A Longitudinal Study of the Impact of America's Choice on Student Performance in Rochester, New York, 1998-2003

Henry May
Jonathan A. Supovitz
David Perda

July 2004

Contents

List of Tables and Figures

Executive Summary

Education is a cumulative process. Yet while students' knowledge and skills are built up over time, educational researchers are rarely afforded the opportunity to examine the effects of interventions over multiple years. This study of the America's Choice school reform design is just such an opportunity. Using 11 years of student performance data from Rochester, NY — which includes several years of data before America's Choice began working in the district — we examine the effects of America's Choice on student learning gains from 1998 to 2003. Employing a sophisticated statistical method called Bayesian hierarchical growth curve analysis with crossed random effects, we compare the longitudinal gains in test performance of students attending America's Choice schools to those of students attending other Rochester schools. Our analytical method allows us to examine student test performance over time, account for the nested structure of students within schools, and account for the very real problem of within-district student mobility.

Through these analyses, we sought to answer three central questions. First, is there evidence that America's Choice increases students' rates of learning and, if so, how big is the increase? Second, does America's Choice improve the performance of particularly low-achieving students? And third, does America's Choice make education more equitable for minority students?

Overall, we found that students in America's Choice schools gained significantly more than did students in other Rochester schools in both reading and mathematics test performance. These differences are moderate in the early-elementary grades (grades 1-3) and stronger in later grades (grades 4-8). In the early-elementary grades, students in America's Choice schools averaged three weeks of additional learning per year in comparison to students in other district schools. In grades 4-8, students in America's Choice schools averaged slightly more than two months of additional learning per year in comparison to students in other district schools.

Low-achieving students performed particularly well under the America's Choice regimen, without having any negative effect on students in the upper quartiles. In our analyses, we found that students in Rochester's America's Choice schools who started out in the lower quartiles learned more than did similar low-performing students in other Rochester schools. At the other end of the spectrum, students in the upper quartiles in America's Choice schools performed indistinguishably from similar students in other district schools.

Minority students in America's Choice schools — African Americans and particularly Hispanics — also learned more than their peers in other district schools. Conversely, the rates of learning for White students in America's Choice schools were no different from other White students from comparison schools. Both African American and Hispanic students who attended America's Choice schools also consistently out-gained White students, reducing the gaps in performance between White and minority students. Thus, while still considerable, the gaps in achievement between White and minority students were reduced for those students attending America's Choice schools.

Introduction

Rarely in educational research do we have access to data that allow us to empirically explore the impacts of programs on individual student learning over long periods of time. This element of CPRE's evaluation of the America's Choice comprehensive school reform design is one of those exceptional cases.

In this study, we use individual-level student performance data from Rochester, NY, linked across 11 years, to examine the longitudinal impact of America's Choice. The modern statistical methods used in these analyses are able to compare the individual performance of America's Choice students over multiple years to both the performance of similar students in other schools, as well as the performance of America's Choice students before attending an America's Choice school.

The research questions we sought to answer were:

- Is there evidence that America's Choice increases students' rates of learning? If so, how big is the increase?

- Does America's Choice help to improve the performance of especially low-achieving students? That is, does America's Choice help to "bring up the bottom?" If so, is this done at the expense of high-achieving students?

- Does America's Choice make education more equitable for minority students? That is, does America's Choice help close the race gap? If so, is this done at the expense of non-minority students?

The Context of Rochester Public Schools

Rochester — located in the northwestern part of New York on Lake Ontario, about one hour from the Canadian border — is a mid-sized urban district. The home of once corporate giants Kodak, Xerox, and Bausch and Lomb, Rochester has long been a thriving metropolis. The Rochester City School District currently educates just over 35,000 pre-K-12 students. In stark contrast to surrounding wealthy and suburban Monroe County, the city has become increasingly poor and minority. Over the past 25 years, the schools have increasingly educated high numbers of poor and minority students. Between 1980 and 2004, for example, the percentage of students on free and reduced-price lunch in the district increased from 22% to 81% today. Minority population in the city increased from 69% in 1990 to 85% today. Rochester ranks 11th nationally in per-capita child poverty, ahead of large urban districts such as New York City, Washington, D.C., Chicago, and Los Angeles. Rochester schools are projected to face enrollment declines of 30% by the end of the decade.

The district has a long history of educational innovation and partnerships. The Rochester teachers' union has long been considered one of the most innovative in the nation, pioneering such efforts as teacher mentor programs and initiatives to professionalize the teaching staff. The district has one of the most highly acclaimed pre-K systems in the nation

Table 1. Demographic Characteristics of Students (Grades 1-8) in Rochester by America's Choice Participation Status, 1998-2003

Student Characteristic	America's Choice Students	Other Rochester Students
Percent Female	49.8	49.4
Percent Male	50.2	50.6
Percent African American	71.7	58.9
Percent Hispanic	17.8	19.6
Percent White	8.9	19.1
Percent Other	1.6	2.4
Percent Special Education	16.6	15.3
Percent Limited English Proficient	87.9	85.5
Percent Receiving Free/Reduced-Price Lunch	95.2	89.2

and several professional development initiatives with local colleges and universities. Rochester also has long experimented with school reform models.

Manuel Rivera, the current superintendent, is the only person to have held the superintendency twice. He was superintendent of Rochester from 1992-1994, before becoming a vice president for Edison Schools. Rivera returned to the district in 2002. His second tenure has focused on reducing and aligning the array of the reforms that are in the district, developing and aligning accountability systems with these initiatives, and navigating the district through declining funding and enrollment.

America's Choice took root in Rochester between Rivera's two tenures. The program was introduced into Rochester in the fall of 1998 when six elementary schools began to implement the design. In the fall of 2000, one middle school also adopted America's Choice. An additional five elementary schools and four middle schools began to implement America's Choice in the fall of 2001. In total, about one-third of the elementary and middle schools in Rochester had experience with

America's Choice by the 2002-2003 school year. A total of 10,165 students attended an America's Choice school for at least one year.

Table 1 shows the demographics of the district for the years following the implementation of America's Choice. The America's Choice students represent a more disadvantaged student group when compared to their peers not receiving America's Choice. America's Choice students were more likely to be receiving free or reduced-priced lunch and were more likely to receive special education services or be classified as Limited English Proficient students. Schools using America's Choice also enrolled a greater proportion of minority students. Almost 90% of America's Choice students were either African American or Hispanic compared to 78.5% of students not receiving America's Choice.

Sample and Measures

The data for this study span 11 years, including all student test scores from the 1992-1993 school year to the 2002-2003 school year. During this period, over 55,000 students enrolled in grades 1-8 in 42 elementary and 10

Figure 1. Number of Years of Test Data Available for Rochester Elementary and Middle Grades Students
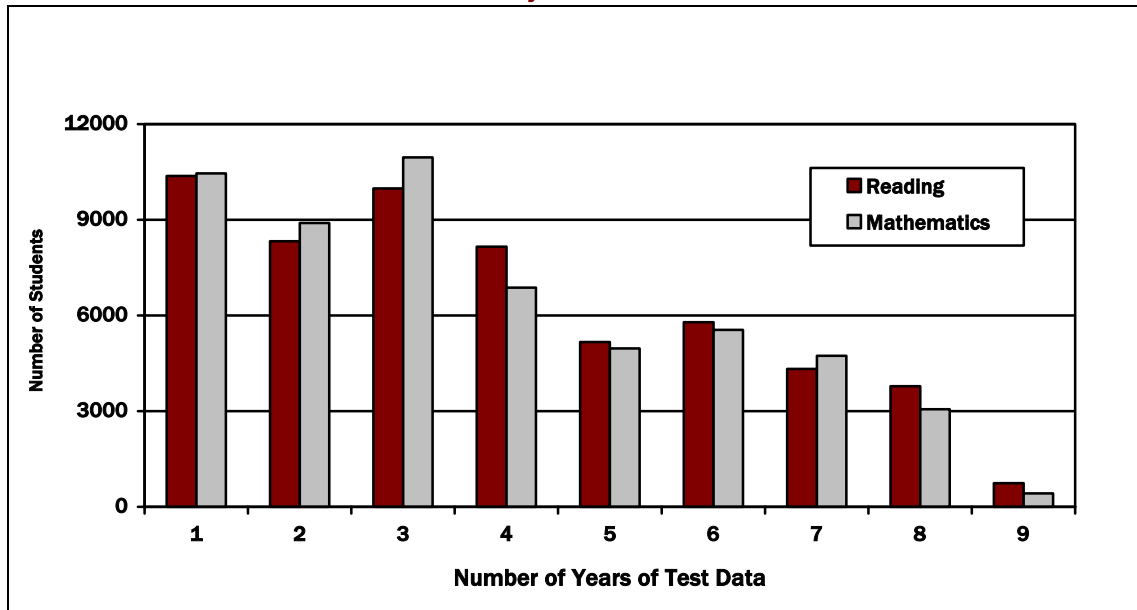
middle schools were tested in reading and mathematics. Because Rochester tested nearly every student in every grade, these data constitute a near census of the population of students enrolled in Rochester elementary and middle schools from 1993 to 2003.

Additional demographic data were also collected for each student, including age, race/ethnicity, poverty status, special education status, and limited English proficiency status. Inclusion of these demographic variables in statistical models allowed us to control for differences in the backgrounds of students in the district.

The demographic and annual test score data for each student were linked across years, so that achievement scores could be tracked over time for each individual student. This enabled us to estimate the amount of learning (i.e., as represented by increases in test scores) that occurred for each student during each year. However, as a result of

student mobility in and out of the district and matriculation beyond eighth grade, most students had less than eight years of data. Figure 1 shows the number of years of data available for the sample of 56,693 students in reading and 55,932 students in mathematics.[1] Approximately 50% of the students in this analysis had four or more years of data.

Rochester used multiple achievement tests during the 11-year span covered by this study. These included the Stanford Achievement Test (SAT-9), the California Achievement Test (CAT-5), the Degrees of Reading Power test (DRP), the New York State assessments (NYS), the New York Pupil Evaluation Program tests (PEP), and the New York Preliminary Competency Test (PCT). The SAT-9, CAT, and DRP are all nationally normed standardized

---

[1] The sample size is slightly lower for mathematics because the NYS grade 4 and grade 8 mathematics scores for 2003 were not yet available at the time of this analysis.

tests. The NYS, PEP, and PCT are New York State assessments. In Rochester, the SAT-9 was the most widely used test in recent years, being used in grades 1-3 and 5-7 from 1999-2000 to the present.

In order to track changes in individual student performance over time, the achievement scores from the different tests had to be placed on the same metric. The metric of the SAT-9 was used so that growth in student performance could be charted over time.[2] The process by which this is usually done is called *test equating*, which involves administering multiple tests to the same students at the same time and determining how the scores on one test match up to the scores from another test. Traditional equating methods were not possible in this context because tests were not administered simultaneously and there were no common items across the tests used in Rochester. Instead, the test scores were equated using a linear rescaling within each grade based upon the mean and standard deviation of the scores on the SAT-9 for each grade level during the 2002-2003 school year (see Appendix A for details). This type of equating falls into the category of *calibration* as defined by Linn (1993). This process imposes the assumption that test scores are similarly distributed over time and across tests, and that the only difference between the tests is the scale of the scores (e.g., a 100-point scale vs. a 500-point scale). In other words, it is assumed that a student would show similar performance relative to other students in the district, regardless of

which test was used to measure performance. This assumption is reasonable for tests that measure the same construct (e.g., reading comprehension) with similar types of items (e.g., multiple choice). Such is the case for the tests used in Rochester. It is important to note that equating was done separately for reading and mathematics scores. In other words, reading scores were not used to predict mathematics scores, or vice versa.

To evaluate the plausibility of the equating assumptions using empirical methods, we estimated correlations between students' scores as they progressed through the grades and compared the correlations from the years in which the tests changed to correlations from the following year. For example, if the correlation between a cohort of second graders' reading scores in 1999 (i.e., DRP scores) and their scores as third graders in 2000 (i.e., SAT-9 scores) was similar to the correlation between the next cohort of second graders' reading scores in 2000 (i.e., SAT-9 scores) and their scores as third graders in 2001 (i.e., SAT-9 scores), then there is evidence to suggest that the two tests have a similar linear relationship and that the DRP from 1999 can be substituted as a nearly equivalent pre-test measure when SAT-9 scores are unavailable. The results of our analyses of grade-to-grade correlations suggest remarkable consistency among the six tests in both reading and mathematics. Table 2 shows correlations between test scores from 1999 to 2001. The analysis is restricted to these years because this was the major transition period in which previously used tests were replaced with the SAT-9.

