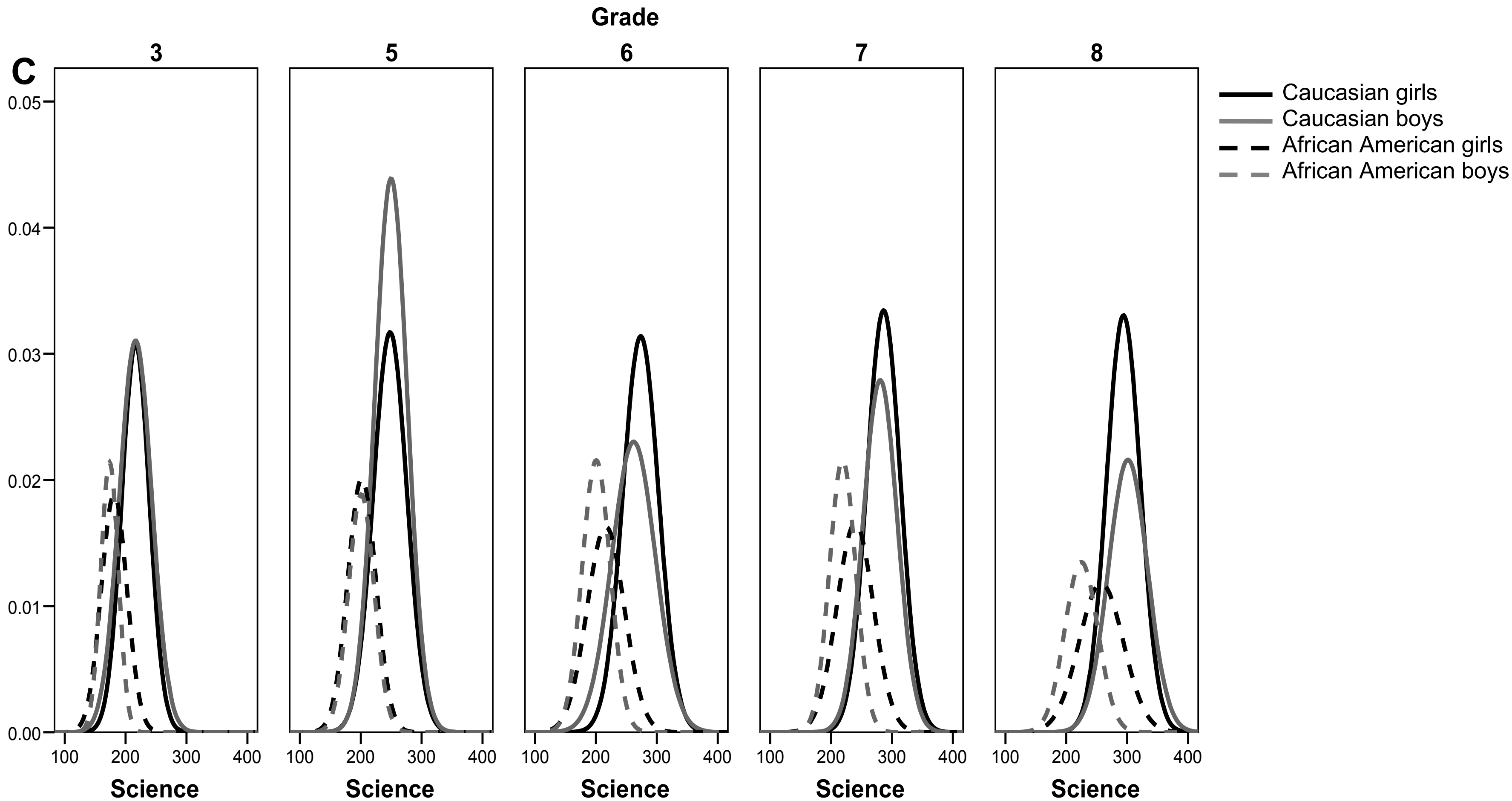


__Figure 5, cont__. Distribution curves by gender and ethnicity subgroups by grade level for language arts (A), mathematics (B), and science (C)



;

Figure 6. Average performance over time by gender and ethnicity subgroups in language arts, mathematics, and science

