Tripartite Growth Trajectories of Reading and Math Achievement: Tracking National Academic Progress at Primary, Middle, and High School Levels

Jaekyung Lee

Am Educ Res J 2010 47: 800 originally published online 19 April 2010
DOI: 10.3102/0002831210365009

The online version of this article can be found at:
http://aer.sagepub.com/content/47/4/800

Published on behalf of

American Educational Research Association

and

SAGE

http://www.sagepublications.com

Additional services and information for American Educational Research Journal can be found at:

Email Alerts: http://aerj.aera.net/alerts

Subscriptions: http://aerj.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions
Tripartite Growth Trajectories of Reading and Math Achievement: Tracking National Academic Progress at Primary, Middle, and High School Levels

Jaekyung Lee
University at Buffalo, State University of New York

This study examines trends in American students' growth trajectories in reading and math achievement over the past three decades. Drawing upon multiple sources of national assessment data, cohort analyses provide new evidence on the stability and change of national academic growth curves. The emerging trends imply a tripartite pattern where American students are gaining ground at the pre/early primary school level, holding ground at the middle school level, and losing ground at the high school level. National progress in reading and math achievement at the pre/early primary school level appears to be offset by declines at the high school level. The study discusses the limitations and challenges of tracking academic growth trajectories across all different levels of education over the long term. It also calls for national P-16 education policy and research efforts toward sustainable academic growth and seamless educational transition.

KEYWORDS: achievement, NAEP, longitudinal studies, human development, testing

Sustained human capital development and academic growth is a key concern for national education policy. Since A Nation at Risk (National Commission on Excellence in Education, 1983), there has been a wealth of research and policy discussion about the state of the American school system and the performance of its students. High school education has often been at the center of these debates, which have resulted in numerous and often-polarized policy recommendations for high school reform (e.g., Berliner & Biddle, 1995; National Education Summit on High Schools, 2005; Powell, Farrar, & Cohen, 1985; Ravitch, 2003; Sizer, 1992). American

Jaekyung Lee is an associate professor of education at the State University of New York at Buffalo, 402 Baldy Hall, Buffalo, NY 14260-1000, and a fellow of the Center for Advanced Study in the Behavioral Sciences at Stanford University; e-mail: JL224@buffalo.edu. His research focuses on educational policy analysis for excellence, equity, and accountability.
high schools have been criticized for performing like a shopping mall where too many courses are offered with too little depth. In the midst of globalization and economic competition, policy also has been driven by comparing achievement in the American education system with achievement in higher performing industrialized countries (Baker, 2003; Lee, 2001). The Trends in International Math and Science Study (TIMSS) and prior international test results showed that the U.S. achievement gap relative to other countries tends to grow larger at the middle and high school levels.1

Given the past two decades of standards-based education reform and test-driven school accountability policy, American educational policymakers are left with a puzzle: Standardized achievement test scores have not risen, while more high school students have taken advanced courses in math and sciences and also more high-stakes tests for promotion and/or graduation. Obviously, high schools do not exist in a vacuum and understanding the state of high schools requires a broader and developmental view of educational trajectory in the whole system, although one may question if there is such a thing that can be called a system. Ravitch (2003) aptly raised these questions: “To what extent are the weaknesses of the lower schools hampering the effectiveness of the high schools? To what extent are the demands of colleges and universities hampering the effectiveness of the high schools?” (p. 3). Without a solid knowledge base of students’ average academic growth trajectory and schools’ value-added contribution in the process, simple answers to these questions are likely to end up simply blaming adjacent levels of education for students’ poor performance or improper readiness, merely passing the burden of correcting identified problems onto others.2

In comparison with secondary education, there is also no shortage of policy and research on academic improvement in early childhood or elementary education. Federal programs such as Head Start and Title I programs went through major changes during the past two decades. A plethora of Title I program reform efforts culminated in the No Child Left Behind Act of 2001 (NCLB) that relies on high-stakes testing and a test-driven school accountability system for improving reading and math achievement of all students.3 The School Readiness Act of 2007 pushes for Head Start reform toward improving teacher quality and strengthening academic focus on school readiness.

While there has been no shortage of policy and research tackling many issues of American students’ academic performance at separate stages of development (e.g., early childhood, childhood, adolescence) or at separate levels of the educational system (e.g., preschool, elementary, secondary, post-secondary), relatively few policy and research agendas have addressed cross-cutting issues among different levels (e.g., academic transition and growth from one level to another). These issues include problems such as locating the weaknesses in the P-16 educational transition pipeline or the gap in the K–12 academic growth trajectory. The need for more systematic research on the full course of academic transition and growth has grown from recent
national education policy movements. K–16 or P–16 initiatives call for the seamless transition of students between levels of schooling. Notwithstanding the lofty goal implied in policy rhetoric, fundamental scientific research and information databases for effective policy planning and intervention are not yet in place. For example, we do not yet have a national-level longitudinal data-set that tracks students’ achievement from preschool or kindergarten through high school, not to mention postsecondary education.

The key questions for developmentally oriented educational policy include whether and how the developmental needs of children or adolescents are met by school learning environments and whether and how the timely enactment of an effective policy intervention can alter the course of children’s developmental trajectories through environmental changes. It should build upon a recognition that the effect of an intervention may depend on the age and developmental needs of the children as well as on the timing, quality, or intensity of the intervention (Foster & Kalil, 2005; Yoshikawa & Hsueh, 2001). While the effects of early interventions such as preschool and full-day kindergarten programs on reading and math achievement are well-established, prior research revealed that those academic benefits are difficult to sustain later on after early intervention (e.g., Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Fusaro, 1997; Gormley, Gayer, Phillips, & Dawson, 2005; Lee, Brooks-Gunn, Schnur, & Liaw, 1990; Ramey & Ramey, 1998; Reynolds, Temple, Robertson, & Mann, 2002; Walston & West, 2004). Furthermore, the effects of an intervention on children may be mediated or moderated by a variety of educational processes and interactions at multiple levels of the school system. Although there is a growing body of longitudinal studies that examine environmental or intervention effects on individual academic growth trajectories, they often lack a dynamic view of the environment and ignore long-term effects of environmental changes over generations.

In light of these concerns, this study addresses these related questions: How does American students’ average reading and math achievement grow over the full course of schooling? Is there any generalizable and consistent pattern of academic growth over generations across the nation? Did the national aggregate patterns of academic growth change over the past three decades and how? Did changes in the growth patterns, if any, vary among different levels of age or grade? While prior research has addressed the first two questions but not the last two, this study attempts to update and expand our knowledge base through more systematic analyses of existing national datasets and critical synthesis of the results. Implications of this study’s findings for educational policy and research will be discussed.

Limitations and Challenges of Research on Academic Growth

The answers to the first two research questions may be provided partly by a review of prior research. There are two lines of research on this topic; one line of research is the longitudinal study of individual-level growth trajectories. For
instance, Benjamin Bloom (1964) published a pioneering meta-analytic synthesis of many longitudinal studies of academic growth trajectory. He concluded that approximately 50% of academic achievement at grade 12 (age 18) has been reached by the end of grade 3 (age 9) and about 75% has been developed by about age 13 (grade 7). Emphasizing the importance of the first few years of school as well as the preschool period in the development of learning and achievement, Bloom called for early interventions for children at academic risk to prevent continued failure throughout the remainder of their school career. However, the literature for his study is outdated and his generalization about this diminishing growth pattern has not been time tested. Further, previous studies of academic growth trajectories often drew on data from a limited age or grade span from relatively small samples of individual students and also relied on norm-referenced measurements of learning (Bloom, 1964; Lichten, 2004). This has threatened the validity of assessing academic growth and restricted the generalizability of study results to the larger population over the entire time span of schooling.

Another line of previous research relied on large cross-sectional samples to examine national-level aggregate academic growth patterns. This line of research handles the limitations of small-scale local studies and has produced more generalizable patterns of academic growth for the national population of students. These studies also found that observed academic growth is generally faster during the earlier grades, followed by a decelerating rate of growth over the course of schooling, even though the patterns of academic growth are not always consistent from subject to subject and from grade to grade (see Beggs & Hieronymus, 1968; CTB/McGraw-Hill, 1997, 2003; Harcourt Educational Measurement, 2002; Lichten, 2004; McGrew & Woodcock, 2001). Since the primary data sources were conventional national norms from test publishers, the test norms had limitations in that they were drawn from cross-sectional samples of different ages or grades’ cohort groups assessed during the same year. Therefore, this line of research was not able to produce accurate pictures of academic growth curves by confounding age or grade effects with cohort effects on growth.