---

[2] The state assessments were scaled with equal mean scores at each grade level. Therefore, average growth using this metric would be represented by a gain of 0 points. Use of the SAT-9 metric allows comparison of program effects to typical rates of annual growth in test scores.

	Grade Levels						
	1→2	2→3	3→4	4→5	5→6	6→7	7→8
Reading							
1999 Lower-Grade Test	CAT-5	DRP	PEP	NYS	DRP	PEP	DRP
2000 Upper-Grade Test	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2000 Lower-Grade Test	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9
2001 Upper-Grade Test	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
1999-2000 Lower-Grade→ Upper-Grade Correlation	.61	.69	.70	.75	.76	.75	.71
2000-2001 Lower-Grade→ Upper-Grade Correlation	.69	.69	.71	.73	.79	.83	.75
Mathematics							
1999 Lower-Grade Test	n/a	CAT-5	PEP	NYS	CAT-5	PEP	CAT-5
2000 Upper-Grade Test	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2000 Lower-Grade Test	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9
2001 Upper-Grade Test	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
1999-2000 Lower-Grade→ Upper-Grade Correlation	n/a	.60	.73	.77	.72	.73	.73
2000-2001 Lower-Grade→ Upper-Grade Correlation	.65	.65	.71	.76	.78	.81	.75

n/a – not available (no test given)
Note. Lower and Upper Grade refer to the two grades listed at the top of each column.

Notice in Table 2 that the correlations for adjacent cohorts taking different combinations of tests are very similar (i.e., there is little difference within each column and subject). Also note that the lower-grade to upper-grade correlations increase as grade level increases (i.e., the correlations increase from left to right within each row). This suggests that the tests share similar pre-test/post-test relationships, and the differences between pre-test and post-test are smaller in later grades.[3]

[3] This is consistent with the differential rates of growth for grades 1-3 and 4-8 mentioned later in this section.

Visual inspection of the distribution of scores on the six tests for each grade and year in this study also supported the equating assumptions. For example, the shapes of the distributions of scores for all grades on all relevant tests in all years were very similar and approximately normal. Figure 2 shows box plots of the rescaled reading achievement scores by grade. This plot shows consistent variance from grade to grade; however, the rate of growth is much steeper from first to third grade as compared to the rate of growth from fourth to eighth grade. A similar difference in the growth rates from first to third grade and fourth to eighth

Figure 2. Box Plots of Rescaled Reading Achievement Scores

grade is also present in the national norms for the SAT-9 in both reading and mathematics (Harcourt, 1997, Table 72). This may be due to actual differences in rates of learning or an artifact of the scaling methodology for the SAT-9. To circumvent this issue, these two grade spans are analyzed separately in this report.

Methods

To analyze these longitudinal data, we used an advanced statistical modeling technique called hierarchical growth curve analysis. This technique enabled us to model the annual growth in individual students' reading and mathematics performance, while adjusting for differences in the demographic characteristics of students. Most importantly, this method allowed us to determine the extent to which differences in students' performance and growth were due to particular individual factors (e.g., minority status) and school-level factors (e.g., participation in America's Choice). Even though America's Choice did not start in Rochester schools until 1998, having

data from 1993 permitted us to examine differences in the rates of growth of students from before America's Choice began in Rochester until deep into and after its implementation.

The Statistical Model

The final model for this analysis was a three-level hierarchical growth curve model (Raudenbush & Bryk, 2002, p. 178) with time points nested within students, and students nested within schools. Student age was used to represent time, and it was centered around nine years old (i.e., the average age of a fourth grader). Student gender, minority status, free or reduced-price lunch eligibility, limited English proficiency status, and special education status were included as control variables predicting student status at age nine, and the annual growth in achievement. All control variables were centered around the grand mean for America's Choice students. This enabled interpretation of the intercepts as representing the growth curve for a student with demographic characteristics that are similar to the typical America's Choice student.

Separate models were fit for the early-elementary grades (grades 1-3) and the upper-elementary and middle grades (grades 4-8). This was done to improve the fit of the model, given that increases in SAT-9 scores were rapid during grades 1-3, then slowed down after third grade. This two-staged trend can be seen in Figure 2. Notice how the growth is linear from grades 1-3, and continues at a slower linear pace in grades 4-8. This *piecewise* approach using two linear growth models was selected instead of a curvilinear growth model in an effort to maintain simple interpretation of the growth parameters. For example, it is easy to understand the linear growth estimate of *points gained per year*, while it is not easy to understand the quadratic growth estimate of *points per year-squared*.

The modeling strategy used in this study compares annual learning gains of America's Choice students to learning gains for students in other schools from 1998-2003, and learning gains for the entire district from 1993-1998. In other words, this model uses a control group *and* multiple baseline measures to estimate effects. This is similar to an *interrupted time series* design with a non-equivalent control group (see Shadish, Cook, & Campbell, 2002). However, the time series in this study is modeled via a hierarchical latent growth curve model instead of a traditional time series model.

It is important to note that this method of analysis produces a time-averaged estimate of the effect of America's Choice individual student achievement growth. In other words, a linear model including the number of years each student has attended an America's Choice school is unable to detect differences in the effect of the

program during the early stages of implementation compared to later stages. The estimate of annual impact on student performance is the same regardless of the number of years the school has been implementing the design. Unfortunately, the production of year-by-year estimates of effect is complicated by the high mobility of students in the Rochester district. More specifically, effects after the first year of implementation may be mitigated by the fact that many students are being exposed to the program for the first time. Therefore, valid estimation of program effects separately for year 1, year 2, year 3, etc., becomes very difficult. Fortunately, we can validate the assumption of consistent effects across cohorts and implementation years by considering exposure to America's Choice in two ways, then comparing the results. First, we estimate the impact of students' exposure to the design (i.e., the number of years attending an America's Choice school), and we do not differentiate between schools that have been implementing America's Choice for differing amounts of time. Next, we estimate the impact of schools' exposure to the design (i.e., the number of years a school has implemented America's Choice), and we do not differentiate between students that have been attending America's Choice schools for differing amounts of time. Additional analyses used to check the fit of the statistical models (e.g., residual analysis) can also help to determine whether changes in achievement are unusually large or small for any given year.

For the model of student exposure to the design, a time-varying covariate indicating the number of years a student had been in an America's Choice school was included in the model. The estimate

for this variable shows the additional change in achievement score above the baseline growth rate for each year in an America's Choice school. Simply stated, this shows the annual impact of the program on individual student achievement. For the model of school exposure to the design, a set of indicator variables showing the number of years a school has been implementing the design was included in the model. These produce estimates of the effect of the program on student performance for each year that the program is in place. In other words, it shows the year-by-year effect of America's Choice. For the advanced reader, the mathematical forms of the models are shown in Appendix A.

Three key pieces of information are produced by the growth-curve models, and all are adjusted to control for differences in the demographic characteristics of America's Choice students and students in other schools. First, the models produce an estimate of the typical annual growth in achievement. In the simplest terms, this shows the average number of points gained per year over multiple years. Second, the models produce estimates of the additional annual growth experienced by students participating in America's Choice. In other words, this shows the number of extra points students gain per year on average while they attend an America's Choice school. Third, by dividing the additional growth attributable to America's Choice by the baseline growth estimate and multiplying by 10 months, we produce an estimate of the number of additional months of learning that students experience each year that they attend an America's Choice school.

Model Estimation and Software

Because students switched schools during the study, the data did not have a pure nested structure; that is, not all the observations for a single student were nested within the same school. Therefore, traditional hierarchical linear modeling estimation methods could not be used. Fortunately, the use of crossed random effects (Rasbash & Goldstein, 1994; Raudenbush, 1993) combined with Bayesian estimation techniques (see Browne, 2002, p. 165) has made possible analyses of data with complex nesting structures. The Bayesian technique used here is a simulation-based method called Markov Chain Monte Carlo (MCMC) using Gibbs Sampling (Geman & Geman, 1984). The model was estimated using MCMC as implemented in *MLwiN 2.0* (Browne, 2002; Rasbash, Browne, Healy, Cameron, & Charlton, 2003). Details of the estimation process are given in Appendix A.

Analysis Strategy

In accordance with our research questions, we conducted a staged analysis. The first stage focused on overall effects on achievement growth, and is separated into effects on individual student growth and year-by-year effects of the program. These analyses include all students in the sample. The second stage explores the effects of the program on the achievement growth of students with different initial performance levels (i.e., the data are reanalyzed by initial performance quartile). Finally, the third stage involves the estimation of different effects for students of different ethnicities. Whereas the first stage of analysis seeks to answer the overarching question of general program effectiveness, the second and third

stages of analysis seek to determine the degree to which America's Choice is successful in closing the gaps between low and high performers and between minority and White students.

Results

In this section, we present the results of our analyses. We first describe the average annual test score gain for students in Rochester and compare this to the national norms for the SAT-9. We then present results of analyses designed to answer the three research questions presented earlier. First, we examine the effect of America's Choice on students' annual gains in general and for each year of implementation. Second, we look more specifically at the effect of America's Choice on the annual gains of students with different initial performance. Third, we look at the effects of America's Choice on the annual gains of students from different ethnicities. In the sections below, we report the impact estimates for each analysis. Appendices B through E contain detailed results for the full models.

Baseline Test Score Gains for All Rochester Students

In order to interpret the meaning of the impact estimates, it is helpful to know the magnitude of typical annual growth in reading and mathematics achievement scores at different grade levels for all students in Rochester. Differences in rates of growth in the early grades versus the later grades could be due to differences in the amount of learning that is typically experienced for different subjects at different grade levels, or they could simply be artifacts of the scale of the

test. Either way, knowledge about the expected rate of growth in SAT-9 scores for all students in the nation helps in benchmarking both the annual rates of learning for Rochester students in general, and also in benchmarking the effects of America's Choice on student learning gains.

In grades 1-3, Rochester students gained 29.7 points per year in reading and 27 points per year in mathematics, on average. These rates of growth are smaller than national norms for the SAT-9 during grades 1-3, which has expected annual rates of growth of 49.9 points in reading and 36.7 points in mathematics.[4] This suggests that Rochester students in grades 1-3 are learning at a rate that is 40% slower than national norms in reading and 26% slower than national norms in mathematics.

In grades 4-8, Rochester students gained 8.6 points per year in reading and 6.2 points per year in mathematics, on average. These rates of growth are also smaller than national norms for the SAT-9 during grades 4-8, which has expected annual rates of growth of 14.1 points in reading and 14.8 points in mathematics (see footnote 4). This suggests that Rochester students in grades 4-8 are learning at a rate that is 39% slower than national norms in reading and 58% slower than national norms in mathematics.

Although the number of points gained on the SAT-9 in grades 1-3 is much larger than in grades 4-8, the

---

[4] The annual growth rates for the national norming sample were calculated using grade equivalents and scale scores published in the *Stanford-9 Spring Norms Book* (Harcourt, 1997).

comparison to national norms shows that, in reading, the difference in the number of points gained is *not* an indication that Rochester students in grades 4-8 are performing worse than Rochester students in grades 1-3. In fact, the annual gains in reading are equally poor for grades 1-3 and 4-8 (with both being about 40% below national norms). In mathematics, however, there is evidence that Rochester students perform worse while in grades 4-8 (58% below national norms) compared to performance during grades 1-3 (26% below national norms).

Mean Effects of America's Choice on Student Achievement Growth

Overall, students attending America's Choice schools in Rochester experienced significantly greater annual gains in both reading and mathematics performance than did similar students in other schools in the district. Table 3 shows annual gains by grade level and subject. It also shows the additional annual gains made by America's Choice students in terms of months of schooling. For both reading and mathematics, students attending America's Choice schools in grades 1-3 experienced an additional seven-tenths of a month of learning each year compared to similar students in other Rochester schools. In grades 4-8, students in America's Choice schools experienced an additional 1.7 months of learning each year in reading, and an additional 2.6 months of learning each year in mathematics, compared to similar students in other Rochester schools.

In terms of points gained on the SAT-9, students in non-America's Choice schools increased their reading scores by 8.6 points per year while in grades 4-8, while students attending America's Choice schools increased their reading scores by 10.1 points per year while in grades 4-8 (1.5 points in addition to the baseline 8.6 points per year).