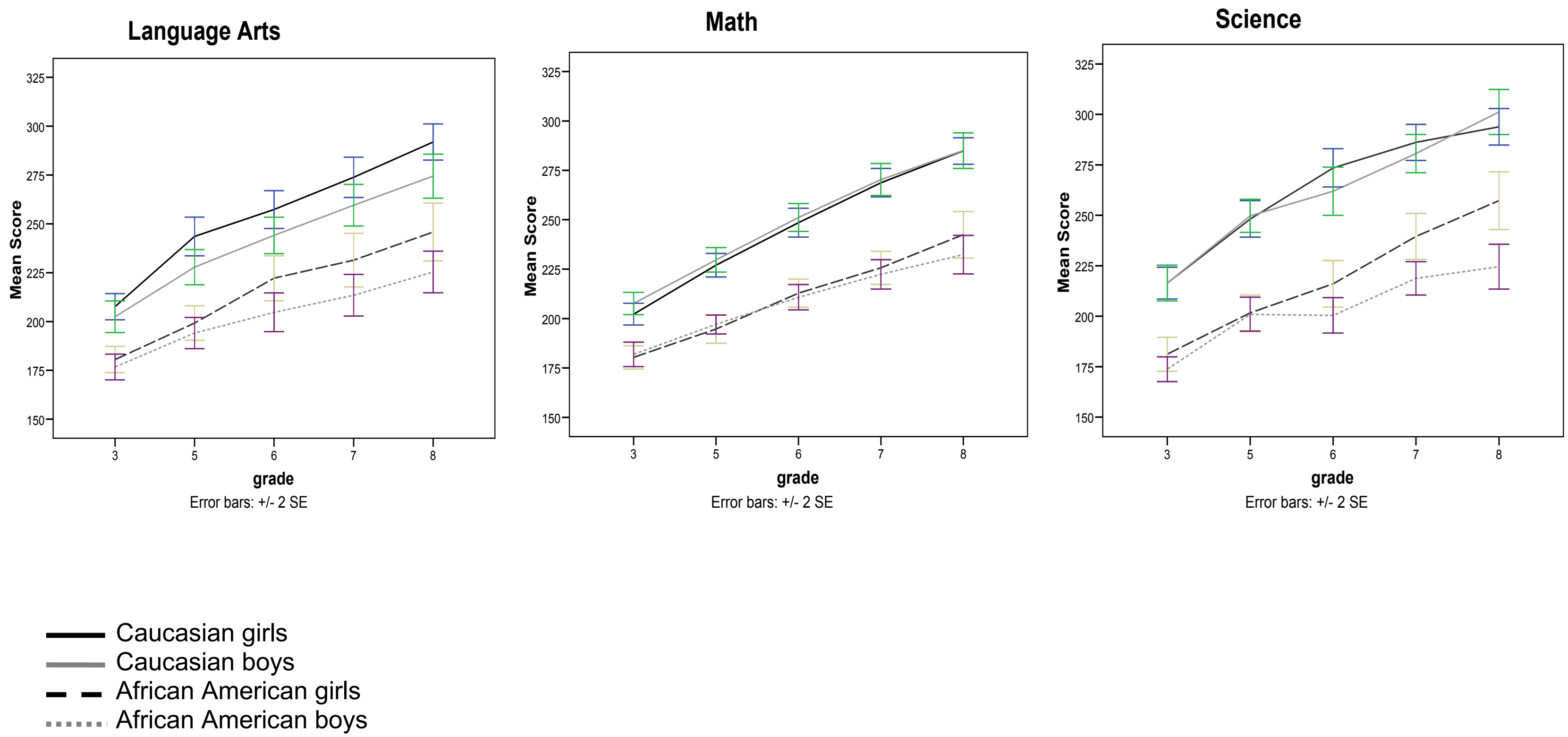


Figure 7. Random 20% samples of individual raw data trajectories for language arts, mathematics, and science