Advances in nationally representative longitudinal data have helped address the limitations with regard to both internal and external validity by allowing researchers to track generalizable growth trajectories within a single cohort group and examine the complex array of factors that influence interindividual and inter-organizational variations in their growth patterns. However, restricting data analysis to a single cohort at a particular time period does not afford a long-term assessment of changes in the growth trajectories between sequential cohorts. While the general forms and characteristics of national aggregate growth curves are known relatively well, no systematic research has been done yet to compare different cohort groups’ growth patterns over the long run. Previous studies of National Assessment of Educational Progress (NAEP) achievement data usually examined trends at the same age level or grade level over time or cross-age/grade differences during
the same year. Only a few studies capitalized on cohort-based tracking of achievement with repeated cross-sectional or quasi-longitudinal samples of NAEP cohorts. For example, prior research has compared math achievement growth patterns between U.S. states (Coley, 1998, 2003) or countries (Lee & Fish, 2008).

Tracking changes of national academic growth over long-term periods would have implications for the malleability of growth curves and the role of educational policy interventions. As children grow older and progress through higher grades in school, their academic growth rate changes and the alterability of the growth curve changes as well. A generalization made by Bloom (1964) argued the following:

The effects of marked differences in the environment are greatest in the period of greatest normal growth and least in the periods of least normal growth. . . . Furthermore, a shift from one environment to another will have greatest consequence in a period of rapid normal growth and will have little effect on the individual during the period of least rapid normal growth. (p. 194)

What complicates this general rule are the possibilities that the environment itself may change over time during the course of child development and that an earlier environmental effect may not necessarily be sustained without the continuing presence of the same environmental factors or replacement by other factors with the same effects.

Despite many previous studies that showed clear and consistent environmental effects on academic growth in general, theoretical models that give direction to studying simultaneous and interactive changes in environmental and developmental trajectories are yet to be constructed. Bronfenbrenner (1979) used the metaphor of “moving macro-systems” in explaining that the macrosystem (e.g., cultural and institutional factors) also undergoes a process of development and that the members of a changing society necessarily experience developmental change in the process. Building on Bronfenbrenner’s ecological perspective of human development as occurring within a dynamic environmental system, we need to study the interplay between the simultaneous trajectories of student achievement outcomes and educational systems/policies. Given the possibility of interplay between two trajectories, one embedded within the other, changes in the elementary, middle, or high school education environment might be related to changes in academic growth trajectories at corresponding age or grade levels. The central questions in this study are whether and how different age or grade-level students’ achievement growth patterns have changed over the past three decades or so.

Method

This study is based on the premise that academic growth in reading and math can be measured on a common scale across grades. This requires
developmental scales that allow us to track students’ academic performance at various developmental levels. Prior research has demonstrated that the use of different metrics may lead to different representations of growth trajectories (Schulz & Nicewander, 1997; Seltzer, Frank, & Bryk, 1994; Yen, 1986). The development of more valid and robust developmental scales has been made available through vertical test equating based on item response theory (IRT) modeling (Kolen & Brennan, 2004; Lord, 1980). The advantage of the use of IRT models to equate tests across grades is that all of the items measuring performance in a particular content domain can be placed on a common scale of difficulty, and therefore, all scores can be placed on a common ability scale. Further, the development of growth curve modeling (e.g., hierarchical linear modeling) has facilitated empirical studies of growth trajectories (Bryk & Raudenbush, 1987; Singer & Willett, 2003). This study capitalizes on advances in these areas to examine both within- and between-cohort variations in national average achievement on common developmental scales.

Samples and Measures

This study utilizes existing national databases that provide repeated cross-sectional, quasi-longitudinal, or longitudinal data for systematic investigation of K–12 students’ growth trajectories in reading and math achievement. Tracking changes in academic achievement through a meta-analytic synthesis of evidence from multiple national databases that are built from representative samples of the national population has the potential to enhance both the scope and power of research (see list of databases in Appendix A, available as supplementary material for this article in the online version of the journal). Reading and math composite test scores across all datasets used in this study are treated as measures of common underlying constructs, that is, reading and math achievement. However, cautions are in order, since content variation among different tests used in different datasets may erode the exchangeability of test gain score information for comparison and synthesis. All of the test gain scores are reported in standard deviation units (σ).

Because age-specific or grade-specific standard deviations are unlikely to have either the same size or the same meaning, this study uses a common standard deviation across ages or grades in each database. This pooled standard deviation is also obtained across all cohorts so that standardized gain scores can be compared among sequential cohorts that participated in taking the same tests. Obviously, measures cannot be made equitable simply by standardizing multiple waves of scores on different occasions to a common standard deviation (Singer & Willett, 2003). In this study, the original test measures already have interval and constant metric properties through appropriate equating. Therefore, it should be noted that standardization here is not done for equating purposes but for facilitating the interpretation of academic
growth in effect size terms as well as the comparison and synthesis of results from different sources of data that employed different achievement scales.

First, this study uses long-term trend NAEP (L-NAEP) reading and math achievement data at ages 9, 13, and 17: 1971–2004 in reading and 1973–2004 in math. This study also uses national NAEP (N-NAEP) reading and math achievement data at grades 4, 8, and 12: 1992–2007 in reading and 1990–2007 in math (see Appendix A online). The standard deviations of NAEP scores vary among age/grade levels. In order to compute standardized gain scores between different ages or grades for each cohort, the national average NAEP scale score in the lower age/grade was subtracted from the higher age/grade average score and then this difference score was divided by the pooled standard deviation from all ages/grades involved.

Second, this study uses a sequence of national longitudinal education datasets collected by the National Center for Education Statistics (NCES) that provide information on individual-level academic growth at the elementary and secondary education levels (see Appendix A for description of the databases). At the elementary school level, the Early Childhood Longitudinal Study–Kindergarten (ECLS-K) and Prospects are used. The ECLS-K, launched in 1998, follows the educational trajectory of students from kindergarten through grade 5. Standardized reading and math gain scores from the ECLS-K are compared with corresponding results from Prospects data collected in the early 1980s. The National Education Longitudinal Study (NELS:88) involves tracking of high school students’ achievement gains (1988–2002). Standardized reading and math gain scores from the NELS:88 are compared with results from its predecessor, the High School and Beyond (HS&B) 1980–82 Study, and also from its successor, the Educational Longitudinal Study (ELS) 2002–04. The National Longitudinal Study of the High School Class of 1972 (NLS-72) is not included in this study because the data include only high school seniors and there were no repeated measures of achievement during their high school period.

Third, this study utilizes well-established national norms from standardized achievement tests—Comprehensive Test of Basic Skills (CTBS) and TerraNova (TN) as developed and normed by CTB/McGraw-Hill (see Appendix A online). TN succeeds CTBS and there is continuity among successive editions of CTBS and TN; these national norms were developed from stratified nationally representative standardization samples of grades 1–12 student populations during 1981, 1988, 1996, and 2000. Despite the limitations of drawing inferences about academic growth from cross-sectional data, these existing national norms from test publishers such as CTBS-TN can provide useful reference points because not only have the tests been widely used in many school districts across the nation, but their norms from nationally representative norming samples cover every grade (unlike NAEP, which covers only three selected grades). Among off-the-shelf standardized achievement tests with available national norms data, CTBS-TN was chosen because it provided
comparable test results from 1981 through 2000. The test equating procedures (i.e., equipercentile linking and IRT common-item equating methods) were designed to provide comparable national average reading and math scores between sequential cohorts at each grade over the past two decades.

Analytic Models and Procedures

A major problem in developmental research is the confounding of age, cohort, and time, which may occur in either cross-sectional or longitudinal designs (Schaie, 1965). In cross-sectional designs, observations are taken on different cohorts but at the same point in time. Consequently, age differences that emerge from cross-sectional data are confounded with cohort differences. This problem may occur with the use of cross-sectional datasets such as CTBS-TN to estimate academic growth trajectories. On the other hand, longitudinal studies may address questions about changes in a given sample or population when measurements are taken on at least two occasions. Given that only one cohort is measured, however, the investigator must consider the possibility that longitudinal change might be a product of some factor unique to the cohort. Another issue with longitudinal studies is the separation of developmental change with time-of-measurement effects such as those owing to changes in instrumentation. This problem applies to the use of longitudinal datasets such as ECLS-K and NELS:88 or repeated cross-sectional datasets such as NAEP.