Table 3. Baseline Growth in Achievement and Effects of America's Choice in Rochester, NY

Growth Estimate	Annual Achievement Growth (standard errors in parentheses)			
	Reading		Mathematics	
	Grades 1-3	Grades 4-8	Grades 1-3	Grades 4-8
Baseline Annual Growth of SAT-9 Scores	29.7*** (0.6)	8.6*** (0.2)	27.0*** (0.7)	6.2*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points Per Year	2.0*** (0.3)	1.5*** (0.1)	1.9*** (0.3)	1.6*** (0.2)
Additional Months of Learning Per Year	+0.7	+1.7	+0.7	+2.6

* p<.05, **p<.01, ***p<.001

Table 4. Baseline Growth in Achievement and Year-by-Year Effects of America's Choice in Rochester, NY

Growth Estimate	Annual Achievement Growth (standard errors in parentheses)			
	Reading		Mathematics	
	Grades 1-3	Grades 4-8	Grades 1-3	Grades 4-8
Baseline Annual Growth of SAT-9 Scores	29.7*** (0.6)	8.6*** (0.3)	26.9*** (0.7)	6.0*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points in Year 1	2.2*** (0.6)	2.8*** (0.3)	3.3*** (0.6)	2.9*** (0.3)
Additional SAT-9 Points in Year 2	-0.5 (0.7)	1.6*** (0.3)	-0.8 (0.7)	1.2*** (0.4)
Additional SAT-9 Points in Year 3	2.6** (0.9)	-0.3 (0.6)	2.7*** (0.9)	-0.4 (0.6)
Additional SAT-9 Points in Year 4	4.6*** (0.9)	2.6*** (0.7)	1.9* (0.9)	2.5*** (0.7)
Additional SAT-9 Points in Year 5	-1.9* (1.0)	-1.4* (0.7)	0.2 (1.0)	-0.7 (0.8)

* p<.05, **p<.01, ***p<.001

It is important to recognize that these estimates of impact on achievement growth are averaged across years. In the next section, we present year-by-year estimates in order to determine the degree to which the mean estimates accurately represent a consistent trend.

Year-by-Year Effects of America's Choice

Efforts to separate the overall effect of America's Choice into year-by-year effects suggested that there is variation in the program impact across years; however, the effects were neither concentrated in the early years of implementation, nor in the later years. Table 4 shows annual impact by grade level and subject. There is some indication that program effects dropped off in the fifth year, especially in reading, where the impact estimate is negative. However, without an in-depth qualitative study or a sixth year of test score data, it is unclear whether this is the result of a systematic downward trend or just another dip in a somewhat

random process. It is also interesting to note that only six schools in this study were using the America's Choice design for five years, and that these schools began implementing the design under a three-year implementation plan. After the third year as an America's Choice school, the program was maintained in these schools through a less involved relationship with NCEE. During this time, according to one district official, "The original prescribed design started to erode in many places." The degree to which the effects of comprehensive school reform can be sustained under *maintenance contracts* is an important issue for future research.

Across the five years of program implementation, the impact estimates are predominantly positive and statistically significant. Therefore, linear projections of student achievement growth based on a time-averaged estimate of program impact can serve as a relatively accurate approximation of changes in individual student

performance during the first few years of implementation.

Visual Projections of Student Achievement Growth

In this section, we introduce visual plots of individual student achievement growth using time-averaged impact estimates from Table 2. In these plots, two lines are used to represent typical achievement trends for America's Choice students (red line) and similar students in other Rochester schools (gray line). This helps to improve the perspective for interpreting the size of the impact estimates, and for gauging the cumulative effects of the program over time.

Figure 3 shows a visual depiction of two prototypical Rochester students with similar demographics who start out the first grade with identical reading achievement scores. The student in an America's Choice school experiences a slightly faster rate of growth than the student in a non-America's Choice

Figure 3. Growth in Reading Achievement of America's Choice Students and Similar Students in Other Schools in Rochester During Grades 1-3
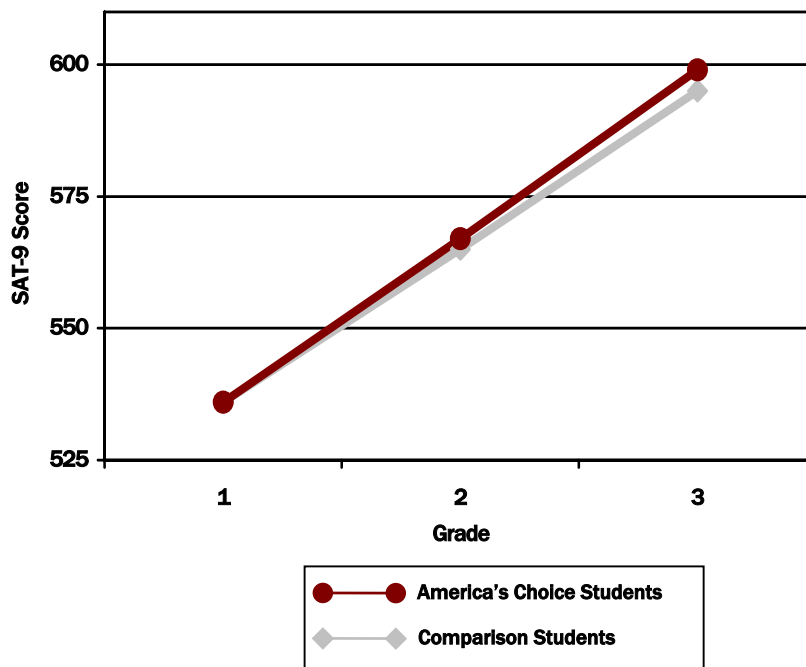
school. After two years, a student attending an America's Choice school would have gained 63.4 points (31.7 points x 2), while a student in a non-America's Choice school would have gained only 59.4 points (29.7 points x 2).

Figure 4 shows a visual depiction of two prototypical Rochester students with similar demographics who start out in the fourth grade with identical achievement scores. The student in an America's Choice school experiences a considerably faster rate of growth than the student in a non-America's Choice school. After four years, a student attending an America's Choice school would have gained 40.4 points (10.1 points x 4), while a student in a non-America's Choice school would have gained only 34.4 points (8.6 points x 4).

Effects of America's Choice on Students with Different Initial Performance Levels

The lowest performing students in America's Choice schools typically experienced the greatest gains relative to similar students in other Rochester schools, although there were effects of America's Choice present at all performance levels. To determine the effects of America's Choice on students at different performance levels, we grouped students into four performance groups (quartiles) in order to compare students in America's Choice schools to those in other Rochester schools within each of the performance quartiles. Students were grouped into performance quartiles based upon their first available achievement score. Because test scores within each grade were similarly distributed across years, the same grade-specific cutoff scores were used for all years. Details of these results are given in Table 5.

Figure 4. Growth in Reading Achievement of America's Choice Students and Similar Students in Other Schools in Rochester During Grades 4-8
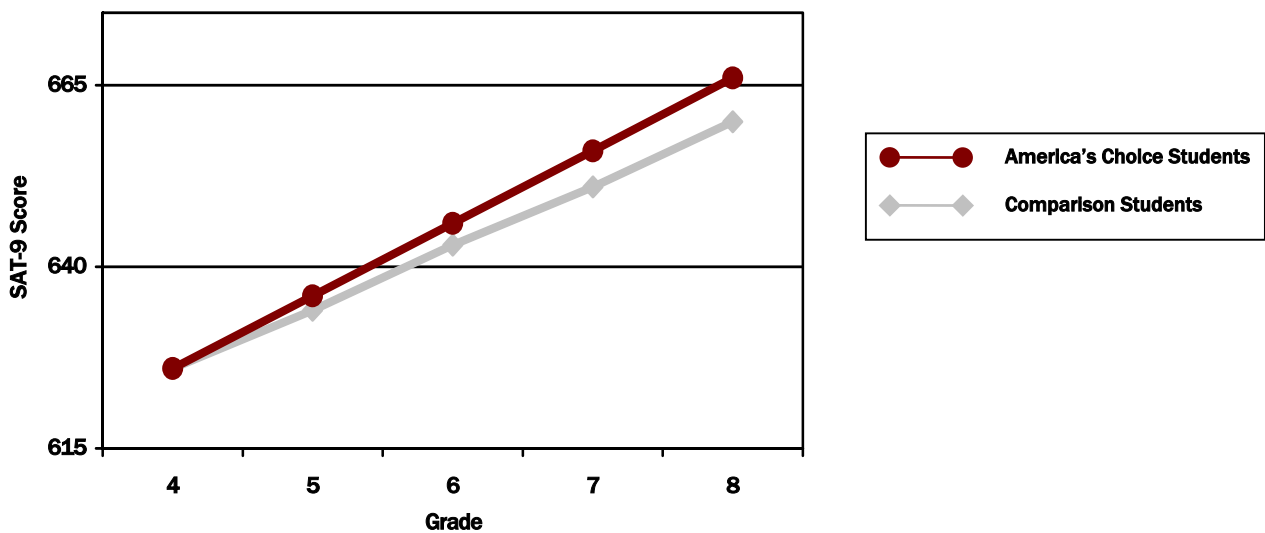
Table 5. Baseline Growth in Achievement and Effects of America's Choice by Initial Performance

	Annual Achievement Growth (standard errors in parentheses)			
	Reading		Mathematics	
	Grades 1-3	Grades 4-8	Grades 1-3	Grades 4-8
First Quartile (Lowest Initial Performance)				
Baseline Annual Growth of SAT-9 Scores	38.0*** (0.6)	8.7*** (0.4)	30.7*** (0.5)	6.3*** (0.3)
Effect of America's Choice				
Additional SAT-9 Points Per Year	3.6*** (0.6)	3.0*** (0.3)	4.4*** (0.6)	2.4*** (0.3)
Additional Months of Learning Per Year	**+0.9**	**+3.4**	**+1.4**	**+3.8**
Second Quartile				
Baseline Annual Growth of SAT-9 Scores	31.6*** (0.4)	8.6*** (0.3)	26.4*** (0.4)	5.4*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points Per Year	-0.5 (0.5)	1.7*** (0.2)	0.8 (0.5)	1.8*** (0.2)
Additional Months of Learning Per Year	~	**+2.0**	~	**+3.3**
Third Quartile				
Baseline Annual Growth of SAT-9 Scores	27.5*** (0.4)	8.4*** (0.3)	24.3*** (0.5)	5.8*** (0.5)
Effect of America's Choice				
Additional SAT-9 Points Per Year	-0.7 (0.5)	0.3 (0.2)	-0.6 (0.5)	0.6* (0.3)
Additional Months of Learning Per Year	~	~	~	**+1.0**
Fourth Quartile (Highest Initial Performance)				
Baseline Annual Growth of SAT-9 Scores	17.9*** (0.5)	8.2*** (0.3)	18.7*** (0.7)	6.0*** (0.7)
Effect of America's Choice				
Additional SAT-9 Points Per Year	-1.2 (0.6)	-0.3 (0.3)	0.2 (0.6)	1.4*** (0.4)
Additional Months of Learning Per Year	~	~	~	**+2.3**

* p<.05, **p<.01, ***p<.001
~ no significant difference in annual gains

America's Choice students in the lowest quartile (percentile rank less than 25) gained significantly more than did the lowest performing students in other district schools. The lowest quartile students in America's Choice schools gained nearly one additional month in reading during grades 1-3 (38 points + 3.6 points per year) and an additional 3.4 months per year in reading during grades 4-8 (8.7 points + 3.0 points per year) during the years that they attended an America's Choice school. These lowest performing students also gained an additional 1.4 months in mathematics during grades 1-3 (30.7 points + 4.4 points per year) and an additional 3.8 months in mathematics during grades 4-8 (6.3 points + 2.4 points per year) for every year that they attended an America's Choice school.

America's Choice students in the second quartile (percentile rank between 25 and 50) had significantly greater gains in both reading and mathematics in grades 4-8, but not in grades 1-3. While in grades 1-3, second-quartile students gained similarly in both America's Choice and other district schools; however, these students gained an additional two months in reading (8.6 points + 1.7 points per year) and an additional 3.3 months in mathematics (5.4 points + 1.8 points per year) while in grades 4-8, compared to their counterparts in other district schools.

Gains for students in the third quartile (percentile rank between 50 and 75) were generally similar between those in America's Choice schools and those in other Rochester schools. Only America's Choice students in grades 4-8 in mathematics exhibited greater learning gains, an additional one month per year (5.8 points + 0.6 points per

year) during the years that they attended an America's Choice school.

For the highest performing students, those in the fourth quartile (percentile rank greater than 75), there were few differences between those in America's Choice schools and those in other Rochester schools. Only in mathematics during grades 4-8 did students in America's Choice schools have greater gains, an additional 2.3 months per year (6.0 points + 1.4 points per year).