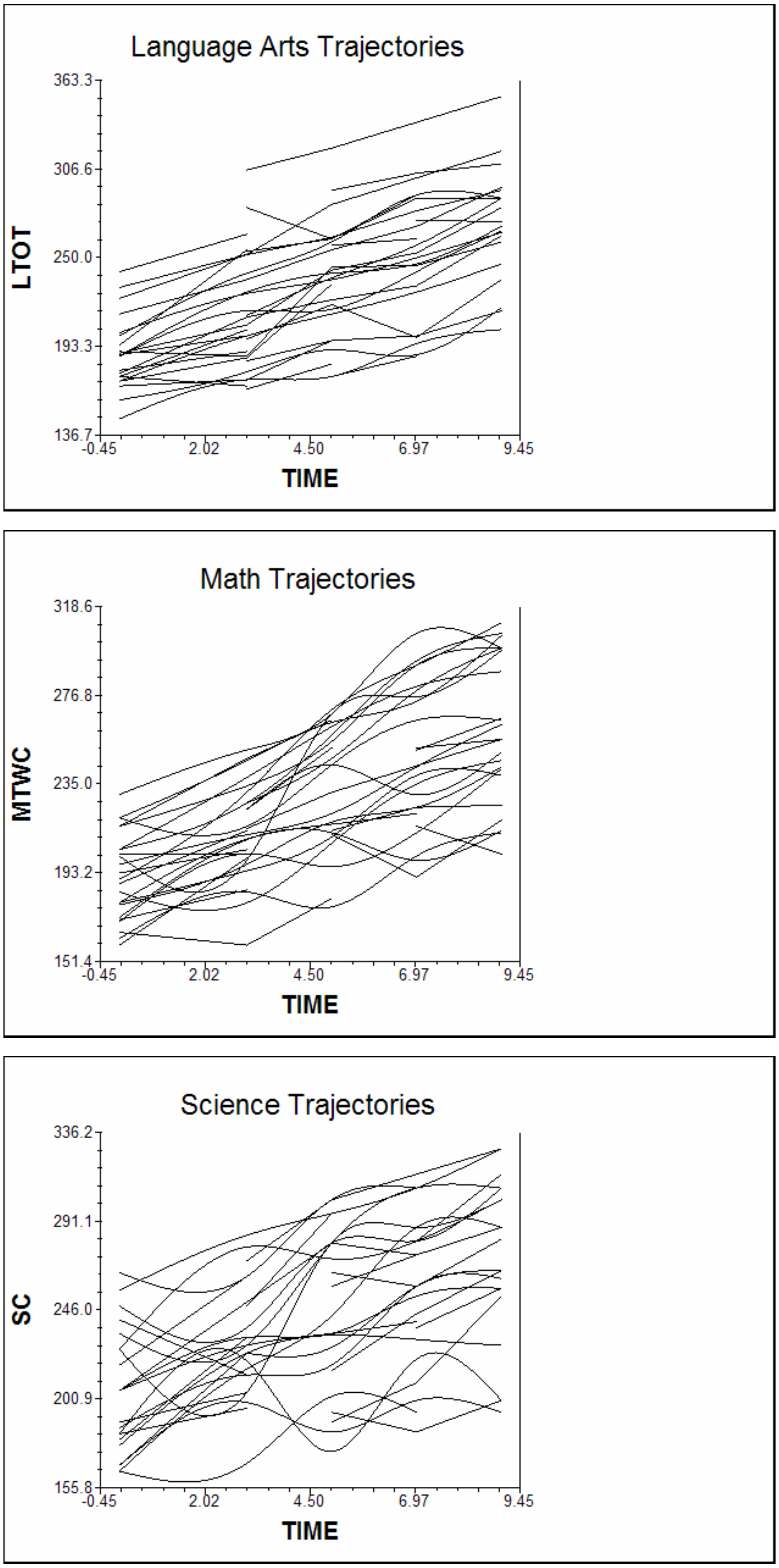


Figure 8. Predicted score as a function of grade, gender, and ethnicity in language arts, mathematics, and science