In order to reduce the respective problems inherent in cross-sectional and longitudinal designs, this study utilizes a cohort-sequential design that includes both elements in combination with a time-lag component (see Schaie, 1965). The time-lag feature facilitates comparisons of different cohorts measured at different times. Each cohort serves as a comparison group for the following cohort. The L-NAEP testing schedule with a 4-year time interval allows for cross-cohort comparison of students’ academic growth from age 9 to 13 or from age 13 to 17. The same design is also applicable to the N-NAEP for cross-cohort comparison of students’ academic growth from grades 4 to 8 and from grades 8 to 12.

Following the design of both the long-term trend NAEP and the national NAEP assessments, we divide academic growth into three stages of education by ages/grades: (1) preschool or primary school level: growth up to age 9 (L-NAEP) or grade 4 (N-NAEP), (2) middle school level: age 9 to age 13 (L-NAEP) or grades 4 to 8 (N-NAEP), and (3) high school level: age 13 to age 17 (L-NAEP) or grades 8 to 12 (N-NAEP). However, child performance younger than age 9 or grade 4 is not available in NAEP and thus the gain score in the preschool or primary school level cannot be estimated. There are 10 sequential cohort groups available in L-NAEP and four sequential cohort groups in N-NAEP, each with at least two repeated measures of reading and math achievement at 4-year time intervals. All national average
scores were standardized by dividing them by the pooled standard deviation from all cohorts included in the study; cohort-specific standard deviation was not preferable for the sake of comparability, since different cohorts at different time periods tend to have slightly different test score variability. Also, 95% confidence intervals (CI) were calculated for each cohort’s standardized gain scores and the statistical significance of the differences between cohorts in their standardized gain scores was tested.

This study employed a weighted least squares (WLS) time-series regression method to identify the national average trajectory of students’ academic growth in reading and math across multiple cohorts in each dataset. Weights were computed by taking the inverse of the standard errors for national average scores. Building upon prior research as well as preliminary analyses that show curvilinear growth patterns, a quadratic growth model was fit to explain national aggregate growth patterns. The model used with N-NAEP and CTBS-TN data includes both linear and quadratic terms of grade for the outcome variable $Y$ for year $t$ and cohort $i$: $Y_{ti} = \pi_{0i} + \pi_{1i}[\text{grade}]_{ti} + \pi_{2i}[\text{grade}]^2_{ti} + e_{ti}$. In this model, the grade variable is centered at eighth grade, which is hypothesized to be a major turning point in the growth trajectory due to the typical transition from elementary to secondary education. Therefore, the linear parameter associated with [grade] represents the instantaneous rate of change at grade = 8, whereas the curvature parameter associated with [grade]$^2$ describes the changing rate of growth over time. Likewise, the model fitted with L-NAEP is also a quadratic growth model in which grade is replaced with age (i.e., growth from age 9 through age 17 as opposed to grade 4 through grade 12). In this model, the linear parameter associated with the [age] predictor represents the rate of change at age = 13 (value of age chosen for centering). The curvature parameter associated with the [age]$^2$ predictor describes the changing rate of growth over time.

Further, this study also employed two-rate linear growth modeling through hierarchical linear modeling (HLM) in order to differentiate growth rates between middle and high school age levels and to examine between-cohort variations in those growth rates (see Raudenbush & Bryk, 2002). The NAEP data are not longitudinal but quasi-longitudinal in the sense that we cannot match individual students but can match cohort groups at the aggregated group level. In particular, this approach permits one to synthetically estimate cross-cohort achievement gains over the period of schooling and use multi-level models that estimate parameters of individual cohorts’ growth trajectories including intercepts and slopes. Two temporal predictors, one for middle school age growth (age 9–13) and another for high school age growth (age 13–17), were used for modeling national aggregate cohort-by-cohort reading and math achievement trajectories ($N = 23$ measures in each subject at level 1 and $N = 10$ sequential cohort groups at level 2) in L-NAEP. This HLM analysis was precision weighted, using the inverse of standard errors as weights for national average reading and math standardized scores.
At level 1 (time level), there are three parameters in the model—(1) status at age 9 ($p_{0i}$), (2) growth rate between age 9 and age 13 ($p_{1i}$), and (3) growth rate between age 13 and age 17 ($p_{2i}$)—to track each cohort group $i$'s average reading or math achievement standardized score $Y$ at age $t$:

$$Y_{ti} = p_{0i} + p_{1i}[\text{middle}]_{ti} + p_{2i}[\text{high}]_{ti} + e_{ti},$$

where [middle] is coded 0 for age 9, 4 for age 13, and 4 for age 17, and [high] is coded 0 for age 9, 0 for age 13, and 4 for age 17.

At level 2 (cohort level), the level-1 intercept and slope coefficients for each of 10 cohort groups are modeled to change as a linear function of their birth year:

$$p_{0i} = \beta_{00} + \beta_{11} [\text{cohort birth year}]_{i} + \mu_{0i}$$

$$p_{1i} = \beta_{10} + \beta_{11} [\text{cohort birth year}]_{i}$$

$$p_{2i} = \beta_{20} + \beta_{21} [\text{cohort birth year}]_{i}$$

In theory, the above multilevel modeling used for cross-cohort comparison of age-based academic growth patterns in long-term NAEP could also be applied to a similar cross-cohort comparison of grade-based growth patterns in the national NAEP and CTBS-TN. However, there were insufficient numbers of cohort groups available for modeling cross-cohort changes in both the N-NAEP and CTBS-TN. Instead, paired comparison of successive cohort groups was conducted through a series of $t$ tests to test for the significance of paired-cohort differences in their average reading and math achievement gains. The small number of currently available cohorts and time points, particularly in the case of N-NAEP, does not provide a robust examination of the trend for tracking changes in achievement gains over time through regression techniques.

Threats to Validity and Ameliorative Strategies

There are potential threats to the validity of investigating the nature and degree of national average academic growth through vertical scaling across ages/grades and drawing inferences about changes in the national growth curve through cross-cohort comparisons. These threats are basically related to issues of measurement and sampling. In this section, some threats to the internal validity of this study and ameliorative strategies are discussed.

First, one of the major methodological issues in growth curve analysis is measurement equivalence (Baltes, Reese, & Nesselroade, 1977; Bergman, Eklund, & Magnusson, 1991; Haertel, 1991; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Petersen, Kolen, & Hoover, 1989; Singer & Willett, 2003). A fundamental assumption of this study is that there is sufficient continuity of both curriculum and assessment across grades K–12 in reading and
math to warrant vertical scaling for a common scale in each subject. NAEP, Prospects, ECLS-K, HS&B, NELS:88, ELS, and CTBS-TN all use IRT-based vertical equating methods for scaling. For example, NAEP has the tradition of using cross-age/grade vertical scaling. This is based on the conception of curriculum as a single, linear progression from low-level skills through higher level applications and also the conception of scale anchor points in terms of contents and processes common to broad ranges of age/grade levels. This vertical scaling facilitated interpretations of scores in terms of age/grade equivalents and afforded comparisons of growth rates. Although the tenability of the psychometric assumptions underlying cross-age/grade vertical scaling has been questioned (Haertel, 1991), this potential problem may be lessened when equating is done between adjacent grades, such as in longitudinal datasets or CTBS-TN test norms. IRT-based vertical equating procedures address this problem by having sufficient test overlap between adjacent test levels and by having students from multiple grades take the tests as a combined norming sample. The longitudinal studies such as NELS that have an explicit goal of measuring change or developmental growth also address the problem through appropriate test design and equating; the emphasis on general conceptual and/or problem-solving skills rather than isolated facts would minimize distortions in the gain scores due to forgetfulness (NCES, 1995).

In addition to cross-age or cross-grade equivalence within each assessment, this study also assumes the equivalence of reading and math achievement measures among different assessments. Fortunately, there was general convergence of measurement goals and target constructs across the studies. ECLS–K explicitly drew on the NAEP content and process frameworks for reading and math assessments with some modifications for relevance to early grades (NCES, 2002). The NAEP assessment goals are similar to those of the ECLS–K and NELS in that both projects aim to assess general knowledge and cognitive skills that schools typically emphasize in each subject. The availability of national standards also facilitated sharing of a common framework for test specifications. In particular, the mathematics frameworks for NAEP, ECLS-K, and TN are based on common curriculum standards from the National Council of Teachers of Mathematics (1989). Despite such common emphasis on general knowledge and problem-solving skills, there are differences between some assessments. For example, L-NAEP math focuses more on traditional, basic procedural skills, whereas N-NAEP math is designed to give more up-to-date student achievement information based on challenging curriculum standards. As acknowledged in the ECLS-K psychometric report (NCES, 2002), identifying typical curriculum objectives in the American education system is difficult because of the decentralized control and constantly evolving nature of curricula. This study is predicated on the existence of a general overlap in the constructs that would allow for
comparisons between the studies but at the same time recognizes unique attributes of each study.