Finally, it is worth noting that there were no cases in any of the quartiles and in either reading or mathematics in which students in other Rochester schools out-gained students in the America's Choice schools after adjusting for differences in student demographics between the two populations of students.

Figure 5 shows a visual depiction of four pairs of prototypical Rochester students with similar demographics who start out the fourth grade with a reading achievement score at the average value for each of four quartiles. The students in America's Choice schools who start out with lower achievement scores experience a faster rate of growth than their counterparts in non-America's Choice schools. This can be seen by the steeper slope of the lines representing America's Choice students from quartiles 1 and 2. Furthermore, the growth of higher performing students in America's Choice schools is equivalent to the growth rates for their counterparts in non-America's Choice schools. This shows that students with lower initial performance experience significant benefits from America's Choice, but not at the expense of higher performing students.

**Figure 5. Growth in Reading Achievement of America's Choice Students and Similar Students in Other Schools in Rochester During Grades 4-8 by Initial Performance Quartile**
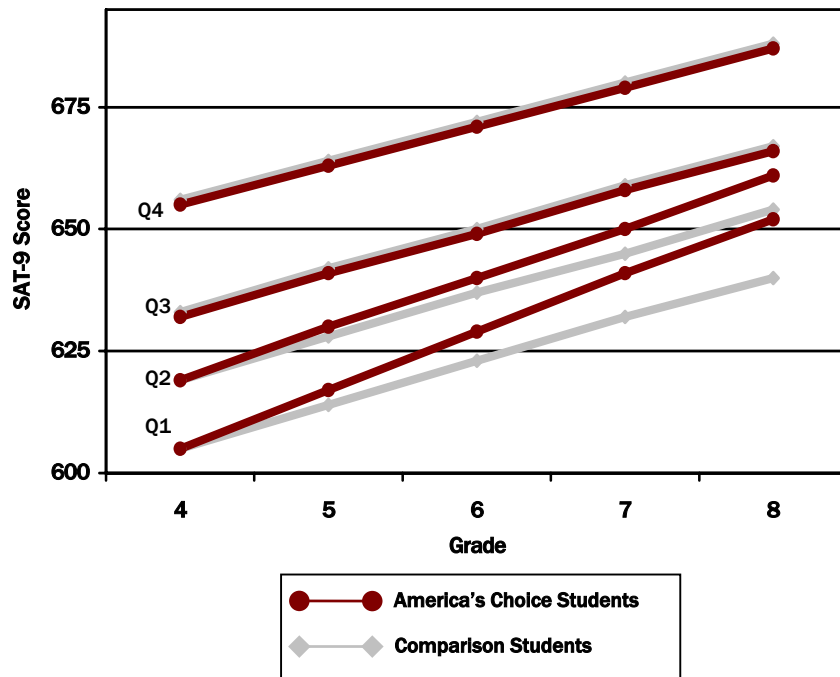
Effects on Students of Different Ethnicities

Minority students exhibited the largest learning gains from their involvement in America's Choice relative to minority students in Rochester's non-America's Choice schools. Both Hispanic and African American students in America's Choice schools typically gained significantly more than their ethnic counterparts in non-America's Choice schools. These results are shown in Table 6.

White students in America's Choice schools did not experience significant additional gains in reading in comparison to White students in other Rochester schools. In mathematics, they gained an additional 2.1 months per year during grades 4-8 (6.1 points + 1.3 points per year).

African American students in America's Choice schools gained an additional two-fifths of a month per year in reading during grades 1-3 (29 points + 1.2 points per year) and an additional 1.4 months per year in reading during grades 4-8 (8.4 points + 1.2 points per year). They also gained an additional one-half of a month per year in mathematics during grades 1-3 (25.9 points + 1.3 points per year) and an additional 1.8 months per year in mathematics during grades 4-8 (6.0 points + 1.1 points per year).

Hispanic students in America's Choice schools gained an additional 1.2 months per year in reading during grades 1-3 (29.0 points + 3.6 points per year) and an additional 2.9 months per year in reading during grades 4-8 (8.4 points + 2.4 points per year). They also gained an additional 1.2 months per year in mathematics during grades 1-3

Table 6. Baseline Growth in Achievement and Effects of America's Choice by Ethnicity

	Annual Achievement Growth (standard errors in parentheses)			
	Reading		Mathematics	
	Grades 1-3	Grades 4-8	Grades 1-3	Grades 4-8
White Students				
Baseline Annual Growth of SAT-9 Scores	29.6*** (0.7)	8.6*** (0.3)	26.9*** (0.7)	6.1*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points Per Year	-0.4 (1.0)	0.8 (0.5)	-0.4 (1.0)	1.3* (0.5)
Additional Months of Learning	~	~	~	+2.1
African American Students				
Baseline Annual Growth of SAT-9 Scores	29.0*** (0.6)	8.4*** (0.3)	25.9*** (0.5)	6.0*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points Per Year	1.2*** (0.4)	1.2*** (0.2)	1.3** (0.4)	1.1*** (0.2)
Additional Months of Learning	+0.4	+1.4	+0.5	+1.8
Hispanic Students				
Baseline Annual Growth of SAT-9 Scores	29.0*** (0.6)	8.4*** (0.3)	25.9*** (0.5)	6.0*** (0.4)
Effect of America's Choice				
Additional SAT-9 Points Per Year	3.6*** (0.6)	2.4*** (0.3)	3.1*** (0.7)	1.6*** (0.3)
Additional Months of Learning	+1.2	+2.9	+1.2	+2.7

* p<.05, **p<.01, ***p<.001
~ no significant difference in annual gains

(25.9 points + 3.1 points per year) and an additional 2.7 months per year in mathematics during grades 4-8 (6.0 points + 1.6 points per year).

Figure 6 shows a visual depiction of three pairs of prototypical White, Hispanic, and African American Rochester students. The minority students in America's Choice schools experience a faster rate of growth than their counterparts in non-America's Choice schools. This can be seen by the steeper slope of the lines representing Hispanic and African American America's Choice students. Furthermore, the growth of White students in America's Choice schools is equivalent to the growth rates for their counterparts in non-America's Choice schools. This shows that minority students experience significant benefits from America's Choice, but not at the expense of non-minority students.

Figure 6. Growth in Reading Achievement of America's Choice Students and Similar Students in Other Schools in Rochester During Grades 4-8 by Ethnicity
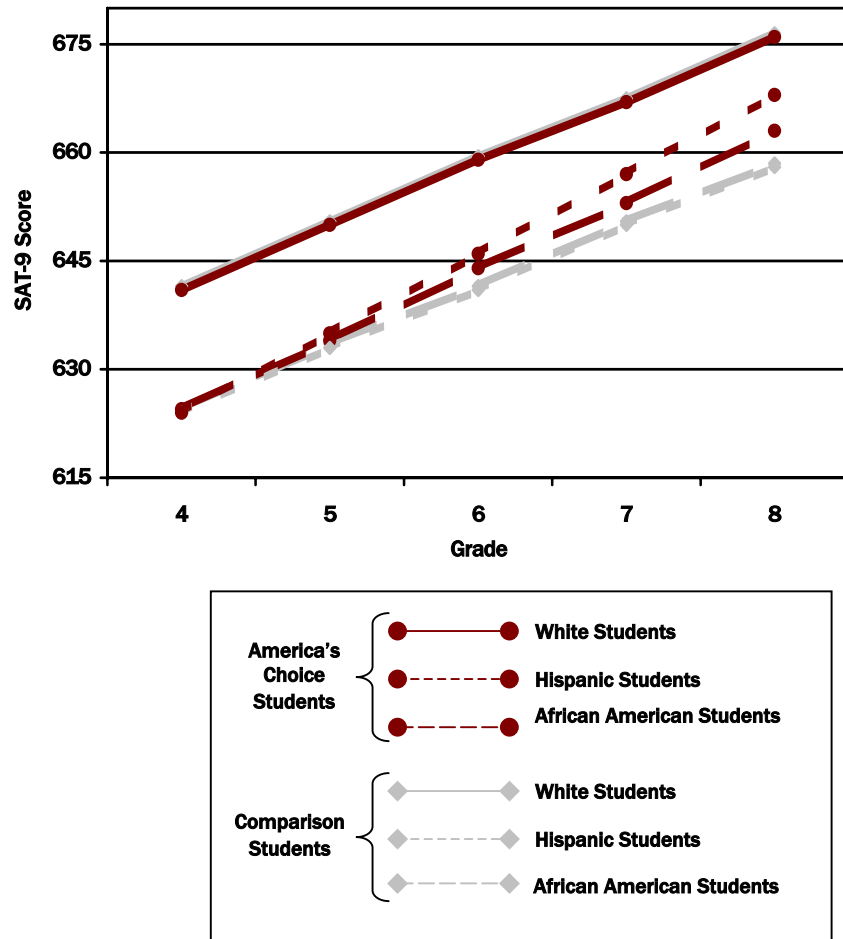
Summary and Implications

These analyses reveal a number of important findings about the effectiveness of America's Choice in Rochester. First and foremost, the overall results indicate that, on average, students in America's Choice schools learned significantly more than did other students in the district, even after adjusting for differences in student demographics. Each year, the magnitude of these results is small to moderate, but they accumulate over the time that students attend America's Choice schools.

Furthermore, the impact of America's Choice seems to be larger in the later grades than in the early grades. In fact, the results are 2-3 times stronger in the later grades (1.7 months in reading and 2.6 months in mathematics in grades 4-8) than in the early grades (0.7 months in both reading and mathematics in grades 1-3). This could be due to more powerful programmatic influences in the later grades; however, the very small baseline gains for students in grades 4-8 suggest another

interpretation. If Rochester students are learning very little each year during grades 4-8, the proportional effect of America's Choice may be an overstatement of its actual effectiveness. That is, even though Rochester students in America's Choice schools are learning more than their counterparts in other Rochester schools, they are by no means learning an overwhelming amount each year.

Nonetheless, the evidence also suggests that America's Choice is particularly powerful for low-performing and minority students. The gains in performance of students in the lowest quartile and African American and particularly Hispanic students who attend America's Choice schools, relative to their counterparts in non-America's Choice schools in the district, are particularly dramatic. It could be that the ethnicity effect and the effect on the lowest achieving students are due to the same phenomenon given the correlation between low performance and minority status. Regardless, the America's Choice strategies of identifying and paying substantial attention to bringing lower performing students up to standard and differentiating instruction for learners at all levels appears to be resulting in substantial achievement gains for the most disadvantaged students.

While this study does provide strong evidence of the longitudinal effects of the America's Choice program, it is not without limitations. First, schools in Rochester were not randomly assigned to America's Choice. Without such an experimental design, we must rely on statistical models to adjust for pre-existing differences between America's Choice schools and comparison schools. The methods used

in this study are thought to be quite effective in accomplishing this goal (Shadish, Cook, & Campbell, 2002); however, the study would be even stronger if America's Choice schools had been selected at random from a pool of eligible schools.

Second, this study does not attempt to directly connect improvements in performance to implementation of specific components of the America's Choice design (as was done in May, Supovitz, & Lesnick, 2004; and Supovitz & May, 2003). Therefore, while the study suggests that the program has positive impacts, it does not provide specific evidence of how and why these positive results were achieved.

The biggest strength of this study stems from an ability to track individual student performance over many years. Although these circumstances have been quite rare in the past, efforts by states to satisfy the requirements of No Child Left Behind should create rich databases that will enable these kinds of powerful longitudinal analyses to become more common in education policy research and program evaluation.

References

Browne, W. J. (2002). *MCMC estimation in MLwiN*. London: Centre for Multilevel Modelling, University of London.

Geman, S., & Geman, D. (1984). Stochastic relaxation Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721-741.

Harcourt. (1997). *Stanford Achievement Test Series, ninth edition, spring norms book*. San Antonio, TX: Author.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83-102.

May, H. (2004, April). *Practical longitudinal analysis of student achievement using cross-classified Bayesian Hierarchical Linear Modeling*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

May, H., Supovitz, J. A., & Lesnick, J. (2004). *The impact of America's Choice on writing performance in Georgia: First-year results*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

National Center on Education and the Economy (NCEE) (n.d.). *America's Choice: Program overview*. Retrieved August 28, 2003 from http://www.ncee.org/acsd/.