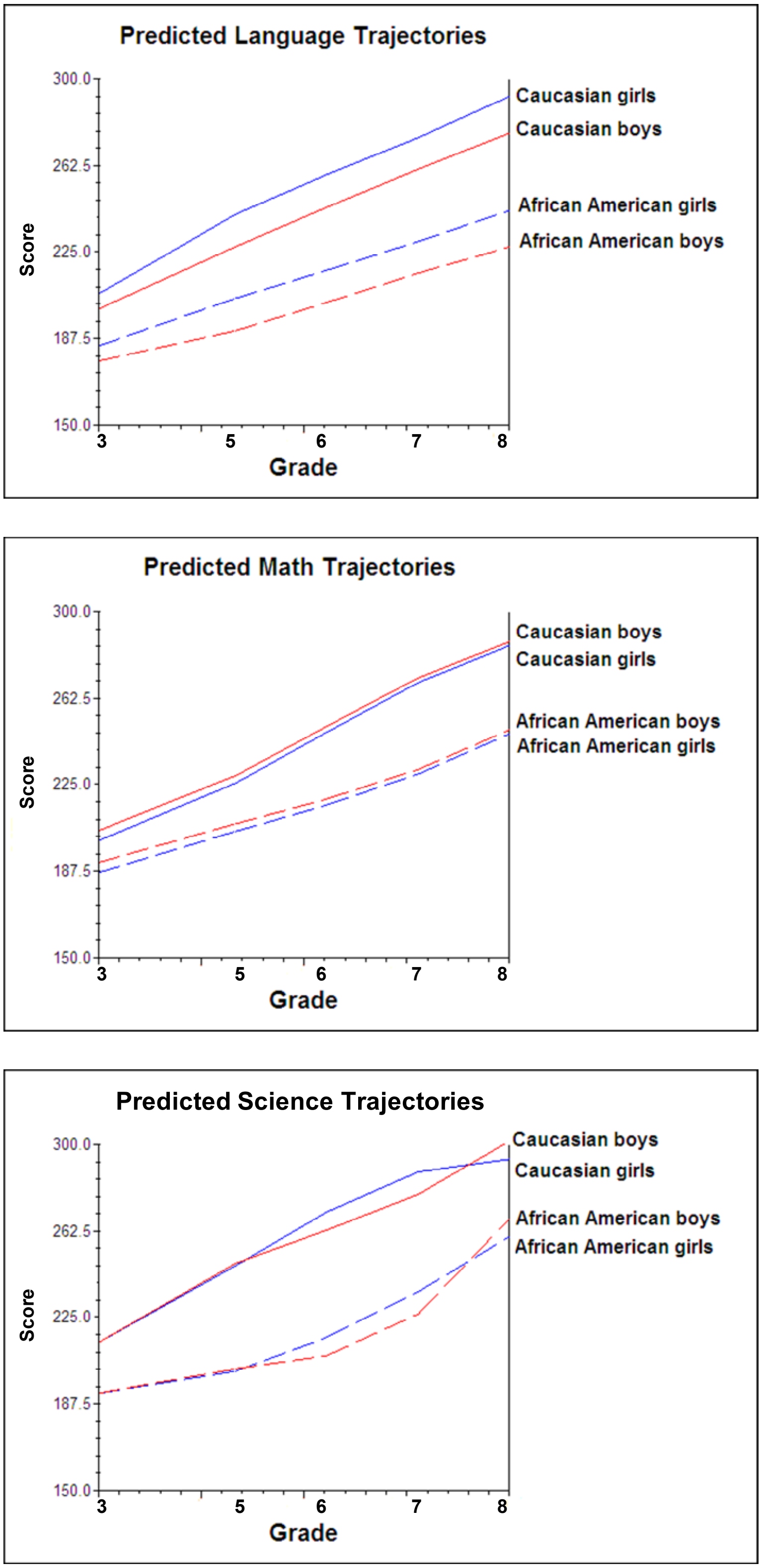


Figure 9. Language arts: effect size change over time for ethnicity (A), income (B), and gender (C).

