EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Volume 12 Number 49 September 18, 2004 ISSN 1068-2341

Age and Achievement*

James B. Grissom
California Department of Education

Citation: Grissom, J. B. (2004, September 18). Age and achievement. Education Policy Analysis Archives, 12(49). Retrieved [date] from http://epaa.asu.edu/epaa/v12n49/.

Abstract
There is continuing controversy about the optimal or appropriate age at which children should start school. The purpose of this study is to examine the relationship between age and achievement. It is an attempt to evaluate the hypothesis that older students fare better academically than their younger classmates. Findings indicate that on average for students in elementary school there is positive linear relationship between age and achievement *for age normal peers*. Even though there is positive linear relationship, the difference in average test scores between the oldest and youngest students is not great and by the time students reach 10th grade the positive linear relationship has disappeared. For overage students there is on average a negative linear relationship between age and achievement at all grade levels. That is, the negative relationship between age and achievement remains constant over time. These results argue against modifying entrance age policies, delaying school entry, implementing transitional kindergarten or first grade programs or retaining students to improve educational achievement. Policies and practices that make students older than their classmates inversely affect their educational achievement.

*The opinions expressed are of the author alone and do not reflect opinion or policy of the California Department of Education.

There is continuing controversy about the optimal or appropriate age at which children should start school. The efficacy of delaying school entry beyond the age that a student can legally enroll in public school has been debated in the research literature. Crosser (1991), Kinard & Reinherz (1986), and La Paro & Pianta (2000) present evidence that older children fare better academically than their younger, age appropriate peers. Uphoff & Gilmore (1985) use research evidence about the relationship between age and achievement as well as other evidence to argue that the older and/or more mature students in a class fare better than younger classmates. In contrast DeMeis & Stearns (1992) and Dietz & Wilson (1985) found no significant relationship between age and achievement. Langer, Kalk, & Searls (1984) found significantly higher achievement of the oldest as compared to the youngest students at age nine but this difference disappeared by age seventeen. Shepard (1997) argues that even if more emotionally mature children do better in school, there are no valid instruments or means to identify these children. The most popular readiness tests (e.g., Light's Retention Scale, the Gesell School Readiness Test, the Gesell Preschool Test, the Brigance K & 1 Screen, the Daberon Screening for School Readiness, the Developmental Indicators for the Assessment of Learning-Revised, and the Missouri Kindergarten Inventory of Developmental Skills) all lack demonstrated validity and reliability to make readiness decisions about individual children. Meisels (1992) argues that when parents practice delayed school entry, younger, legally enrolled students may be disadvantaged. When first graders who are barely 6 years old are compared to 7½-year olds, the 6 year olds, functioning at a developmentally appropriate level, seem immature. Immaturity is one of the major reasons given for grade retention (Abidin, Golladay, & Howerton, 1971; Niklason, 1984). The controversy continues despite the fact that no matter where the legal entry age is set there will be older and younger children (i.e., relative to each other) in a class.

The incidence of delayed school entry and transitional kindergarten and first grade programs has been fueled by the belief among parents and educators that the curriculum is too difficult (Shepard & Smith, 1988). The academic expectations at each grade level have been raised and so the curriculum is being pushed down. That is, what was once expected of students in first grade is now expected of students in kindergarten. As a means to "protect" children from the "more demanding" curriculum some educators recommend (Uphoff & Gilmore, 1985) and some parents practice delayed school entry (Graue & DiPerna, 2000). It is thought that older more mature children are better prepared developmentally (i.e., both cognitively and emotionally) to handle the more rigorous curriculum.

In the current climate of accountability teachers are being held responsible for student outcomes on standardized tests. Teachers are also responsible for making sure students have mastered the material in the current grade so that students are prepared to master material in the next grade. The result is that teachers want more homogenous and cognitively advanced classrooms *so that the outcomes for which they are held accountable can be more efficiently produced* (Smith & Shepard, 1987). In their own self-interest, teachers want classrooms of students that will be successful on standardized tests. Teachers therefore favor policies and practices (e.g., delayed school entry, changing the legal age of entry, and transitional kindergarten programs) they believe produce such classrooms. The negative education consequences of homogeneous grouping and tracking have been described in the research literature. *Homogeneous ability grouping enhances the achievement ability of the fast group and retards the achievement of the slower group* (Smith & Shepard, 1987). These consequences run counter to the democratic goals of education in the United States. In addition, as long as classrooms are made up of 20 to 30+ students, efforts to reduce heterogeneity will largely be futile. There will always be variability in terms of educational achievement.

There is a strong belief among parents and educators that grade retention allows children time to mature cognitively to handle the more rigorous curriculum (Byrnes, 1989; Combs & Tanner, 1993; Smith, 1989). It is also seen by teachers as another way to reduce heterogeneity. In spite of these beliefs, research on grade retention indicates that retained students do less well academically when compared to recommended but not retained students (Holmes, 1989; Holmes & Matthews, 1984). Even so, grade retention remains an accepted academic intervention to raise student achievement. One certain effect of grade retention is that it makes students older than their grade level peers.