Second, another major threat arises when we attempt to draw inferences about schooling effects based simply on the changes in academic growth patterns through within- and between-cohort tracking of achievement. Under the quasi-longitudinal NAEP sample design, changes in sampling processes between two assessment rounds (e.g., grade 4 and grade 8) may result in unexpected demographic changes for the same cohorts. Alternatively, real changes in cohort composition may result from migration during the 4-year time interval, although students and their families are much less likely to move in and out of the nation than their states or districts. A sensitivity analysis was conducted in order to check demographic changes (race/ethnicity, poverty, English proficiency, and disability status) and their influences on N-NAEP test score gains. The results showed very small changes with statistically nonsignificant differences for all of the matched cohort gain score estimates. On the other hand, between-cohort differences remain an issue since different cohort groups could have gone through different family and social changes. This study applied a statistical decomposition method to address this threat by sorting out the effects of demographic changes on N-NAEP achievement trends from the effects of other unmeasured variables or structural changes; these results will be discussed later when explaining potential confounders of the achievement trends observed (see Appendix B in the online version of this journal).

Cross-cohort comparison of academic growth patterns between different national longitudinal datasets poses more difficulties in sorting out true cohort differences from other confounding factors. Differences may come from many factors including student sample eligibility, sampling rates, sample sizes, design effects, and test content and design. At the elementary school level, differences in test design, as well as sampling, may affect the comparability of results between the Prospects and ECLS-K data sets. For the high school-level datasets, threats related to the comparison of HS&B, NELS:88, and ELS:2002 may be less serious because these high school data sets were all designed by NCES; full information on their similarities and differences are available for data users in the NCES technical report (see Ingels, Pratt, Rogers, Siegel, & Stutts, 2005).

Results

National Average Reading and Math Achievement Growth Trajectories

Figures 1 and 2 show national aggregate reading and math achievement growth trajectories among 10 sequential cohort groups based on the long-term trend NAEP national average age 9-13-17 scores. All scores were standardized with pooled standard deviations and centered to the average scores of baseline year 1971 for reading and of baseline year 1973 for math. WLS
time-series regression analyses of these standardized gain scores show that the quadratic growth model fits the data well in both reading and math. The estimated growth model parameters and model fit statistics based on pooled data across all L-NAEP cohort groups are summarized in Table 1. The estimate of the average age 13 achievement score in reading and math was 1.25 \( \sigma \) \((p < .001)\) and 1.64 \( \sigma \) \((p < .001)\), respectively; accumulated growth between age 9 and age 13 was about 1.5 standard deviations. The estimates of average growth rate at age 13 in reading and math were about .24 \( \sigma \) \((p < .001)\) and about .29 \( \sigma \) \((p < .001)\), respectively; students at age 13 gain about a quarter to one third of a standard deviation per year. The estimates of curvature in the growth curve were –.01 \( \sigma \) \((p < .001)\) in both reading and math; growth rates decelerate by 1\% of a standard deviation per year.

Figure 1. Long-term trend NAEP cross-cohort national average reading achievement growth trajectories in pooled standard deviation units at ages 9, 13, and 17 (data points marked by cohort birth years).
model fit statistics based on pooled data across all N-NAEP cohorts are summarized in Table 1. The estimates of the average grade 8 achievement scores (accumulated gains between grade 4 and grade 8) in reading and math were 1.31 \( \sigma \) \((p < .001)\) and 1.61 \( \sigma \) \((p < .001)\), respectively. The estimates of average growth rate at grade 8 in reading and math were .24 \( \sigma \) \((p < .001)\) and .26 \( \sigma \) \((p < .001)\), respectively. The estimate of curvature in the growth curve was –.02 \( \sigma \) \((p < .001)\).

Figures 3 and 4 show national aggregate reading and math achievement growth trajectories based on the CTBS-TN grades 1–12 national norms from four sequential standardization samples. As with NAEP data, a time-series regression analysis of the CTBS-TN data shows curvilinear growth patterns in both reading and math. The estimated growth model parameters and model fit statistics based on pooled data across all four editions are summarized in Table 1. The estimates of average grade 8 achievement in reading and math were 2.72 \( \sigma \) \((p < .001)\) and 4.18 \( \sigma \) \((p < .001)\), respectively; accumulated growth from grade 1 through grade 8 (grade 1 was the lowest grade available

---

*Figure 2. Long-term trend NAEP cross-cohort national average math achievement growth trajectories in pooled standard deviation units at ages 9, 13, and 17 (data points marked by cohort birth years).*

---

813
in these data) was 3 to 4 standard deviations. The estimates of the average growth rate at grade 8 in reading and math were \(0.25 \sigma(p < 0.001)\) and \(0.31 \sigma(p < 0.001)\), respectively; students at grade 8 gain about a quarter to one third of a standard deviation of achievement per year. The estimates of the quadratic terms that capture curvature in the reading and math growth curves were \(-0.02 \sigma(p < 0.001)\) and \(-0.04 \sigma(p < 0.001)\), respectively; the growth rates decelerate by 2% and 4% of a standard deviation per year squared.

Several patterns emerge from the synthesis of national aggregate reading and math achievement growth trajectories across L-NAEP, N-NAEP, and CTBS-TN. First, there is a curvilinear growth pattern with an incremental deceleration of growth rate over the entire period of schooling. Growth in the preschool or primary school level is faster than growth at the middle school level, which in turn outpaces growth at the high school level. The national average math achievement gain per grade is about one \(\sigma\) in the lower grades (1–4), about a half \(\sigma\) in the middle grades (5–8), and about a quarter \(\sigma\) in the higher grades (9–12). It appears that the learning gain reduces by about half from one schooling level to the next. Second, academic growth tends to be relatively faster in math than in reading, at least during the period of schooling. Total academic growth from K through grade 12

<table>
<thead>
<tr>
<th>Data</th>
<th>Reading</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-NAEP</td>
<td>(\hat{Y} = 1.25 + .24 (\text{age}) - .01 (\text{age})^2) (N = 23) (2 or 3 repeated measures per cohort from 10 cohorts at ages 9, 13, 17) (R^2 = .995, F(2, 20) = 2278.014, p &lt; .001)</td>
<td>(\hat{Y} = 1.64 + .29 (\text{age}) - .01 (\text{age})^2) (N = 23) (2 or 3 repeated measures per cohort from 10 cohorts at ages 9, 13, 17) (R^2 = 0.771, F(2, 20) = 367.086, p &lt; .001)</td>
</tr>
<tr>
<td>N-NAEP</td>
<td>(\hat{Y} = 1.31 + .24 (\text{grade}) - .02 (\text{grade})^2) (N = 9) (2 or 3 repeated measures per cohort from 4 cohorts at grades 4, 8, 12) (R^2 = .996, F(2, 6) = 953.392, p &lt; .001)</td>
<td>(\hat{Y} = 1.61 + .26 (\text{grade}) - .02 (\text{grade})^2) (N = 9) (2 or 3 repeated measures per cohort from 4 cohorts at grades 4, 8, 12) (R^2 = .949, F(2, 6) = 75.675, p &lt; .001)</td>
</tr>
<tr>
<td>CTBS-TN</td>
<td>(\hat{Y} = 2.72 + .22 (\text{grade}) - .02 (\text{grade})^2) (N = 48) (12 measures per edition from 4 serial editions at grades 1–12) (R^2 = .960, F(2, 45) = 570.417, p &lt; .001)</td>
<td>(\hat{Y} = 4.18 + .27 (\text{grade}) - .04 (\text{grade})^2) (N = 48) (12 measures per edition from 4 serial editions at grades 1–12) (R^2 = .977, F(2, 45) = 983.003, p &lt; .001)</td>
</tr>
</tbody>
</table>
is estimated to be in the range of 5 to 6 standard deviations in reading and 6 to 7 standard deviations in math. While the overall pace of growth differs between math and reading, both subjects follow the pattern of diminishing rate of growth with increased age or grade levels.