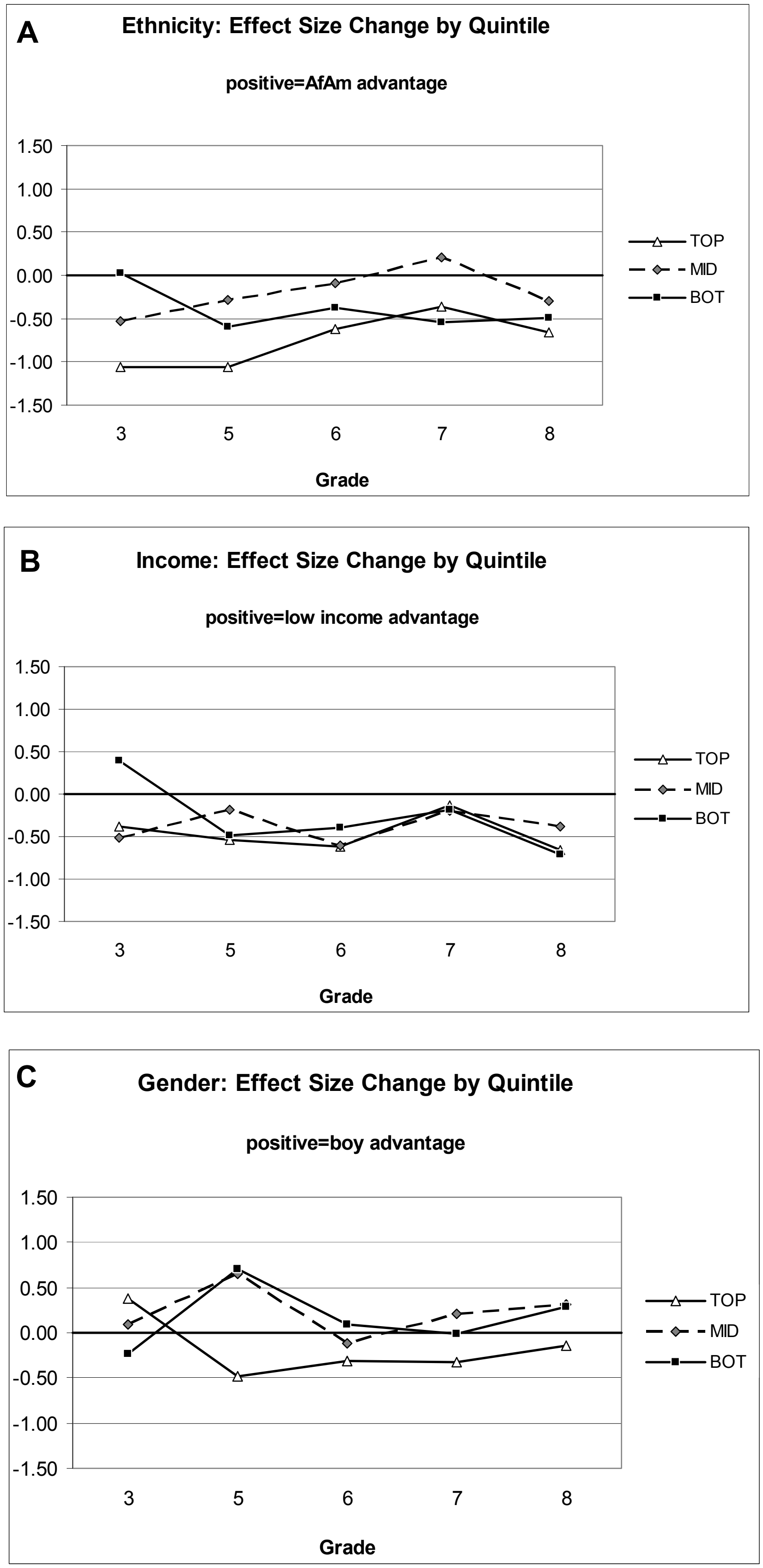


Figure 10. Language arts: student ratio change over time for ethnicity (A), income (B), and gender (C).