Angrist & Krueger (1992) tested the hypothesis that there is inverse relationship between educational attainment and age at school entry. That is, students who enter school at an older age drop out after having completed less schooling (i.e., because they are legally able to do so) than students who enter school at a younger age. The argument by Angrist & Krueger (1992) is that leaving early reduces the number of years of schooling and thus educational attainment. Grissom & Shepard (1989) present evidence that retained and/or overage students are more likely to drop out of school than students not retained and/or not overage after controlling for achievement differences. Policies and practices that make students older than their classmates increase the likelihood that these students will leave school early.

Most students will be older than their age normal peers for three reasons: they started school late, spent two years in a transitional kindergarten or first grade program, or were "flunked" and forced to repeat a grade. The belief remains strong that these three academic interventions benefit students academically and in other ways, despite evidence to the contrary. The belief is so strong that there are laws that mandate grade retention as the preferred remediation for low-achieving students. For example, California Education Code, Section 48070.5 (d) states that,

> If … a pupil is performing below the minimum standard for promotion, the pupil shall be retained in his or her current grade level unless the pupil's regular classroom teacher determines in writing that retention is not the appropriate intervention for the pupil's academic deficiencies.

The major purpose of this study is to examine the relationship between age and achievement. It evaluates the hypothesis that older students fare better academically (e.g., score higher on standardized tests) than their younger classmates and it also evaluates the hypothesis that children are protected (i.e., from an unrealistic and harsh curriculum) and benefit from delayed school entry, transitional programs, and/or grade retention.

A secondary purpose is to present some evidence on the extent of academic "red-shirting" in California public schools. Researchers sometimes refer to the practice of delaying school entry as academic "red-shirting." Although available data cannot identify why students are overage, an examination of age distributions may allow some inferences.

Researchers (Brent, May, & Kundert, 1996; Bracy, 1989) have reported on an increasing trend to delay school entry for age-eligible children. Though Bellisimo, Sacks, & Mergendoller, (1995) reported a drop in this trend in California between 1989 and 1991, there is reason to believe that that academic "red-shirting" is on the rise due to the increasing curriculum demands of kindergarten and first grade (Graue & DiPerna, 2000). Researchers also indicate that males are more likely to experience delayed school entry than females (Bellisimo, Sacks, & Mergendoller, 1995; Brent, May, & Kundert, 1996; Graue & DiPerna, 2000; May & Kundert, 1995) and that parents identified as having higher socio-economic status (SES) are more likely to delay school entry than those parents identified as lower SES (Bellisimo, Sacks, & Mergendoller, 1995).

Most studies of entrance age practices are based on data collected at the district level, which limits the ability to generalize to larger populations. In contrast, Langer, Kalk, & Searls, (1984) had nationally representative data (i.e., NAEP) but had to infer school entry practices based on student age and school entry policies. Graue & DiPerna (2000) addressed weaknesses of earlier studies by using sample selection strategies that allowed them to collect data as to when students started school and make inferences at the state level (i.e., Wisconsin).

This study is based on data collected on students enrolled in California public schools in grades 2 through 11 and suffers the limitations of Langer, Kalk, & Searls, (1984). That is, data were not collected until second grade and so entrance age practices have to be inferred from age distributions and statewide entrance age policies.

Method

Each spring California administers a series of standardized achievement tests known as the Standardized Testing and Reporting (STAR) program. Tests are administered to all public school students enrolled in grades 2 through 11. As part of the testing program, demographic information, including birth date, is collected. STAR tests were administered first in the spring of 1998. From 1998 to 2002 a norm-referenced standardized test, the Stanford Achievement Test version 9 (SAT/9) form T, was administered as part of the STAR program. This study uses data from tests administered in the spring of 1998 through 2002.

Students in California need to be five years old by December 2 to enroll in kindergarten or six years old by December 2 to enroll in first grade. For students tested in 2002, student age on December 1, 2001 (i.e., the December before spring testing) was calculated in months. As stated, the youngest students tested in spring 2002 were second graders. The youngest second graders were students who turned 7 close to the cutoff date of December 2. The age of the youngest second graders was 85 months[1]. Second grade students who were 85 months

---

[1] Students who were $7 \times 12 = 84$ months old were students who turned seven on December 1. Although these were the youngest students in the cohort, they were not included in the analyses. The study concerns test scores by month. The youngest students born in December would only include students born on one day (i.e., December 1). The number of second grade students born on December 1, 1994 and tested in spring 2002 was 856. The total number of students tested in spring 2002 was 485,796. Including this small number as a unique category is not helpful in making overall inferences. Adding students born on one day in December to those born in November is not going to change the general relationship

old were students with November 1994 birth dates. Second grade students who were 86 months old were students with October 1994 birth dates and so on.