Cross-Cohort Changes in Academic Growth Trajectories

Have these national aggregate patterns of growth changed over the past three decades? HLM analysis of the L-NAEP national aggregate data suggests significant cross-cohort changes in both reading and math achievement growth trajectories (see Table 2). In reading, cohort birth year as a level-2 variable turned out to be a significant predictor of the cross-cohort variations in some parts of the growth curve: positive change for age 9 status ($\beta_{01} = .004, p < .05$), no change for age 9–13 growth rate ($\beta_{11} = -.0004, p = .25$), and negative change for age 13–17 growth rate ($\beta_{21} = -.0008, p < .05$). These results suggest that a significant gain in reading achievement did occur until age 13, while the growth stalled between age 9 and 13 and slowed down between age 13 and age 17 over the period of 1971 to 2008. In math, cohort birth year was a significant predictor of cross-cohort differences in the growth curves as a whole: positive change for age 9 status ($\beta_{01} = .025, p < .001$), and negative change for age 9–13 growth ($\beta_{11} = -.002, p < .001$) and age 13–17 growth ($\beta_{21} = -.0020, p < .01$). Over the period of 1973 to 2008, 9-year-olds’ math achievement increased by about 0.7 of
a standard deviation on average (.02 \sigma \times 35 \text{ years} = .7 \sigma). At the same

time, math achievement growth rates decreased slightly between ages 9 and 13 as well as ages 13 and 17 (–.002 \sigma \times 35 \text{ years} = –.07 \sigma). The results suggest that there was a significant upgrading of math achievement gains in the pre/early primary school years, while this earlier stage of progress was counterbalanced by indiscernibly slow deterioration of later growth during the middle and high school years.

Table 2
Two-Rate Linear Growth Models of National Cross-Cohort Reading and Math Achievement Trajectories: Hierarchical Linear Model Analyses of Long-Term Trend NAEP Data

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (age 9 achievement)</td>
<td>.099**</td>
<td>.240**</td>
</tr>
<tr>
<td>Cohort birth year effect on intercept</td>
<td>.004*</td>
<td>.024***</td>
</tr>
<tr>
<td>Slope for [middle] (age 9–13 growth rate)</td>
<td>.287***</td>
<td>.351***</td>
</tr>
<tr>
<td>Cohort birth year effect on [middle] slope</td>
<td>–.000</td>
<td>–.002***</td>
</tr>
<tr>
<td>Slope for [high] (age 13–17 growth rate)</td>
<td>.192***</td>
<td>.253***</td>
</tr>
<tr>
<td>Cohort birth year effect on [high] slope</td>
<td>–.001*</td>
<td>–.002**</td>
</tr>
</tbody>
</table>

Note. *p < .05. **p < .01. ***p < .001.
In brief, the national average reading and math achievement growth patterns appear to have changed significantly between the older and younger cohorts of students over the past three decades or so. The results suggest that national progress in the growth of students’ reading and math achievement, as an indicator of a value-added effect of schooling, was uneven at different stages of development or levels of schooling. In the following sections, this study takes an in-depth look at the trends separately at each age or grade level.

1. Gaining ground in pre/early primary school-level academic growth. The L-NAEP data for age 9 show that there was approximately a .28 σ reading gain and approximately a .65 σ math gain for 9-year-olds over the past 35 years or so. One way to understand the magnitude of this gain is to assess the gain in light of the pace of academic development at this age level. Since an average student’s reading or math gain per year is approximately 1 standard deviation during these primary school years, the long-term progress means that today’s 9-year-olds perform about 3 months ahead of their counterparts of the same age three decades ago in reading and about 7 months ahead in math.

In reading, there was modest progress in the 1970s, which was followed by a decade of decline and then a slow movement of recovery since the late 1990s. The 9-year-olds’ reading achievement on the L-NAEP recorded a –.15 σ loss during 1980–1990 and a .25 σ gain between 1990 and 2004. In math, there was progress at two time periods with some intermittent stagnation; one spurt occurred in the 1980s and then there was another increase after the late 1990s. The 9-year-olds’ math achievement on the L-NAEP recorded a .32 σ gain during 1982–1992 and a .33 σ gain between 1992 and 2004.

The fourth graders’ performance trend on the N-NAEP assessment during 1990–2007 shows a .14 σ reading gain and a .87 σ math gain. As with the L-NAEP data, much greater gain is recorded in math than in reading. At the same time, N-NAEP data appear to record much faster growth than the L-NAEP in math during the same time period. This may be because N-NAEP is better aligned with current national curriculum standards and thus is better positioned to capture more recent academic development. However, a different pattern is found in reading where both tests record only modest gains.

Despite the overall significance of NAEP achievement gains at age 9 or grade 4, it is unknown how much of this change occurred before or after children’s school entry. In the absence of preschool achievement trend data, it is not possible to disentangle the cumulated academic growth up to age 9 or grade 4 into the periods before versus after school entry. However, some insight can be gained through the comparison of national early childhood longitudinal datasets.

Comparison of ECLS-K data with Prospects data shows some progress in early elementary school-level academic growth in the 1990s. There is approximately a 10-year gap between the Prospects and ECLS-K cohorts; Prospects
used the CTBS to measure growth from grades 1 to 3 (1992–1994) and from grades 3 to 5 (1991–1993), whereas ECLS-K used its own standardized test to measure growth from grades 1 to 3 (2000–2002) and then from grades 3 to 5 (2002–2004). There was a statistically significant but modest degree of change in the size of gains. For reading, the grade 1 to 3 standardized gain increased from 1.63 $\sigma$ (95% CI = 1.60–1.66) to 2.16 $\sigma$ (95% CI = 2.11–2.2) and the grade 3 to 5 standardized gain also increased from .54 $\sigma$ (95% CI = .51–.56) to .94 $\sigma$ (95% CI = .89–.99). For math, the grade 1 to 3 standardized gain decreased from 2.26 $\sigma$ (95% CI = 2.23–2.29) to 1.92 $\sigma$ (95% CI = 1.87–1.96) and the grade 3 to 5 standardized gain increased from .87 $\sigma$ (95% CI = .84–.90) to 1.03 $\sigma$ (95% CI = .98–1.07). By and large, the patterns suggest that at least some part of the national progress as seen on the NAEP reading and math assessments for either age 9 or grade 4 students would have occurred after students’ school entry.

In addition, the analysis of cross-cohort national achievement norms from CTBS/3 through TerraNova/2 (1981–2000) shows a .18 $\sigma$ gain in reading and a .35 $\sigma$ gain in math for grade 1–4 students during the past two decades (see Figures 3 and 4). While the gains were evident at these early grades, it is not clear whether they occurred primarily before or after school entry. Nevertheless, the size of these .18 and .35 $\sigma$ gains is largely consistent with the L-NAEP age 9 gains for reading and math during the same time period; these findings are consistent with the results obtained from the long-term NAEP data that preschool or primary school-level achievement gains were greater in math than in reading.

2. Holding ground in middle school-level academic growth. In the L-NAEP data, there was either no or a relatively small change in the middle school age cohort’s gains from age 9 to age 13 (see Figure 5). In reading, the standardized gain was 1.14 $\sigma$ (95% CI = 1.08–1.21) in 1984–1988 and 1.14 $\sigma$ (95% CI = 1.08–1.21) in 2004–2008. In math, the standardized gain was 1.52 $\sigma$ (95% CI = 1.42–1.62) in 1992–1996 and 1.2 $\sigma$ (95% CI = 1.13–1.27) in 2004–2008. Similar patterns of stability are observed in the national NAEP assessment, which shows statistically insignificant change in the middle school cohort’s achievement gain from grade 4 to grade 8 (see Figure 6): 1.31 $\sigma$ (95% CI = 1.24–1.38) in 1994–1998 to 1.25 $\sigma$ (95% CI = 1.23–1.27) in 2003–2007 for reading; and 1.53 $\sigma$ (95% CI = 1.45–1.60) in 1992–1996 to 1.43 $\sigma$ (95% CI = 1.40–1.45) in 2003–2007 for math. Although both L-NAEP and N-NAEP show declines in recent years, these changes seem too small to have current implications for educational policy and practice.

Currently, there are no national longitudinal datasets available for tracking cohort-sequential changes in middle school-level achievement gains over the long run. ECLS-K tracks reading and math gains from grade 5 to grade 8, but there is no predecessor that allows for comparison of academic growth in the same grade span. Instead, the cross-cohort analysis of national achievement norms from CTBS/3 through TerraNova/2 (1981–2000) shows a .19 $\sigma$ gain
in reading and a .40 $\sigma$ gain in math for grades 5 to 8 during the past two decades (see Figures 3 and 4). The size of these middle school-level gains are very close to those of the corresponding primary school-level gains reported in the previous section (.18 $\sigma$ and .35 $\sigma$). This equivalence also suggests that the early elementary grade achievement gains have been sustained but not enhanced at the middle school level, particularly in reading.