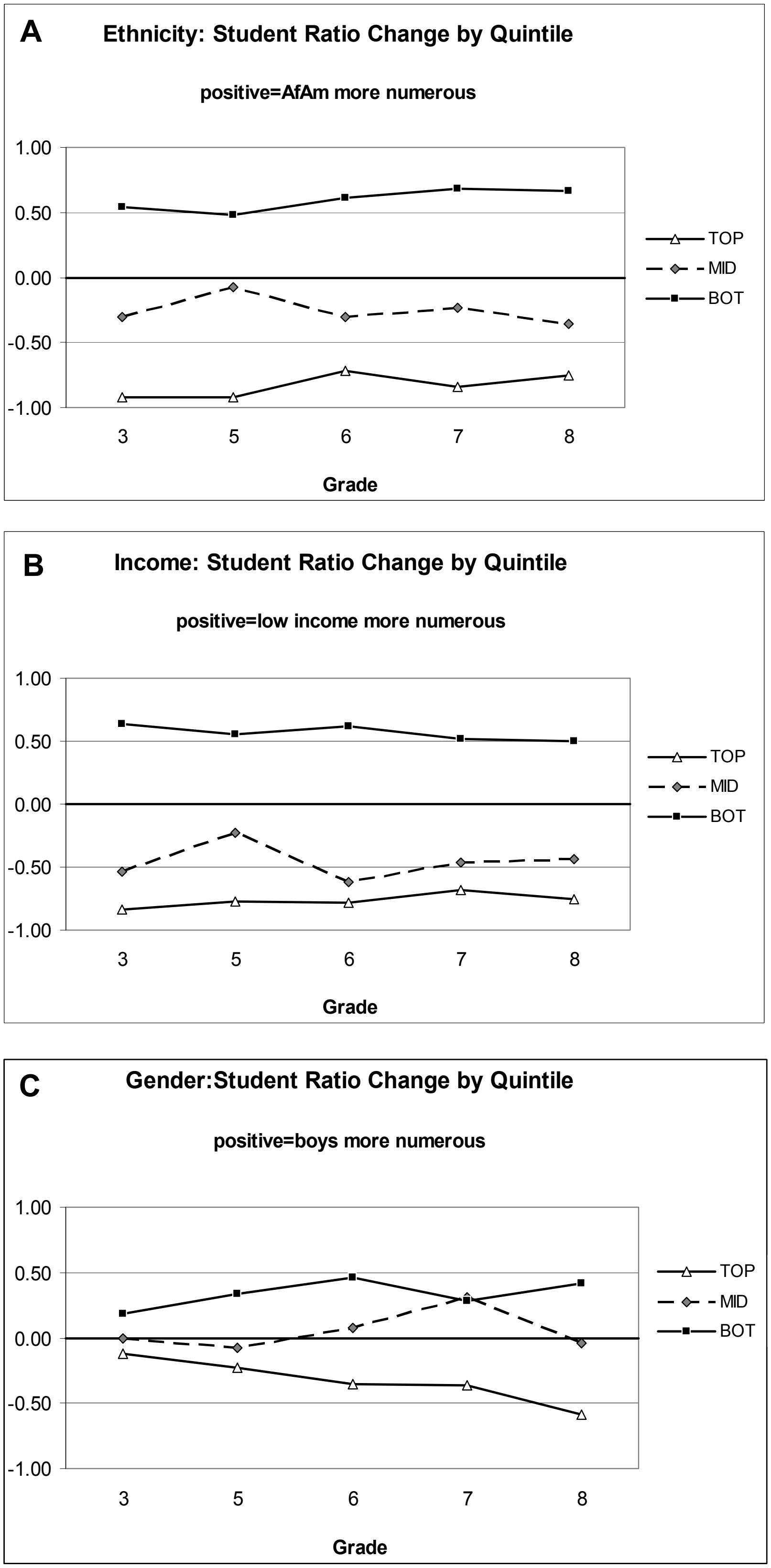


Figure 11. Mathematics: effect size change over time for ethnicity (A), income (B), and gender (C).