Rasbash, J., Browne, W., Healy, M., Cameron, B., & Charlton, C. (2002). *MLwiN* (Version 2.00) [Computer Software]. London: Centre for Multilevel Modelling, University of London.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19, 337-350.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational and Behavioral Statistics*, 18, 321-349.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edition). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

Supovitz, J. A., & May, H. (2003). *The relationship between teacher implementation of America's Choice and student learning in Plainfield, New Jersey*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

About the America's Choice Design

The America's Choice school design is a K-12 comprehensive school reform model developed by the National Center on Education and the Economy (NCEE). America's Choice is a well-established school reform model currently being implemented in over 500 schools across the nation. America's Choice focuses on raising academic achievement by providing a rigorous standards-based curriculum and safety nets for all students. A stated goal of the design is "to ensure that every student is successful on state and local assessments and prepared for college" (NCEE, n.d.).

America's Choice does not offer schools a script or a paint-by-numbers approach to reformed instruction. America's Choice recognizes that the pace of change will vary from school to school and, thus, the model does not have a rigid implementation schedule. Rather, the core of the America's Choice design contains a set of principles about the purpose of schooling and how schools should operate, and it provides a set of tools for building a program based on those principles. These essential principles and tools include:

- High expectations for all students with communication of those expectations through explicit performance standards that are aligned to assessments and include examples of student work that meet the standards. These New Standards Performance Standards provide a common set of expectations for students and teachers.

- The implementation of standards-based literacy, mathematics, and skills blocks, which happen every day for every child, and dramatically change teaching and learning in every classroom. The rituals and routines associated with these blocks are designed to prepare students to deal with demanding content and become independent learners.

- A common core curriculum that is aligned with the standards. Through the America's Choice literacy workshops, core assignments, and foundations of advanced mathematics, school life is organized around a core curriculum. These curricular components are designed to help students develop key skills, convey core concepts, and apply what they know.

- Ongoing assessment of students in order to inform daily instruction through the use of standards-based assessments, including the New Standards Reference Examination, which are aligned with the standards and the core curriculum and provide detailed feedback to teachers and students about student skill levels in relation to the standards.

- School-embedded, ongoing, teacher professional development led by a full-time literacy coach and designed to strengthen teachers' knowledge of the America's Choice approach to teaching and learning. This includes learning how to conduct a close analysis of their students' work in relation to standards, and using this knowledge to develop lessons calibrated to the needs of different students.

- "Safety nets," including tutoring and course recovery programs, that are structured into the school day and school year, and that provide students with extensive support and multiple opportunities to achieve the standards.

- A school leadership team, led by the principal and subject-matter coaches,

that coordinates implementation through a variety of means. These include setting performance targets and analyzing student work on a variety of measures, training teachers, adjusting school schedules, and implementing safety-net programs to provide time for students to receive additional instruction.

In order to become an America's Choice school, over 80% of a school's faculty must indicate their commitment to the America's Choice design and agree to implement the program over three years. Each school must assign personnel as coaches to lead the implementation of the design, and a parent/community outreach coordinator to ensure that students get needed support services.

About CPRE's Evaluation of America's Choice

The Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania was contracted by NCEE in 1998 to conduct the external evaluation of the America's Choice school design. Each year, CPRE designs and conducts a series of targeted studies on the implementation and impacts of the America's Choice design.

The publication of this report follows the release of several other studies by CPRE on the implementation and impact of America's Choice across the nation using both qualitative and quantitative methods. Those previous reports can be viewed as separate pieces or as complements to the information presented in this report.

The purpose of CPRE's evaluation is to provide formative feedback to NCEE and America's Choice schools about emerging trends in the implementation of the design, and to seek evidence of the impacts of the

design using accepted high standards of evaluation design and analysis methodologies.

CPRE's evaluation of America's Choice is guided by three overarching evaluation questions. First, is America's Choice being carried out in the manner envisioned — that is, how are teachers and school administrators understanding and implementing the many facets of the America's Choice reform design? Second, as a result of their implementation of America's Choice, are the instructional practices of teachers changing in ways that would improve student learning? Third, do America's Choice schools experience improvements in student achievement, and to what degree can changes in student performance be attributed to the design? Within this framework, annual evaluation studies target specific aspects of the America's Choice design for more in-depth investigation.

To address these questions, the CPRE evaluation team gathers a broad array of qualitative and quantitative data to develop a rich and valid picture of the implementation process over time and to capture the impacts of the design on students and teachers. Data sources include:

- Surveys of teachers and administrators in America's Choice schools nationwide.

- Site visits to schools across the nation to observe classroom instruction, examine implementation artifacts, and interview teachers, students, and school administrators.

- Telephone interviews with NCEE staff, school faculty members, and school and district administrators.

- Document reviews.

- Observations of national, regional, and school-level professional development.

- Collection of student performance measures, including state and local tests, the New Standards Reference Examination, and more authentic samples of student work products.

After data collection, CPRE evaluation team members analyze the data using appropriate qualitative and quantitative research techniques in order to identify patterns of intended and unintended consequences and to detect effects of the design on students, teachers, and schools. The results are reported in a series of thematic evaluation reports that are released each year.

Additional Reading on America's Choice

The following reports are currently available from CPRE. Print copies are available at no cost by emailing cpre@gse.upenn.edu, or by calling 215-573-0700. Copies can also be downloaded at www.cpre.org/Research/Research_Project_America's_Choice.htm.

- **The Impact of America's Choice on Writing Performance in Georgia: First-Year Results** (Henry May, Jonathan A. Supovitz, and Joy Lesnick, July 2004)

- **Mapping a Course for Improved Student Learning: How Innovative Schools Systematically Use Student Performance Data to Guide Improvement** (Jonathan A. Supovitz and Valerie Klein, November 2003)

- **Teacher and Coach Implementation of Writers Workshop in America's Choice Schools, 2001 and 2002** (Amy J. Bach and Jonathan A. Supovitz, October 2003)

- **The Heart of the Matter: The Coaching Model in America's Choice Schools** (Susan M. Poglinco, Amy J. Bach, Kate Hovde, Sheila Rosenblum, Marisa Saunders, and Jonathan A. Supovitz, May 2003)

- **The Relationship Between Teacher Implementation of America's Choice and Student Learning in Plainfield, New Jersey** (Jonathan A. Supovitz and Henry May, January 2003)

- **Impact of America's Choice on Student Performance in Duval County, Florida** (Jonathan A. Supovitz, Brooke Snyder Taylor, and Henry May, October 2002)

- **Implementation of the America's Choice Literacy Workshops** (Jonathan A. Supovitz, Susan M. Poglinco, and Amy J. Bach, April 2002)

- **Instructional Leadership in a Standards-Based Reform** (Jonathan A. Supovitz and Susan M. Poglinco, December 2001)

- **Moving Mountains: Successes and Challenges of the America's Choice Comprehensive School Reform Design** (Jonathan A. Supovitz, Susan M. Poglinco, and Brooke Snyder, March 2001)

- **America's Choice Comprehensive School Reform Design: First-Year Implementation Evaluation Summary** (Thomas Corcoran, Margaret Hoppe, Theresa Luhm, and Jonathan A. Supovitz, February 2000)

Appendix A. Technical Details of Statistical Analyses

This technical appendix provides details of the statistical methods used in this study. The first section focuses on the process by which achievement scores from multiple assessments were linked across years. The second section focuses on the statistical analysis of growth in student achievement and the impact of America's Choice on annual achievement growth.

Linking Scores from Different Tests

During the time period covered by this study, the Rochester City School District used six different assessments to measure student performance in reading and mathematics. Table A1 shows these tests by subject, year, and grade in which they were used.

Table A1. Achievement Tests Used by Subject, Year, and Grade

Subject/Year	Grade							
	1	2	3	4	5	6	7	8
Reading								
1993	n/a	n/a	PEP	n/a	n/a	PEP	n/a	n/a
1994	CAT-5	DRP	PEP	DRP	DRP	PEP	DRP	PCT
1995	CAT-5	DRP	PEP	DRP	DRP	PEP	DRP	PCT
1996	CAT-5	DRP	PEP	DRP	DRP	PEP	DRP	PCT
1997	CAT-5	DRP	PEP	DRP	DRP	PEP	DRP	PCT
1998	CAT-5	DRP	PEP	DRP	DRP	PEP	DRP	PCT
1999	CAT-5	DRP	PEP	NYS	DRP	PEP	DRP	NYS
2000	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2001	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2002	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2003	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
Mathematics								
1993	n/a	n/a	PEP	n/a	n/a	PEP	n/a	n/a
1994	CAT-5	CAT-5	PEP	CAT-5	CAT-5	PEP	CAT-5	CAT-5
1995	CAT-5	CAT-5	PEP	CAT-5	CAT-5	PEP	CAT-5	CAT-5
1996	CAT-5	CAT-5	PEP	CAT-5	CAT-5	PEP	CAT-5	CAT-5
1997	CAT-5	CAT-5	PEP	CAT-5	CAT-5	PEP	CAT-5	CAT-5
1998	n/a	CAT-5	PEP	CAT-5	CAT-5	PEP	CAT-5	CAT-5
1999	n/a	CAT-5	PEP	NYS	CAT-5	PEP	CAT-5	NYS
2000	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2001	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2002	SAT-9	SAT-9	SAT-9	NYS	SAT-9	SAT-9	SAT-9	NYS
2003	SAT-9	SAT-9	SAT-9	n/a	SAT-9	SAT-9	SAT-9	n/a

n/a – not available or no test given
Note. SAT-9 = Stanford Achievement Test; CAT-5 = the California Achievement Test; DRP = the Degrees of Reading Power test; NYS = the New York State assessments; PEP = the New York Pupil Evaluation Program tests; and PCT = the New York Preliminary Competency Test.

The process by which different tests were linked across years and grades was as follows. First, the distribution of test scores was examined for each test by year and grade to verify approximate normality (i.e., bell-shaped curve). Next, the means and standard deviations of SAT-9 scores were calculated for each grade and year from 2000-2003. These means and standard deviations were nearly identical across years for a given grade, so only the means and standard deviations for 2003 were used from this point forward.

The SAT-9 was administered in May of each year to students in grades 1, 2, 3, 5, 6, and 7. Therefore, expected scores for tests administered at other times during the year and in grades 4 and 8 had to be interpolated based upon the growth curve for observed mean SAT-9 scores. This was accomplished using piecewise linear regression models of mean reading and mathematics scores over time. The visual representation in Figure 2 (see page 6) shows how expected scores at any point during the year can be approximated by following the regression line across time between observed measurements. The fit of these models to the data was quite good, with all but one explaining nearly 100% of the variance in district mean achievement scores over time.

The last step in the equating process was to rescale the scores for each cohort of students within each subject and grade so that the mean and standard deviation of scores for each year and within a particular grade equaled the mean and standard deviation of the observed SAT-9 scores for that grade in 2003. The resultant scores were plotted again for each year and grade to verify consistency in the distribution and to identify unreasonable score values or outliers. There were no outliers or other abnormalities found.

Because the district mean of the rescaled scores for any given grade did not vary across time, any district-wide acceleration or deceleration of achievement growth is removed from the data. However, as noted above, there was no evidence of any district-wide changes in annual growth, given that the observed SAT-9 district means for any one grade were nearly identical for 1999-2003. For example, the district average SAT-9 score for third grade was 601 for four years straight. Similarly, the district average SAT-9 score was 640 in fifth grade for four years straight. This suggests that the performance of third graders in Rochester in 2000, on average, was nearly identical to that of third graders in Rochester from any other year, and that the same is true for fifth graders (i.e., the cohorts are interchangeable). Therefore, it is unlikely that the results of these analyses would be different if district-wide changes in growth rates were included in these analyses. Ultimately, the rescaling methods produce models that are best suited to detect differences in relative performance of individual students, which is exactly what is needed to separate out the impact of America's Choice on individual growth in achievement.

Modeling Growth in Individual Student Achievement

The models for these analyses were three-level hierarchical growth curve models (Raudenbush & Bryk, 2002) using crossed random effects and Bayesian estimation. SAT-9 reading or mathematics scores were the dependent variables, with test scores nested within students, and students nested within schools. Separate models were fit for the early-elementary grades (grades 1-3) and the upper-elementary and middle grades (grades 4-8) to reflect the change in growth trajectories from the early-elementary grades to later grades. Student age, centered around nine years, was used to represent time. Student gender, minority status, free or

reduced-price lunch eligibility, limited English proficiency status, and special education status were included as control variables predicting student status at age nine, and the slope of the growth curve. All control variables, including dummy variables, were centered around the grand mean for America's Choice students to enable interpretation of the intercepts as representing the growth curve for a student with demographic characteristics that are similar to the typical America's Choice student. In the model estimating impact of student exposure to the design, a time-varying covariate indicating the number of years a student had been in an America's Choice school was also included in the model. The estimate for this variable shows the change in growth of SAT-9 scores for each year in an America's Choice school (i.e., the impact of the program on individual rates of learning). In the model estimating impact of school exposure to the design, a set of five dummy variables indicate whether each student had been in an America's Choice school during each of the five years of implementation. The estimates for these variables show the year-by-year impact of the America's Choice program.