3. Losing ground in high school-level academic growth. In the long-term trend NAEP data, there was a significant reduction of the high school-age cohort’s gains from age 13 to age 17 (see Figure 5). In reading, the standardized
gain was .91 \(\sigma\) (95% CI = .85–.97) in 1984–1988, and this all-time record gain dropped to .64 \(\sigma\) (95% CI = .56–.71) in 2000–2004. In math, the greatest standardized gain was about 1.18 \(\sigma\) (95% CI = 1.09–1.26) in 1990–1994, and then it dropped gradually to .83 \(\sigma\) (95% CI = .76–.90) in 2004–2008. A similar pattern of declines is observed by the national NAEP assessment, which also shows a significant drop in the high school cohort's achievement gain from grade 8 to grade 12 (see Figure 6): .83 \(\sigma\) (95% CI = .77–.88) in 1994–1998 to .64 \(\sigma\) (95% CI = .58–.69) in 1998–2002 for reading; and 1.06 \(\sigma\) (95% CI = .98–1.13) in 1992–1996 to .82 \(\sigma\) (95% CI = .74–.90) in 1996–2000 for math. Initially, these differences appear small, since the change in effect size is approximately .2 to .3 in standard deviation units. In considering their true practical import, however, these changes must be examined in context. Considering the fact that high school students grow in their reading and math achievement about a quarter of 1 standard deviation (.25 \(\sigma\)) per year on average, the magnitude of these changes amounts to the loss of up to 1 year’s worth of schooling and thus signifies a substantial slowdown of high school students’ academic growth relative to past national norms. This trend seems to be equally present in both subjects.

The stability and change of academic growth in reading and math during high school over the past two decades or so can be further investigated by using two additional data sources: (1) national longitudinal datasets and (2) CTBS-TN national norms. First, the high school cohort growth trend was examined by comparing results from successive rounds of national longitudinal achievement datasets at the high school level. The HS&B 10th-grade cohort’s gain during 1980–1982 shows a .36 \(\sigma\) gain (95% CI = .33–.39) in reading and a .19 \(\sigma\) gain (95% CI = .16–.22) in math from grade 10 to grade 12. Ten years later, the NELS:88 8th-grade cohort’s 1990–1992 test results show national average 10th- to 12th-grade gains of .22 \(\sigma\) in reading (95% CI = .17–.27) and .32 \(\sigma\) in math (95% CI = .27–.38). Further, the ELS 10th-grade cohort’s 2002–2004 test results show a .34 \(\sigma\) gain (95% CI = .30–.38) from grade 10 to grade 12 in math. By and large, the findings from these longitudinal assessments administered at different times suggest that math gains have not changed, whereas reading gains decreased.

Second, cross-cohort comparison of national high school students’ reading and math achievement norms from CTBS/3 through TerraNova/2 (1981–2000) show a .05 \(\sigma\) loss in reading and a .37 \(\sigma\) gain in math for grades 9 through 12 during the past two decades (see Figures 3 and 4). The comparison of these high school-level gains with corresponding middle school-level gains as reported in the previous section (.19 \(\sigma\) and .40 \(\sigma\)) reveals a substantial decline in reading but not in math. This supports the findings from the L-NAEP data that the elementary or middle grade achievement gains have not been sustained at the high school level, particularly in reading.

In addition, college entrance exam trends in the past 10 years do not show any measurable progress, mirroring the NAEP trends. The SAT score trends show slight losses in critical reading (down 3 points = –.03 \(\sigma\)) and slight gains...
Validation of Observed Changes in the Growth Trajectories

In this section, threats to the validity of observed changes in the national reading and math achievement trajectories at different age or grade levels are examined and discussed. Do the findings of this study reveal real changes or reflect statistical artifacts? Specifically, exploratory analyses are conducted with regard to some factors possibly confounding the observed changes, including ceiling effects, regression to the mean, demographic changes between past and current cohort groups, and the correlation between achievement and ability measures. Furthermore, if the changes are real, also examined is whether they are practically meaningful and significant to warrant policy discussion. Expectations based on the norms and standards of academic growth and discrepancies between expected and actual growth are then discussed.

First of all, there may be important ceiling effects or limits to achievement gains that could explain the observed trajectory. Prior research has shown no evidence of ceiling effects for NAEP, unlike some state assessments (Koedel & Betts, 2009; Linton & Kester, 2003). Indeed, it is very unlikely that ceiling effects would inhibit further growth for average high school students on the current NAEP scale, since the national average NAEP 12th-grade achievement has been below 300 throughout all rounds of NAEP assessment over the past decades on a scale of 0 to 500. For example, N-NAEP 2005 12th-grade national reading assessment results show an approximate range of 172 to 400 (mean ± 3 standard deviations = 286 ± 38) such that there was significantly large room for further growth even among relatively high-performing students on the N-NAEP scale. For NELS and ELS, special adaptive testing procedures were designed to minimize both floor and ceiling effects that typically distort gain scores (NCES, 1995).

Further, the study examined skewness in the test score distributions as an indicator of ceiling effects; a skewness value between ±1 is considered acceptable for most psychometric purposes. The analysis of 17-year-old data or 12th-grade data reveals slightly negative skewness (e.g., skewness = −.19 for reading and −.07 for math in L-NAEP age 17; −.39 for reading and −.20 for math in NELS 12th-grade; −.10 for math in ELS 12th grade). However, negative skewness is not unique to those high school samples but also present in some elementary school samples (e.g., skewness = −.22 for reading and −.15 for math in N-NAEP age 9; −.30 for reading and −.25 for math in ECLS-K grade 3). More importantly, the nature and extent of skewness among high school cohorts remains highly consistent over the study period between earlier and later NAEP as well as between NELS and...
ELS. These results suggest that at least the standardized reading and math achievement tests used in this study, with appropriate design and scaling, do not set age/grade-specific or time-bound limits on gain scores for students including high school students.\textsuperscript{14}

If ceiling effects prevail at the higher grade level, high school students may not be losing ground, especially if their growth trajectories are steeper than what we could expect at upper grades. The problem is that we currently do not have an unequivocal benchmark to determine whether the observed gain is greater or less than what we might expect. How much growth is adequate for different age or grade levels? One possible source of a growth benchmark may come from current NAEP performance standards. NAEP cut scores for the “Proficient” achievement level suggest desired progress from 238 at grade 4 to 281 at grade 8 (43 point gain) in reading and from 249 at grade 4 to 299 at grade 8 (50 point gain) in math. If one compares the actual NAEP 1996–2000 4th- to 8th-grade average gain scores (49 points in reading and 51 points in math) with desired proficiency gain scores, it appears that students meet derived standards of growth on average. NAEP cut scores for Proficiency also grow between grades 8 and 12, rising from 281 to 302 (22 point gain) in reading and from 299 to 336 (37 point gain) in math. Actual NAEP 1996–2000 8th- to 12th-grade national average gain scores (23 points in reading and 29 points in math) are similarly close to desired growth. However, considering the fact that the current NAEP national average 4th- and 8th-grade scores are substantially below the proficiency cut scores, the current pace of academic growth among both middle and high school students is actually not sufficient for staying on track to proficiency by the time of high school graduation.

Regarding the regression to the mean threat, there is no clear indication that the deceleration of the high school cohort’s growth rate is primarily due to the improvement of initial achievement status or the acceleration of prior growth rate. Because the cross-sectional nature of NAEP data precludes direct investigation of this problem, this study examined the patterns in national longitudinal data. ECLS-K was used to examine the relationship between 5th-grade test scores (as an indicator of middle school initial status) and 5th- to 8th-grade gain scores (as an indicator of middle school growth rate). NELS was also used to examine the relationship between 8th-grade scores (as an indicator of high school initial status) and 8th- to 12th-grade gain scores (as an indicator of high school growth rate). The correlations turned out to be generally modest in both NELS ($r = -.19$ for reading and $r = -.03$ for math) and ECLS-K ($r = -.23$ for reading and $r = -.44$ for math). Further, the correlation between fall kindergarten scores and K–8 gain scores in ECLS-K was tenuous ($r = -.004$ for reading and $r = .09$ for math).

Another potential threat to drawing inferences about student achievement gains from low-stakes tests such as NAEP would be the lack of motivation for students, particularly 12th graders, to do well on the tests (Brophy & Carole,
2005; Kiplinger & Linn, 1995/1996; O’Neil et al., 1992). However, if the lack of motivation for 12th graders has equally affected all high school cohorts over the entire period of L-NAEP or N-NAEP, it would only influence the estimation of high school growth rate relative to the lower grade level growth rates within each cohort, but it would not affect the estimation of cross-cohort changes in high school growth rates. Future research needs to address this problem by directly measuring and adjusting for the motivation factor in NAEP.