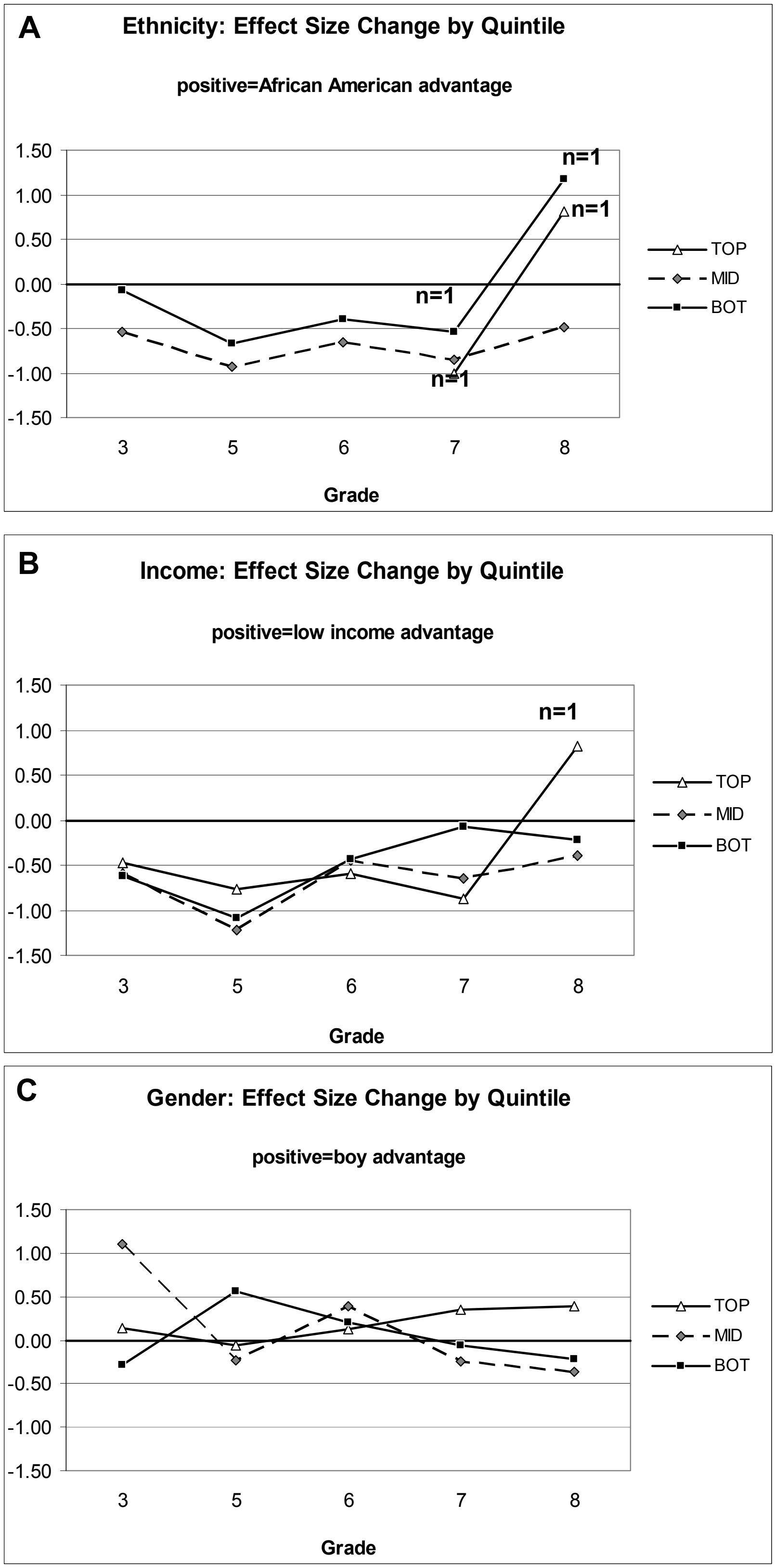


Figure 12. Mathematics: student ratio change over time for ethnicity (A), income (B), and gender (C).

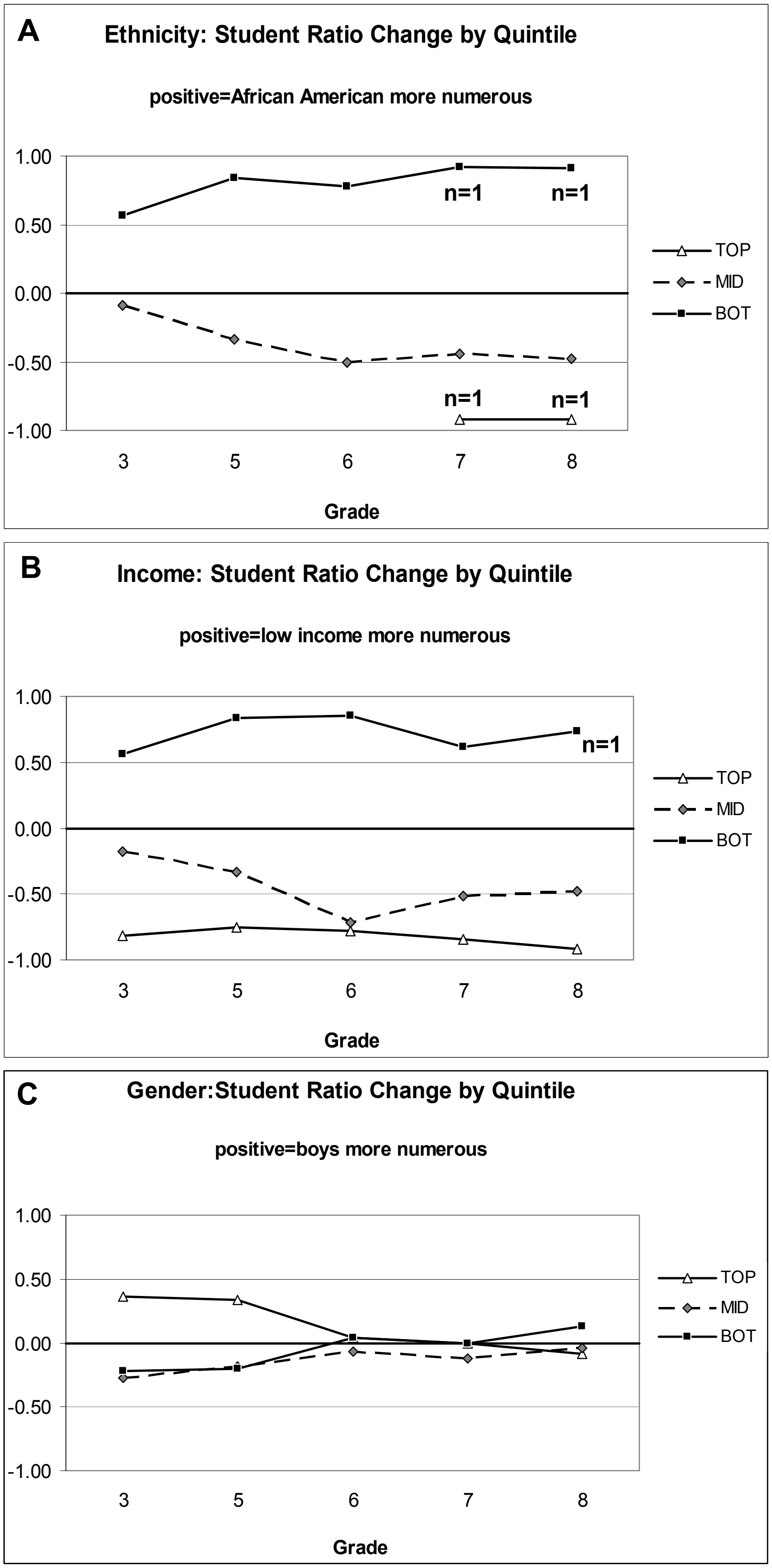


Figure 13. Science: effect size change over time for ethnicity (A), income (B), and gender (C).