Because students switched schools during the study, the data did not have a pure nested structure; that is, not all the observations for a single student were nested within the same school. Therefore, traditional hierarchical linear modeling methods could not be used. This is because the assumption of nesting, when applied to longitudinal data, presumes that students do not switch classrooms or schools at any time during the study. When mobility does occur, a traditional hierarchical growth curve model would break students' growth curves into independent pieces, one for each school attended. This can result in serious bias of coefficient estimates (Browne, 2002; May, 2004).

Fortunately, the use of crossed random effects (Rasbash & Goldstein, 1994; Raudenbush, 1993) combined with Bayesian estimation techniques (see Browne, 2002, p. 165) made possible analyses of data with complex nesting structures.[5] All Rochester public schools in existence from 1992-2003 were included in the analyses. Alternative education programs were combined into a single cluster. The total school-level sample size for each model was 43 schools for grades 1-3 and 50 schools for grades 4-8. The student-level sample sizes ranged from about 10,000 students in the quartile analyses, to over 40,000 students in the overall models. The number of test scores included in the analyses ranged from about 25,000 individual scores in the quartile models to over 130,000 in the overall models. In each model, individual test scores were treated as cross-classified within students and schools.

Using the notation of Browne, Goldstein, and Rasbash (2001), the mathematical form of the model estimating the impact on individual rates of learning was as follows.

---

[5] The use of crossed random effects assumes that only the current school affects the current test score. In other words, achievement gains in a single year are assumed to be attributable to the current school and individual characteristics; that is, prior schools are assumed to have negligible influence on current gains.

$$
\begin{aligned}
y_t ={}& \beta_0 + \beta_1 \text{MALE} + \beta_2 \text{MINORITY} + \beta_3 \text{SPCED} + \beta_4 \text{LEP} + \beta_5 \text{FRL} \\
& + \beta_6 \text{AGE}' + \beta_7 (\text{AGE}' \times \text{MALE}) + \beta_8 (\text{AGE}' \times \text{MINORITY}) + \beta_9 (\text{AGE}' \times \text{SPCED}) \\
& + \beta_{10} (\text{AGE}' \times \text{LEP}) + \beta_{11} (\text{AGE}' \times \text{FRL}) + \beta_{12} (\text{YEARS IN AC}) \\
& + \left(\text{AGE}' \times u^{(3)}_{1, \text{School}(t)}\right) + \left(\text{AGE}' \times u^{(2)}_{1, \text{Student}(t)}\right) + u^{(3)}_{0, \text{School}(t)} + u^{(2)}_{0, \text{Student}(t)} + e_t
\end{aligned}
$$

$$
\begin{bmatrix} u^{(3)}_{1, \text{School}(t)} \\ u^{(3)}_{0, \text{School}(t)} \end{bmatrix} \sim N\left(0, \Omega_u^{(3)}\right) : \Omega_u^{(3)} = \begin{bmatrix} \Omega^{(3)}_{u0,0} & \\ \Omega^{(3)}_{u0,1} & \Omega^{(3)}_{u1,1} \end{bmatrix}
$$

$$
\begin{bmatrix} u^{(2)}_{1, \text{Student}(t)} \\ u^{(2)}_{0, \text{Student}(t)} \end{bmatrix} \sim N\left(0, \Omega_u^{(2)}\right) : \Omega_u^{(2)} = \begin{bmatrix} \Omega^{(2)}_{u0,0} & \\ \Omega^{(2)}_{u0,1} & \Omega^{(2)}_{u1,1} \end{bmatrix}
$$

$$
[e_t] \sim N\left(0, \Omega_e\right)
$$

The mathematical form of the model estimating the year-by-year impact of America's Choice was as follows.

$$
\begin{aligned}
y_t ={}& \beta_0 + \beta_1 \text{MALE} + \beta_2 \text{MINORITY} + \beta_3 \text{SPCED} + \beta_4 \text{LEP} + \beta_5 \text{FRL} \\
& + \beta_6 \text{AGE}' + \beta_7 (\text{AGE}' \times \text{MALE}) + \beta_8 (\text{AGE}' \times \text{MINORITY}) + \beta_9 (\text{AGE}' \times \text{SPCED}) \\
& + \beta_{10} (\text{AGE}' \times \text{LEP}) + \beta_{11} (\text{AGE}' \times \text{FRL}) \\
& + \sum_{year=1}^{5} \beta_{(11+year)} \text{IN AC}_{year,t} \\
& + \left(\text{AGE}' \times u^{(3)}_{1, \text{School}(t)}\right) + \left(\text{AGE}' \times u^{(2)}_{1, \text{Student}(t)}\right) + u^{(3)}_{0, \text{School}(t)} + u^{(2)}_{0, \text{Student}(t)} + e_t
\end{aligned}
$$

$$
\begin{bmatrix} u^{(3)}_{1, \text{School}(t)} \\ u^{(3)}_{0, \text{School}(t)} \end{bmatrix} \sim N\left(0, \Omega_u^{(3)}\right) : \Omega_u^{(3)} = \begin{bmatrix} \Omega^{(3)}_{u0,0} & \\ \Omega^{(3)}_{u0,1} & \Omega^{(3)}_{u1,1} \end{bmatrix}
$$

$$
\begin{bmatrix} u^{(2)}_{1, \text{Student}(t)} \\ u^{(2)}_{0, \text{Student}(t)} \end{bmatrix} \sim N\left(0, \Omega_u^{(2)}\right) : \Omega_u^{(2)} = \begin{bmatrix} \Omega^{(2)}_{u0,0} & \\ \Omega^{(2)}_{u0,1} & \Omega^{(2)}_{u1,1} \end{bmatrix}
$$

$$
[e_t] \sim N\left(0, \Omega_e\right)
$$

As mentioned previously, the age variable is centered around nine years of age and all control variables are centered around the grand mean for America's Choice schools. Therefore, β_0 shows the expected achievement score for a nine-year-old from a comparison school that has characteristics that are identical to the average America's Choice student. This is referred to as *baseline* in the tables in Appendices B through E. The estimate for β_6 shows the expected annual gain in achievement score for a student from a comparison school that has characteristics that are identical to the average America's Choice student. This is referred to as *annual change* in the tables in Appendices B through E. The estimates for $\beta_{12\text{-}16}$ show the additional gain in achievement score for an America's Choice student. This is referred to as *additional annual*

growth while in America's Choice or as the year-by-year *effect of America's Choice* in the tables in Appendices B through E.

The Bayesian estimation technique used here is a simulation-based method called Markov Chain Monte Carlo using Gibbs Sampling (Geman & Geman, 1984). The model was estimated via *MLwiN 2.0* (Browne, 2002; Rasbash, Browne, Healy, Cameron, & Charlton, 2003). Due to the complexity of the models, a long burn-in period of 10,000 iterations preceded a chain of 50,000 iterations in which plausible values were drawn from the posterior distributions of parameters at every fifth iteration, yielding 10,000 plausible values for each parameter. Point estimates and standard errors are calculated as the mean and standard deviation, respectively, of the plausible values from each posterior distribution.

Appendix B. Full Results for Models of Overall Effects of America's Choice on Growth in Individual Student Achievement

Table B1. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	594.9***	1.32
Gender (Male)	-2.1***	0.37
Minority (non-White)	-16.7***	0.54
Special Education	-37.7***	0.51
Limited English Proficient	7.9***	0.48
Free or Reduced-Priced Lunch	-22.9***	0.68
Annual Change in Reading Scores — Grades 1-3		
Intercept	29.7***	0.62
Gender (Male)	1.2***	0.27
Minority (non-White)	-0.5	0.35
Special Education	1.1***	0.32
Limited English Proficient	-1.9***	0.31
Free or Reduced-Priced Lunch	-0.7	0.45
Additional Annual Growth While in America's Choice	2.0***	0.34

Random Effect	Variance Component	
Baseline School Reading Scores	75.0	18.55
Change in School Reading Scores — Grades 1-3	15.9	4.14
Baseline Student Reading Scores	884.5	10.16
Change in Student Reading Scores — Grades 1-3	133.6	3.78
Residual within Growth Curve	551.4	4.16

$\tilde{}p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$

Table B2. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	617.0***	0.71
Gender (Male)	0.3	0.37
Minority (non-White)	-17.4***	0.53
Special Education	-33.3***	0.56
Limited English Proficient	11.9***	0.47
Free or Reduced-Priced Lunch	-20.8***	0.65
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.6***	0.25
Gender (Male)	-0.6***	0.09
Minority (non-White)	-0.2	0.13
Special Education	0.7***	0.13
Limited English Proficient	-0.6***	0.11
Free or Reduced-Priced Lunch	0.4**	0.15
Additional Annual Growth While in America's Choice	1.5***	0.14

Random Effect	Variance Component	
Baseline School Reading Scores	22.9	5.22
Change in School Reading Scores — Grades 4-8	3.0	0.68
Baseline Student Reading Scores	784.7	9.87
Change in Student Reading Scores — Grades 4-8	14.4	0.47
Residual within Growth Curve	248.7	1.36

*~p<.10, *p<.05, **p<.01, ***p<.001*

<p style="text-align:center; color:#8B0000;">Table B3. Multilevel Growth Curve Results of the Model Predicting
Baseline Mathematics Scores and Change in Mathematics Scores
from the First Grade to the Third Grade</p>

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	588.5***	1.58
Gender (Male)	3.3***	0.38
Minority (non-White)	-16.7***	0.55
Special Education	-42.7***	0.52
Limited English Proficient	3.2***	0.49
Free or Reduced-Priced Lunch	-17.0***	0.69
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	27.0***	0.72
Gender (Male)	0.4~	0.21
Minority (non-White)	-0.9**	0.32
Special Education	-0.8**	0.28
Limited English Proficient	0.2	0.29
Free or Reduced-Priced Lunch	-1.8***	0.40
Additional Annual Growth While in America's Choice	1.9***	.35

Random Effect	Variance Component	
Baseline School Mathematics Scores	90.2	21.88
Change in School Mathematics Scores — Grades 1-3	19.2	5.07
Baseline Student Mathematics Scores	913.6	10.31
Change in Student Mathematics Scores — Grades 1-3	25.1	2.34
Residual within Growth Curve	588.6	4.17

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table B4. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	624.0***	1.16
Gender (Male)	2.6***	0.42
Minority (non-White)	-18.4***	0.58
Special Education	-40.5***	0.61
Limited English Proficient	7.1***	0.52
Free or Reduced-Priced Lunch	-18.1***	0.71
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	6.2***	0.41
Gender (Male)	-0.3**	0.09
Minority (non-White)	0.1	0.13
Special Education	3.6***	0.13
Limited English Proficient	-0.4**	0.11
Free or Reduced-Priced Lunch	0.8***	0.16
Additional Annual Growth While in America's Choice	1.6***	0.16

Random Effect	Variance Component	
Baseline School Mathematics Scores	57.2	12.58
Change in School Mathematics Scores — Grades 4-8	7.7	1.70
Baseline Student Mathematics Scores	891.0	11.81
Change in Student Mathematics Scores — Grades 4-8	9.1	0.45
Residual within Growth Curve	297.7	1.55

~$p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$

Appendix C. Full Results for Models of Year-by-Year Effects of America's Choice on Student Achievement

Table C1. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade and Year-by-Year Effects of America's Choice

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	594.9***	1.33
Gender (Male)	-2.1***	0.38
Minority (non-White)	-16.7***	0.53
Special Education	-37.8***	0.51
Limited English Proficient	7.9***	0.48
Free or Reduced-Priced Lunch	-23.0***	0.68
Annual Change in Reading Scores — Grades 1-3		
Intercept	29.7***	0.65
Gender (Male)	1.2***	0.24
Minority (non-White)	-0.6~	0.35
Special Education	1.1***	0.32
Limited English Proficient	-1.9***	0.31
Free or Reduced-Priced Lunch	-0.7	0.44
Effect of America's Choice		
Year 1	2.2***	0.61
Year 2	-0.5	0.68
Year 3	2.6**	0.86
Year 4	4.6***	0.92
Year 5	-1.9~	1.02

Random Effect	Variance Component	
Baseline School Reading Scores	74.5	18.41
Change in School Reading Scores — Grades 1-3	15.9	4.14
Baseline Student Reading Scores	884.0	10.12
Change in Student Reading Scores — Grades 1-3	134.1	3.70
Residual within Growth Curve	551.2	4.16