For validation of the N-NAEP cross-cohort achievement gains, this study attempted to address the effects of demographic changes in student cohort populations over the past two decades. During the pre-NCLB period (between the 1998 and 2002 cohorts for the N-NAEP reading and between the 1996 and 2000 cohorts for the N-NAEP math), there was a modest but significant increase in the estimated percentages of Hispanic and poor students among the NAEP 4th-grade and 8th-grade student populations. There were also increases in the percentages of English language learners (ELLs) and students with disabilities (SDs), but the changes were not large enough to be statistically significant. There was a significant increase in the percentage of poor students in the 12th-grade student population. Since all of these groups with population gains included relatively larger proportions of lower achieving students, the total effects of these demographic changes on reading and math achievement were somewhat negative. Table 3 summarizes the results of cross-cohort adjustment of the N-NAEP achievement gains for student populations’ demographic changes.

While demographic changes in the late 1990s and 2000s predicted negative achievement gains (i.e., drops) across all grades, the elementary school-level cohort groups (grades 4 and 8) gained more than predicted based on the demographic changes, whereas the high school-level cohort (grade 12) gained less than predicted. Table 3 shows adjusted gains, that is, the differences between the actual and predicted gains by grade; the adjusted gains of both 4th and 8th graders were positive, whereas the adjusted gains of 12th graders were negative. It appears that these trends emerged before the enactment of NCLB and continued afterward.

Lastly, it is noteworthy that American students’ academic achievement trends do not simply reflect concurrent changes in their cognitive ability or academic aptitude as measured by IQ tests. Flynn’s estimate of total American IQ gain was about .40 $\sigma$ for the past two decades (i.e., 3 points per decade on the WAIS). Flynn’s earlier analysis of IQ gains from 1947 to 1972 on the WISC, the WISC-R, and the WAIS for ages 7 to 17 (Flynn, 1984) and the most recent updated analysis of American IQ gains on the WISC from 1932 to 2002 (Flynn & Weiss, 2007) show that IQ gains occur across all school ages and the size of the gains does not vary significantly with age level. Lynn (1998) also noted that the IQ gains were observed very early, among preschool children, and thus claimed that it supports a nutrition hypothesis—better child nutrition improves their cognitive
ability—rather than competing hypotheses such as school effects or other environmental change effects (e.g., video games, radio, and TV). Flynn (1998) also noted that IQ gains occurred mostly on fluid IQ test measures (e.g., Raven’s type or performance tests) as opposed to crystallized IQ test measures (e.g., vocabulary, information, math); the former is supposed to be more impervious to the direct effects of schooling than is the latter. Therefore, both the trends of relatively even IQ test score gains across all ages and the lack of IQ gains on crystallized intelligence measures imply that IQ versus achievement test score trends may not share common causes.

**Discussion**

Successful education for human capital development demands continuous and adequate academic progress between all levels of schooling, from preschool to elementary, from elementary to high school, and from high school to college. Our understanding of American students’ academic growth can be improved with synthetic or sequential cohort analyses of multiple national

---

**Table 3**

National NAEP (N-NAEP) Reading and Math Achievement Gain Scores for Grades 4, 8, and 12 With and Without Adjustment for Demographic Changes Before and After NCLB

<table>
<thead>
<tr>
<th>Subject and Period</th>
<th>Grade</th>
<th>Actual</th>
<th>Predicted</th>
<th>Difference (actual – predicted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading—pre-NCLB (1998–2002)</td>
<td>4</td>
<td>3.4</td>
<td>-2.6</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.5</td>
<td>-2.1</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-3.2</td>
<td>-1.4</td>
<td>-1.7</td>
</tr>
<tr>
<td>Reading—post-NCLB (2003–2007)</td>
<td>4</td>
<td>2.9</td>
<td>-0.5</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.6</td>
<td>-1.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math—pre-NCLB (1996–2000)</td>
<td>4</td>
<td>1.6</td>
<td>-1.4</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.7</td>
<td>-1.1</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-1.8</td>
<td>-1.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>Math—post-NCLB (2003–2007)</td>
<td>4</td>
<td>5.0</td>
<td>-0.2</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.5</td>
<td>-0.9</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Predicted gain scores are obtained from the multiple regression of student achievement on key demographic factors (i.e., race/ethnicity, poverty, English proficiency, disability status). The measures of demographic changes between paired N-NAEP cohort groups are plugged into the estimated regression models to predict changes in student achievement. NA = 2003–2007 NAEP 12th-grade math gain scores are not available due to the incomparability of 2003 and 2007 scores. The 2007 math assessment was based on a new framework and thus results could not be placed on the old NAEP scale.
data sources collected over the past three decades or so, including national education assessments and test publisher norms that use developmental scales of child achievement in core subjects. Overall, the national trends of academic growth trajectories in reading and math show remarkable consistency and stability across different tests and cohorts over the long term. American students’ academic growth curves in schools are characterized by a substantial level of cumulative achievement gains, approximately 6 to 7 standard deviations from the beginning of kindergarten through the end of high school. The comparison of standardized reading and math gains clearly suggests that the pace of growth is faster in math than it is in reading, across the age/grade levels investigated, although the overall growth patterns characterized by decelerating rates of growth over ages/grades are very similar between the two subjects. In math, typical growth curves may be summarized as a tripartite pattern of achievement gains over the full course of schooling—1 standard deviation gain in the primary grades (K–4), a half standard deviation gain in the middle grades (5–8), and a quarter standard deviation gain in the high school grades (9–12).

The diminishing rate of academic growth means that longer time is needed to achieve the same amount of learning gain at the higher age/grade level, and it is attributable to the interaction of two underlying forces of academic growth, that is, human development and curriculum development. One stems from a decreasing rate of growth in child cognitive capacity for acquiring new knowledge and skills at the older ages, and the other stems from the increasing difficulty and complexity of school curricula and instruction at the higher grades. While the slower academic growth in the higher ages or grades is likely to be the norm in most countries and cultures, previous international comparisons of math achievement implied that American students experience relatively faster deterioration of growth in middle and high school in comparison with high-achieving industrial countries (Lee & Fish, 2008; NCES, 1996, 1997, 1998; Stevenson & Stigler, 1992). Why do American students lag further behind during their middle and/or high school years? Despite several decades of school reform efforts, no systematic research has yet been done to examine the nature and degree of changes in the U.S. national academic growth curves and to explore the effects exerted by social and educational policies, if any. This study calls for long-term cross-cohort comparison studies beyond conventional short-term single cohort studies.

Have there been any changes in the nature and extent of national academic growth trajectories in reading and math? This study finds that while the form of U.S. aggregate academic growth curves remains highly stable over the past three decades, there were relatively small but noteworthy shifts, particularly since the late 1990s. It is worth noting that not only did the faster growth occur “within” cohorts at the lower ages/grades but further acceleration of growth also occurred “between” cohorts at the lower ages/grades. A closer look into NAEP and other national test trends by students’ grade or age level reveals these complex and mixed patterns of changes.
The pace of preschool or primary school-level achievement gain tends to accelerate, whereas the pace of middle or high school-level achievement gain tends to remain stagnant or slow down, respectively. While it is not clear how much of the progress in the early child achievement gains occurred before or after school entry, it appears that an acceleration of the growth curve during the preschool or primary school period subsides during the later middle or high school periods. It appears that middle school-level students are caught in the middle as their academic growth pattern tends to follow a path of middle ground between the two opposite trends shown by students in pre/early primary school and high school.

One outstanding question is how much of the observed changes in the national academic growth curves are authentic. After considering potential major threats to validity such as ceiling effects and demographic changes, a tripartite pattern still emerges. All in all, the findings of this study suggest that American students may be gaining ground at the pre/primary school level, holding ground at the middle school level, and losing ground at the high school level. While the causes of these mixed trends may be multiple and complex, they cannot simply be dismissed as artifacts of the measurement and statistical approaches employed in this particular study. However, we need to watch closely how this emerging pattern may develop further and, at the same time, explore ways to not only sustain but also transfer primary school-level success to the middle and high school levels to boost their academic growth. Subsequent studies may examine social and educational policies that differentially influence academic growth curves at different levels of schooling. If the relationship between changes in student outcomes and changes in policy environments were reciprocal, then the challenge would be to distinguish the causes and effects of observed trends.