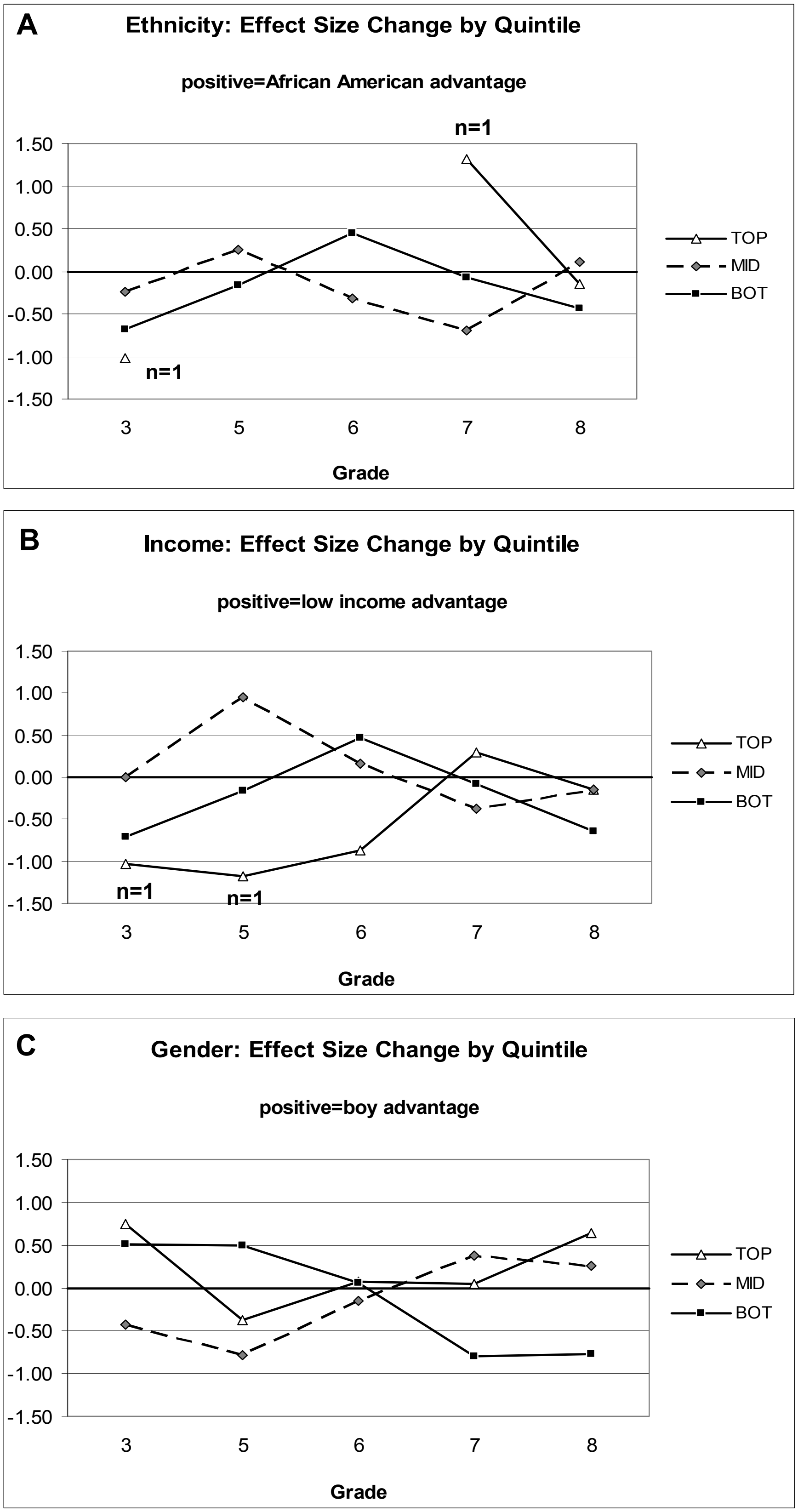


Figure 14. Science: student ratio change over time for ethnicity (A), income (B), and gender (C).