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table C2. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade and Year-by-Year Effects of America's Choice

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	617.0***	0.71
Gender (Male)	0.3	0.37
Minority (non-White)	-17.4***	0.54
Special Education	-33.3***	0.56
Limited English Proficient	11.8***	0.48
Free or Reduced-Priced Lunch	-20.9***	0.65
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.6***	0.26
Gender (Male)	-0.6***	0.08
Minority (non-White)	-0.2	0.13
Special Education	0.7***	0.13
Limited English Proficient	-0.6***	0.11
Free or Reduced-Priced Lunch	0.4**	0.15
Effect of America's Choice		
Year 1	2.8***	0.29
Year 2	1.6***	0.33
Year 3	-0.3	0.62
Year 4	2.6***	0.69
Year 5	-1.4~	0.73

Random Effect	Variance Component	
Baseline School Reading Scores	23.4	5.31
Change in School Reading Scores — Grades 4-8	3.0	0.68
Baseline Student Reading Scores	785.2	10.01
Change in Student Reading Scores — Grades 4-8	14.4	0.47
Residual within Growth Curve	248.6	1.37

*~p<.10, *p<.05, **p<.01, ***p<.001*

**Table C3. Multilevel Growth Curve Results of the Model Predicting
Baseline Mathematics Scores and Change in Mathematics Scores
from the First Grade to the Third Grade
and Year-by-Year Effects of America's Choice**

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	588.5***	1.45
Gender (Male)	3.3***	0.38
Minority (non-White)	-16.8***	0.55
Special Education	-42.8***	0.51
Limited English Proficient	3.1***	0.50
Free or Reduced-Priced Lunch	-17.1***	0.69
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	27.0***	0.70
Gender (Male)	0.4~	0.21
Minority (non-White)	-1.0**	0.32
Special Education	-0.8**	0.28
Limited English Proficient	0.2	0.29
Free or Reduced-Priced Lunch	-1.8***	0.40
Effect of America's Choice		
Year 1	3.3***	0.62
Year 2	-0.8	0.67
Year 3	2.7**	0.85
Year 4	1.9*	0.90
Year 5	0.2	0.99

Random Effect	Variance Component	
Baseline School Mathematics Scores	89.5	21.86
Change in School Mathematics Scores — Grades 1-3	19.3	5.07
Baseline Student Mathematics Scores	913.2	10.36
Change in Student Mathematics Scores — Grades 1-3	25.1	2.33
Residual within Growth Curve	588.9	4.15

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table C4. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade and Year-by-Year Effects of America's Choice

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	624.1***	1.15
Gender (Male)	2.6***	0.42
Minority (non-White)	-18.4***	0.58
Special Education	-40.5***	0.61
Limited English Proficient	7.1***	0.52
Free or Reduced-Priced Lunch	-18.1***	0.72
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	6.0***	0.41
Gender (Male)	-0.3***	0.09
Minority (non-White)	0.1	0.13
Special Education	3.6***	0.13
Limited English Proficient	-0.4***	0.11
Free or Reduced-Priced Lunch	0.8***	0.16
Effect of America's Choice		
Year 1	2.9***	0.31
Year 2	1.2**	0.37
Year 3	-0.4	0.64
Year 4	2.5***	0.73
Year 5	-0.7	0.85

Random Effect	Variance Component	
Baseline School Mathematics Scores	57.1	12.54
Change in School Mathematics Scores — Grades 4-8	7.8	1.70
Baseline Student Mathematics Scores	892.7	11.58
Change in Student Mathematics Scores — Grades 4-8	9.2	0.45
Residual within Growth Curve	297.3	1.57

*~p<.10, *p<.05, **p<.01, ***p<.001*

Appendix D. Full Results for Models of Effects of America's Choice on Growth in Individual Student Achievement by Initial Performance Quartile

Table D1. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade — Quartile 1

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	566.2***	1.00
Gender (Male)	-1.1~	0.62
Minority (non-White)	-3.7***	1.04
Special Education	-18.9***	0.65
Limited English Proficient	8.9***	0.76
Free or Reduced-Priced Lunch	-1.2	1.66
Annual Change in Reading Scores — Grades 1-3		
Intercept	38.0***	0.63
Gender (Male)	-0.5	0.43
Minority (non-White)	-2.0**	0.73
Special Education	-2.7***	0.46
Limited English Proficient	-0.1	0.53
Free or Reduced-Priced Lunch	-0.8	1.11
Additional Annual Growth While in America's Choice	3.6***	0.57

Random Effect	Variance Component	
Baseline School Reading Scores	31.9	8.95
Change in School Reading Scores — Grades 1-3	12.2	3.78
Baseline Student Reading Scores	372.6	14.72
Change in Student Reading Scores — Grades 1-3	67.5	5.95
Residual within Growth Curve	832.6	11.46

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D2. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade — Quartile 2

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	585.0***	0.83
Gender (Male)	-0.6	0.60
Minority (non-White)	-4.6***	0.93
Special Education	-20.9***	0.80
Limited English Proficient	1.1	0.80
Free or Reduced-Priced Lunch	-5.5***	-1.31
Annual Change in Reading Scores — Grades 1-3		
Intercept	31.5***	0.37
Gender (Male)	0.2	0.33
Minority (non-White)	-2.0***	0.51
Special Education	-7.0***	0.43
Limited English Proficient	0.0	0.44
Free or Reduced-Priced Lunch	-2.9***	0.73
Additional Annual Growth While in America's Choice	-0.5	0.48

Random Effect	Variance Component	
Baseline School Reading Scores	23.7	6.60
Change in School Reading Scores — Grades 1-3	4.2	1.31
Baseline Student Reading Scores	528.1	12.85
Change in Student Reading Scores — Grades 1-3	55.3	2.49
Residual within Growth Curve	364.6	4.36

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D3. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade — Quartile 3

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	602.3***	0.91
Gender (Male)	0.0	0.60
Minority (non-White)	-6.6***	0.85
Special Education	-24.3***	1.05
Limited English Proficient	1.1	0.81
Free or Reduced-Priced Lunch	-9.8***	1.10
Annual Change in Reading Scores — Grades 1-3		
Intercept	27.3***	0.39
Gender (Male)	0.8*	0.34
Minority (non-White)	-2.9***	0.48
Special Education	-8.8***	0.58
Limited English Proficient	0.4	0.46
Free or Reduced-Priced Lunch	-4.4***	0.62
Additional Annual Growth While in America's Choice	-0.6	0.51

Random Effect	Variance Component	
Baseline School Reading Scores	26.1	7.46
Change in School Reading Scores — Grades 1-3	4.3	1.35
Baseline Student Reading Scores	508.2	12.18
Change in Student Reading Scores — Grades 1-3	54.4	2.38
Residual within Growth Curve	338.2	4.09

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D4. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade — Quartile 4

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	627.6***	1.08
Gender (Male)	-0.3	0.64
Minority (non-White)	-14.7***	0.81
Special Education	-24.2***	1.61
Limited English Proficient	0.9	0.84
Free or Reduced-Priced Lunch	-16.8***	0.90
Annual Change in Reading Scores — Grades 1-3		
Intercept	17.5***	0.54
Gender (Male)	0.6	0.40
Minority (non-White)	-3.4***	0.51
Special Education	-5.7***	1.03
Limited English Proficient	-1.2*	0.53
Free or Reduced-Priced Lunch	-2.8***	0.56
Additional Annual Growth While in America's Choice	-1.2~	0.63

Random Effect	Variance Component	
Baseline School Reading Scores	38.2	10.60
Change in School Reading Scores — Grades 1-3	8.0	2.41
Baseline Student Reading Scores	514.3	14.01
Change in Student Reading Scores — Grades 1-3	49.1	5.74
Residual within Growth Curve	472.2	7.09

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D5. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade — Quartile 1

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	596.7***	0.96
Gender (Male)	0.4	0.71
Minority (non-White)	-1.7	1.20
Special Education	-18.2***	0.79
Limited English Proficient	16.2***	0.86
Free or Reduced-Priced Lunch	-0.4	1.78
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.7***	0.37
Gender (Male)	-0.3~	0.17
Minority (non-White)	-0.5~	0.30
Special Education	0.5**	0.19
Limited English Proficient	-1.0***	0.21
Free or Reduced-Priced Lunch	0.4	0.41
Additional Annual Growth While in America's Choice	3.0***	0.25

Random Effect	Variance Component	
Baseline School Reading Scores	30.7	8.17
Change in School Reading Scores — Grades 4-8	5.0	1.26
Baseline Student Reading Scores	557.7	18.63
Change in Student Reading Scores — Grades 4-8	17.1	1.08
Residual within Growth Curve	290.3	3.15

*~p<.10, *p<.05, **p<.01, ***p<.001*

**Table D6. Multilevel Growth Curve Results of the Model Predicting
Baseline Reading Scores and Change in Reading Scores
from the Fourth Grade to the Eighth Grade — Quartile 2**

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	610.8***	0.74
Gender (Male)	1.6**	0.59
Minority (non-White)	-6.5***	0.94
Special Education	-15.0***	0.90
Limited English Proficient	1.0	0.79
Free or Reduced-Priced Lunch	-5.6***	1.28
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.6***	0.27
Gender (Male)	-0.8***	0.15
Minority (non-White)	-0.1	0.25
Special Education	0.1	0.24
Limited English Proficient	-0.1	0.20
Free or Reduced-Priced Lunch	0.6~	0.31
Additional Annual Growth While in America's Choice	1.7***	0.22

Random Effect	Variance Component	
Baseline School Reading Scores	18.9	4.85
Change in School Reading Scores — Grades 4-8	2.6	0.67
Baseline Student Reading Scores	337.2	11.66
Change in Student Reading Scores — Grades 4-8	10.3	0.75
Residual within Growth Curve	231.3	2.48

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D7. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade — Quartile 3

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	623.9***	0.78
Gender (Male)	1.5**	0.57
Minority (non-White)	-7.6***	0.83
Special Education	-17.9***	1.17
Limited English Proficient	-0.6	0.79
Free or Reduced-Priced Lunch	-7.7***	1.04
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.4***	0.26
Gender (Male)	-0.6***	0.15
Minority (non-White)	0.1	0.22
Special Education	-0.1	0.31
Limited English Proficient	-0.2	0.20
Free or Reduced-Priced Lunch	-0.1	0.27
Additional Annual Growth While in America's Choice	0.3	0.23

Random Effect	Variance Component	
Baseline School Reading Scores	21.4	5.46
Change in School Reading Scores — Grades 4-8	2.1	0.56
Baseline Student Reading Scores	355.6	10.57
Change in Student Reading Scores — Grades 4-8	11.4	0.72
Residual within Growth Curve	215.2	2.30

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D8. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade — Quartile 4

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	646.3***	1.03
Gender (Male)	0.8	0.67
Minority (non-White)	-13.0***	0.81
Special Education	-13.4***	2.03
Limited English Proficient	-2.0*	0.90
Free or Reduced-Priced Lunch	-14.5***	0.91
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.2***	0.32
Gender (Male)	-0.6***	0.17
Minority (non-White)	-0.3	0.21
Special Education	-1.6**	0.55
Limited English Proficient	-0.1	0.24
Free or Reduced-Priced Lunch	-0.6*	0.24
Additional Annual Growth While in America's Choice	-0.3	0.30

Random Effect	Variance Component	
Baseline School Reading Scores	35.1	8.69
Change in School Reading Scores — Grades 4-8	3.2	0.86
Baseline Student Reading Scores	561.5	14.83
Change in Student Reading Scores — Grades 4-8	14.6	0.94
Residual within Growth Curve	243.6	2.67

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D9. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the First Grade to the Third Grade — Quartile 1

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	557.0***	1.20
Gender (Male)	1.5*	0.63
Minority (non-White)	-4.8***	1.10
Special Education	-25.5***	0.67
Limited English Proficient	6.3***	0.84
Free or Reduced-Priced Lunch	0.3	1.61
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	30.7***	0.50
Gender (Male)	1.3***	0.37
Minority (non-White)	-1.2~	0.67
Special Education	-1.9***	0.40
Limited English Proficient	1.7***	0.51
Free or Reduced-Priced Lunch	0.4	0.99
Additional Annual Growth While in America's Choice	4.4***	0.55

Random Effect	Variance Component	
Baseline School Mathematics Scores	46.0	12.58
Change in School Mathematics Scores — Grades 1-3	7.1	2.21
Baseline Student Mathematics Scores	525.6	14.88
Change in Student Mathematics Scores — Grades 1-3	25.6	2.32
Residual within Growth Curve	666.0	7.89