It needs to be noted that if one looks only at high school achievement trends based on the 12th-grade results, one would get the impression that the current generation of high school students performs as well as the earlier generation 30 years ago. On the other hand, if one examines the entire growth pattern from elementary through high school, one will realize that American high schools take academically better-prepared students than the earlier generation but fail to help them flourish much further. It is crucial to understand not only how well high schools perform but also their performance in the context of the other levels of education: the elementary/middle schools that prepare students for secondary education and the postsecondary institutions for which most students in high school are preparing. Subsequent national assessment studies such as NAEP need to extend the conventional K–12 time frame to preschool and postsecondary levels of education and to monitor the development of core knowledge and skills that influence success or failure in transition across all levels of P–16 education. Further, it is desirable to develop common benchmarks that would add new meaning to achievement gain score analyses, such that the adequacy of...
growth could be determined on the basis of forward-looking curriculum standards in addition to past norms.

The findings of this study offer insights about the alterability of national academic growth curves over generations. It took approximately one generation to raise the pre/early primary school curve by a quarter of a standard deviation in reading and by a half of a standard deviation in math. In other words, the current generation of primary school students seems to learn faster (about a quarter of a school year in reading and a half of a school year in math) than the previous generation. However, this generational learning gap would diminish by the time the students graduate from high school. Although the greatest improvement in the national growth curve occurred at the early childhood stage, this study showed that those gains would not be sustained necessarily through the later stage of adolescence. It supports the conventional proposition that the environment would have its greatest effects during the period of most rapid academic growth, and conversely, as the developmental curve reaches asymptotes it becomes more resistant to change. However, the relatively slower academic growth at the higher ages/grades should not be interpreted as implying that we should invest more energy and resources in early childhood or elementary education than in secondary or postsecondary education. Early intensive preschool and primary school interventions should be accompanied by follow-through enhanced support at the middle and high school levels. Otherwise, the effect of an early school intervention is likely to fade out over time and the pace of early achievement gains becomes harder to maintain at the upper level of education. This study calls for national P–16 education policy and research efforts toward sustainable academic growth and seamless educational transition.

Notes

1On the TIMSS math assessment, American 4th graders performed above the international average, whereas the 8th and 12th graders scored below the international average (National Center for Education Statistics, 1996, 1997, 1998). Previous international research also has revealed that the mathematics achievement gap between Asian (e.g., Chinese, Japanese) and U.S. students exists before their school entry and that these gaps widen during schooling (Stevenson & Stigler, 1992; Uttal, Lummis, & Stevenson, 1988; Wang & Lin, 2005).

2For instance, a case study suggests that a majority of 8th-grade or 10th-grade students in a state who earned passing scores on their state English and math assessments for promotion to 9th grade or high school graduation were not ready for college-level work; college readiness was operationally defined by receiving a grade of C or higher in the corresponding first-year college courses such as English composition and algebra (ACT, 2005b). On the other hand, a longitudinal study documented a sharp decline in adolescents’ confidence and interest in learning mathematics when students were making the transition from elementary to junior high school (Eccles, Lord, & Midgley, 1991). The problem with college readiness should be viewed as an issue of sustainable academic growth and transition across all levels of schooling rather than an isolated high school problem per se.

3NCLB substantially strengthened the scope and intensity of test-driven external accountability provisions by testing all students in grades 3 through 8 and one high school grade annually, targeting all schools as opposed to Title 1 schools only, and imposing
more stringent requirements (e.g., meeting performance target for all subgroups) with real threats of punitive and corrective actions.

4The terms “K–16” or “P–16” describe a policy movement to strengthen educational achievement from preschool through completion of the college degree. For example, the 2005 National Education Summit on High Schools established a five-point *Action Agenda for Improving America’s High Schools* to build stronger P–16 systems, with a particular emphasis on strengthening high schools. More than half of the states currently have K–16 initiatives under way and most include policies and activities aimed at early outreach, student preparation for college, and improvement of teacher quality (see www.sheeo.org/k16/k16-home.htm). The policy was often framed as a response to the demands of the 21st-century economy requiring that even students who proceed directly to work after high school possess college-level skills (ACT, 2005a). However, this market-driven claim for “College for All” has been challenged with cautions against the tendency toward narrow vocationalism (Grubb & Lazerson, 2005).

5Specifically, the national or state average achievement gain on NAEP over the 1996–2000 period was obtained by comparing NAEP scores from the 1996 fourth-grade assessment with NAEP scores from the 2000 eighth-grade assessment. However, no systematic research has been done to track the full course of academic growth across schooling levels and to compare sequential cohort groups’ growth patterns over the long run. One of those studies examined cross-cohort differences in within-cohort achievement gains between the fourth and eighth grades and concluded that the degree of academic growth did not change between the 1970s and 1990s (Coley, 1998). However, the study had limitations in that it only examined gains during the middle school grades between two distant cohorts and did not take into account the influences of demographic and social changes during the interim period.

6Although children’s school grades do not correlate perfectly with their ages and the meaning of grades is complicated by local variations in grade configuration, students in the age range of 10 to 14 are usually in either an elementary school with K–8 grades or a middle school with a typical grade range of 5 to 8. Hence, they are referred to as middle school-level students.

7Although children may be able to take standardized achievement tests as early as age 3 to 5, it is difficult to obtain comparable NAEP baseline scores for younger children, as this assessment was not designed for use with this group. The NAEP scales do not have an absolute value of zero achievement. While there were attempts to develop achievement scales with an absolute zero, their applicability to NAEP is questionable. For example, Thurstone (1928) proposed a method to approximate an absolute zero point for the measurement of intelligence based on the linear relationship between the mean and standard deviation values by age, but this method is not appropriate for IRT-based NAEP scales, which show relatively stable variances across ages.

8Despite the limitation of small sample size, the study provides unbiased estimates. It has passed critical tests for violations of the assumptions of weighted least squares (WLS) regression and hierarchical linear modeling (HLM) analyses. For the independence assumption, Durbin-Watson statistic was used to test for correlated residuals and there was only a modest degree of serial correlation, falling within the region of ignorance. For the normality assumption, differences between expected and observed Mahalanobis distance measures for residuals were assessed, and the Q-Q plot was approximately linear without any evident outliers, suggesting that the assumption is largely tenable. For the homogeneity of variances assumption, chi-square tests supported the adequacy of that assumption.

9The question is what would have happened to gain scores if student demographics had remained the same across all cohorts? The estimated amount of changes to unstandardized gain scores between different cohorts were too small to be significant (on the NAEP 0–500 scale with approximate standard deviation of 35 points): .98 up for math gain between 1996 fourth-grade and 2000 eighth-grade cohorts; 1.39 down for math gain between 2003 fourth-grade and 2007 eighth-grade cohorts; .36 up for math gain between 1998 fourth-grade and 2002 eighth-grade cohorts; 1.20 up for reading gain between 2003 fourth-grade and 2007 eighth-grade cohorts. For example, there were very small demographic differences (e.g., 2% increase in Hispanics and 4% increase in Asians) between the 1996 fourth-grade and the 2000 eighth-grade samples such that
this demographic matching for cross-cohort comparison would not bring about significant change (with estimated change value of .98) to unadjusted math achievement gain.

\(^{10}\) The decomposition method was an adaptation of the Oaxaca-Blinder econometric method, which was originally designed for estimating the effects of discrimination on group differences (Blinder, 1973; Oaxaca, 1973).

\(^{11}\) Standardized scores in Figures 3 and 4 are group mean differences, in standard deviation units, relative to TerraNova 2nd edition (standardized in 2001) grade 1 average, which was chosen as the reference point and assigned the value of zero.

\(^{12}\) Data are for seniors who took the SAT any time during their high school years through March of their senior year. If a student took a test more than once, the most recent score was used. Possible scores on each part of the SAT range from 200 to 800 and the standard deviation is about 110. The critical reading section was formerly known as the verbal section. The College Board cautions that the SAT score drops in both reading and math 2 years in a row (2006 and 2007) may be an artifact of recent changes in the format of the exam and its possible influence on students' test-taking behaviors.

\(^{13}\) The size of these gains in standard deviation units is only about .02 \(\sigma\) (standard deviation of ACT scores is about 4.7 points).

\(^{14}\) This pattern is also supported by evidence from tests designed for the full range of ages in the entire life span. For example, Woodcock-Johnson achievement tests provide developmental evidence of continuing growth beyond the high school level. The results show that observed abilities increase in reading and math during the traditional years of schooling and tend to grow until they reach the plateau around ages 25 to 30 that go substantially beyond compulsory education age (McGrew & Woodcock, 2001).


831


Manuscript received August 18, 2008
Final revision received December 24, 2009
Accepted January 2, 2010