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D10. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the First Grade to the Third Grade — Quartile 2

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	582.7***	0.91
Gender (Male)	0.6	0.61
Minority (non-White)	-5.0***	0.96
Special Education	-21.8***	0.85
Limited English Proficient	0.5	0.80
Free or Reduced-Priced Lunch	-5.1***	1.27
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	26.4***	0.38
Gender (Male)	0.2	0.36
Minority (non-White)	-1.8**	0.56
Special Education	-7.8***	0.51
Limited English Proficient	0.3	0.48
Free or Reduced-Priced Lunch	-3.1***	0.73
Additional Annual Growth While in America's Choice	0.8	0.51

Random Effect	Variance Component	
Baseline School Mathematics Scores	27.7	7.53
Change in School Mathematics Scores — Grades 1-3	4.2	1.31
Baseline Student Mathematics Scores	498.7	13.17
Change in Student Mathematics Scores — Grades 1-3	66.6	2.81
Residual within Growth Curve	403.0	4.92

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D11. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the First Grade to the Third Grade — Quartile 3

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	600.4***	1.02
Gender (Male)	0.4	0.62
Minority (non-White)	-7.6***	0.85
Special Education	-23.6***	1.04
Limited English Proficient	-1.0	0.80
Free or Reduced-Priced Lunch	-9.1***	1.10
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	24.3***	0.50
Gender (Male)	0.3	0.37
Minority (non-White)	-3.1***	0.51
Special Education	-8.6***	0.63
Limited English Proficient	-0.3	0.48
Free or Reduced-Priced Lunch	-4.5***	0.65
Additional Annual Growth While in America's Choice	-0.6	0.52

Random Effect	Variance Component	
Baseline School Mathematics Scores	35.7	9.69
Change in School Mathematics Scores — Grades 1-3	8.1	2.37
Baseline Student Mathematics Scores	516.8	13.00
Change in Student Mathematics Scores — Grades 1-3	72.8	3.17
Residual within Growth Curve	369.6	4.52

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D12. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the First Grade to the Third Grade — Quartile 4

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	621.3***	1.18
Gender (Male)	1.4*	0.64
Minority (non-White)	-9.9***	0.79
Special Education	-23.6***	1.46
Limited English Proficient	0.3	0.81
Free or Reduced-Priced Lunch	-12.5***	0.91
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	18.7***	0.65
Gender (Male)	0.5	0.40
Minority (non-White)	-2.3***	0.50
Special Education	-8.3***	0.91
Limited English Proficient	-0.5	0.51
Free or Reduced-Priced Lunch	-3.5***	0.57
Additional Annual Growth While in America's Choice	0.2	0.65

Random Effect	Variance Component	
Baseline School Mathematics Scores	50.0	13.38
Change in School Mathematics Scores — Grades 1-3	14.3	4.06
Baseline Student Mathematics Scores	479.9	13.40
Change in Student Mathematics Scores — Grades 1-3	38.2	2.60
Residual within Growth Curve	469.4	5.90

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D13. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade — Quartile 1

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	602.2***	1.11
Gender (Male)	-0.6	0.79
Minority (non-White)	-4.0**	1.35
Special Education	-25.0***	0.88
Limited English Proficient	11.4***	0.97
Free or Reduced-Priced Lunch	1.0	1.92
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	6.3***	0.32
Gender (Male)	0.1	0.18
Minority (non-White)	-0.3	0.32
Special Education	3.0***	0.20
Limited English Proficient	-0.8***	0.22
Free or Reduced-Priced Lunch	0.1	0.42
Additional Annual Growth While in America's Choice	2.4***	0.26

Random Effect	Variance Component	
Baseline School Mathematics Scores	42.4	10.82
Change in School Mathematics Scores — Grades 4-8	3.4	1.00
Baseline Student Mathematics Scores	705.7	21.57
Change in Student Mathematics Scores — Grades 4-8	17.8	1.10
Residual within Growth Curve	344.9	3.51

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D14. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade — Quartile 2

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	621.4***	1.00
Gender (Male)	0.9	0.63
Minority (non-White)	-6.2***	1.01
Special Education	-15.8***	1.00
Limited English Proficient	3.0***	0.81
Free or Reduced-Priced Lunch	-3.3**	1.26
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	5.4***	0.35
Gender (Male)	-0.4**	0.15
Minority (non-White)	0.1	0.24
Special Education	1.3***	0.24
Limited English Proficient	-0.2	0.19
Free or Reduced-Priced Lunch	-0.2	0.29
Additional Annual Growth While in America's Choice	1.8***	0.24

Random Effect	Variance Component	
Baseline School Mathematics Scores	39.3	9.49
Change in School Mathematics Scores — Grades 4-8	4.9	1.21
Baseline Student Mathematics Scores	368.0	12.41
Change in Student Mathematics Scores — Grades 4-8	5.1	0.66
Residual within Growth Curve	263.2	2.65

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D15. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade — Quartile 3

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	633.7***	1.23
Gender (Male)	0.9	0.64
Minority (non-White)	-6.4***	0.88
Special Education	-20.9***	1.24
Limited English Proficient	-0.3	0.82
Free or Reduced-Priced Lunch	-8.1***	1.12
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	5.8***	0.46
Gender (Male)	-0.1	0.15
Minority (non-White)	-0.5*	0.21
Special Education	1.8***	0.29
Limited English Proficient	0.0	0.19
Free or Reduced-Priced Lunch	0.4	0.27
Additional Annual Growth While in America's Choice	0.6*	0.27

Random Effect	Variance Component	
Baseline School Mathematics Scores	64.4	14.79
Change in School Mathematics Scores — Grades 4-8	9.0	2.11
Baseline Student Mathematics Scores	393.0	12.74
Change in Student Mathematics Scores — Grades 4-8	3.2	0.68
Residual within Growth Curve	253.8	2.65

*~p<.10, *p<.05, **p<.01, ***p<.001*

Table D16. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade — Quartile 4

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	652.0***	1.81
Gender (Male)	1.1	0.72
Minority (non-White)	-10.9***	0.86
Special Education	-22.6***	1.88
Limited English Proficient	-3.2***	0.91
Free or Reduced-Priced Lunch	-14.3***	0.98
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	6.0***	0.67
Gender (Male)	0.1	0.17
Minority (non-White)	-0.6**	0.21
Special Education	0.4	0.45
Limited English Proficient	0.3	0.23
Free or Reduced-Priced Lunch	-0.3	0.24
Additional Annual Growth While in America's Choice	1.4***	0.38

Random Effect	Variance Component	
Baseline School Mathematics Scores	147.2	32.69
Change in School Mathematics Scores — Grades 4-8	21.4	4.76
Baseline Student Mathematics Scores	528.4	16.26
Change in Student Mathematics Scores — Grades 4-8	3.9	0.87
Residual within Growth Curve	291.7	3.19

*~p<.10, *p<.05, **p<.01, ***p<.001*

Appendix E. Full Results for Models of Overall Effects of America's Choice on Growth in Individual Student Achievement by Ethnicity

Table E1. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the First Grade to the Third Grade by Ethnicity

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	594.9***	1.32
Gender (Male)	-2.1***	0.38
Minority (non-White)	-17.0***	0.55
Special Education	-37.7***	0.51
Limited English Proficient	8.2***	0.49
Free or Reduced-Priced Lunch	-22.9***	0.68
Annual Change in Reading Scores — Grades 1-3		
Intercept	29.6***	0.66
Gender (Male)	1.2***	0.24
Minority (non-White)	-0.7*	0.35
Special Education	1.1***	0.32
Limited English Proficient	-1.8***	0.31
Free or Reduced-Priced Lunch	-0.7	0.44
Additional Annual Growth While in America's Choice		
White Students	-0.4	1.00
African American Students	1.2**	0.39
Hispanic Students	3.6***	0.64
Other Students[A]	8.0***	1.96

Random Effect	Variance Component	
Baseline School Reading Scores	74.4	18.3
Change in School Reading Scores — Grades 1-3	15.9	4.1
Baseline Student Reading Scores	883.8	10.1
Change in Student Reading Scores — Grades 1-3	133.4	3.7
Residual within Growth Curve	551.6	4.18

$\tilde{}p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$
[A] *Results not presented in main text due to small sample size (< 3% of sample)*

Table E2. Multilevel Growth Curve Results of the Model Predicting Baseline Reading Scores and Change in Reading Scores from the Fourth Grade to the Eighth Grade by Ethnicity

Fixed Effect	Estimate	Standard Error
Baseline Reading Scores		
Intercept	617.0***	0.71
Gender (Male)	0.3	0.37
Minority (non-White)	-17.4***	0.52
Special Education	-33.3***	0.56
Limited English Proficient	11.9***	0.48
Free or Reduced-Priced Lunch	-20.8***	0.65
Annual Change in Reading Scores — Grades 4-8		
Intercept	8.6***	0.25
Gender (Male)	-0.6***	0.09
Minority (non-White)	-0.2	0.13
Special Education	0.7***	0.13
Limited English Proficient	-0.6***	0.11
Free or Reduced-Priced Lunch	0.4**	0.15
Additional Annual Growth While in America's Choice		
White Students	0.8~	0.45
African American Students	1.2***	0.16
Hispanic Students	2.4***	0.30
Other Students[A]	6.4***	0.98

Random Effect	Variance Component	
Baseline School Reading Scores	22.7	5.16
Change in School Reading Scores — Grades 4-8	3.0	0.68
Baseline Student Reading Scores	784.5	10.01
Change in Student Reading Scores — Grades 4-8	14.5	0.47
Residual within Growth Curve	248.6	1.36

~$p<.10$, *$p<.05$, **$p<.01$, ***$p<.001$
[A] Results not presented in main text due to small sample size (< 3% of sample)

Table E3. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the First Grade to the Third Grade by Ethnicity

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	588.3***	1.41
Gender (Male)	3.3***	0.38
Minority (non-White)	-17.1***	0.56
Special Education	-42.7***	0.52
Limited English Proficient	3.5***	0.50
Free or Reduced-Priced Lunch	-17.1***	0.68
Annual Change in Mathematics Scores — Grades 1-3		
Intercept	26.9***	0.69
Gender (Male)	0.4~	0.21
Minority (non-White)	-1.1***	0.32
Special Education	-0.8**	0.28
Limited English Proficient	0.3	0.29
Free or Reduced-Priced Lunch	-1.8***	0.40
Additional Annual Growth While in America's Choice		
White Students	-0.4	1.03
African American Students	1.3***	0.39
Hispanic Students	3.1***	0.72
Other Students[A]	18.1***	2.59

Random Effect	Variance Component	
Baseline School Mathematics Scores	89.6	21.66
Change in School Mathematics Scores — Grades 1-3	19.2	5.04
Baseline Student Mathematics Scores	911.4	10.33
Change in Student Mathematics Scores — Grades 1-3	25.0	2.27
Residual within Growth Curve	589.0	4.16

~p<.10, *p<.05, **p<.01, ***p<.001
[A] Results not presented in main text due to small sample size (< 3% of sample)

Table E4. Multilevel Growth Curve Results of the Model Predicting Baseline Mathematics Scores and Change in Mathematics Scores from the Fourth Grade to the Eighth Grade by Ethnicity

Fixed Effect	Estimate	Standard Error
Baseline Mathematics Scores		
Intercept	624.2***	1.11
Gender (Male)	2.6***	0.41
Minority (non-White)	-18.4***	0.59
Special Education	-40.5***	0.61
Limited English Proficient	7.2***	0.51
Free or Reduced-Priced Lunch	-18.1***	0.72
Annual Change in Mathematics Scores — Grades 4-8		
Intercept	6.1***	0.42
Gender (Male)	-0.3***	0.09
Minority (non-White)	0.1	0.13
Special Education	3.6***	0.13
Limited English Proficient	-0.4***	0.11
Free or Reduced-Priced Lunch	0.8***	0.16
Additional Annual Growth While in America's Choice		
White Students	1.3**	0.49
African American Students	1.1***	0.18
Hispanic Students	1.6***	0.33
Other Students[A]	11.0***	1.11

Random Effect	Variance Component	
Baseline School Mathematics Scores	57.8	12.66
Change in School Mathematics Scores — Grades 4-8	7.6	1.66
Baseline Student Mathematics Scores	891.5	11.68
Change in Student Mathematics Scores — Grades 4-8	9.2	0.44
Residual within Growth Curve	297.6	1.55

˜p<.10, *p<.05, **p<.01, ***p<.001
[A] Results not presented in main text due to small sample size (< 3% of sample)