Second grade students who were 96 months old were students with December 1993 birth dates. Students with December 1993 birth dates were the oldest age normal peers[2].

Second grade students who were 97 months or older had been retained. Student demographic information does not contain information as to why students were held back. That is, there is no information as to whether students started school late (i.e., academic "red-shirting"), spent two years in a transitional kindergarten or first grade program, or were retained in grade (i.e., "flunked") in kindergarten or first grade. Students who were 97 months old have November 1993 birth dates. These were the retained (i.e., held back for one of the three stated reasons) students whose birth dates are closest to the cut-off date. Students who were 98 months old were the retained students with October 1993 birth dates and so on.

For each age (in months) the average SAT/9 total reading and mathematics scores were calculated in normal curve equivalent (NCE) units. The average test score by age provides an indicator of the relationship between age and achievement.

Given the errors in self-report data and the desire to avoid discussion about students who are very young or very old relative to their age-normal peer group, the full age range was truncated. For second grade students the age range was truncated to students who were 85 to109 months old. The oldest students are 24 months or two years older than the youngest students.

Results

Age and Achievement

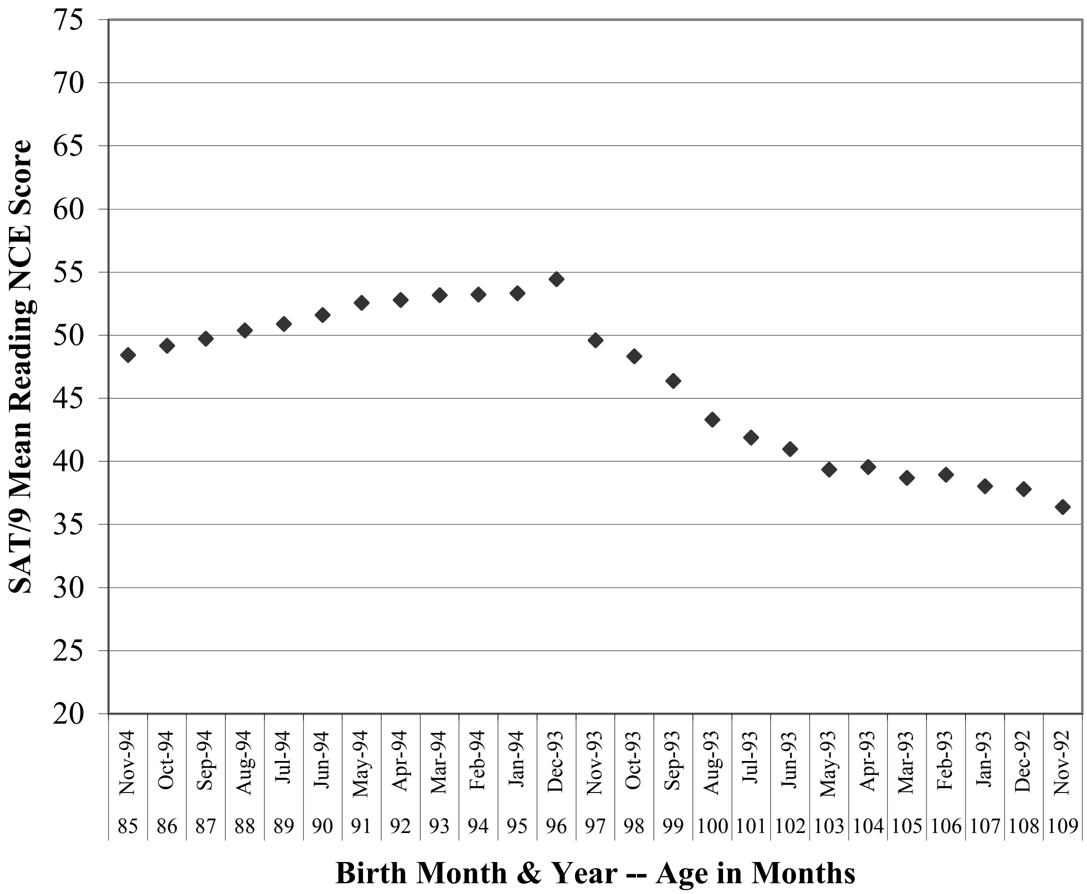
Figure 1 shows the SAT/9 average total reading NCE score for students in grade 2 by age in months who were tested in spring 2002.

---

between age and achievement. In fact, the average normal curve equivalent (NCE) score for students born on December 1, 1994 is 48.5. The average NCE score for students born in November 1994 is 48.4. Since these values are essentially the same, it was easier to exclude them from the analyses than determine how to include them.

[2] Normal age peers are students who start school as soon as they are legally able.

Figure 1. SAT/9 mean total reading score by age in months: Grade 2 STAR 2002, n = 455,638



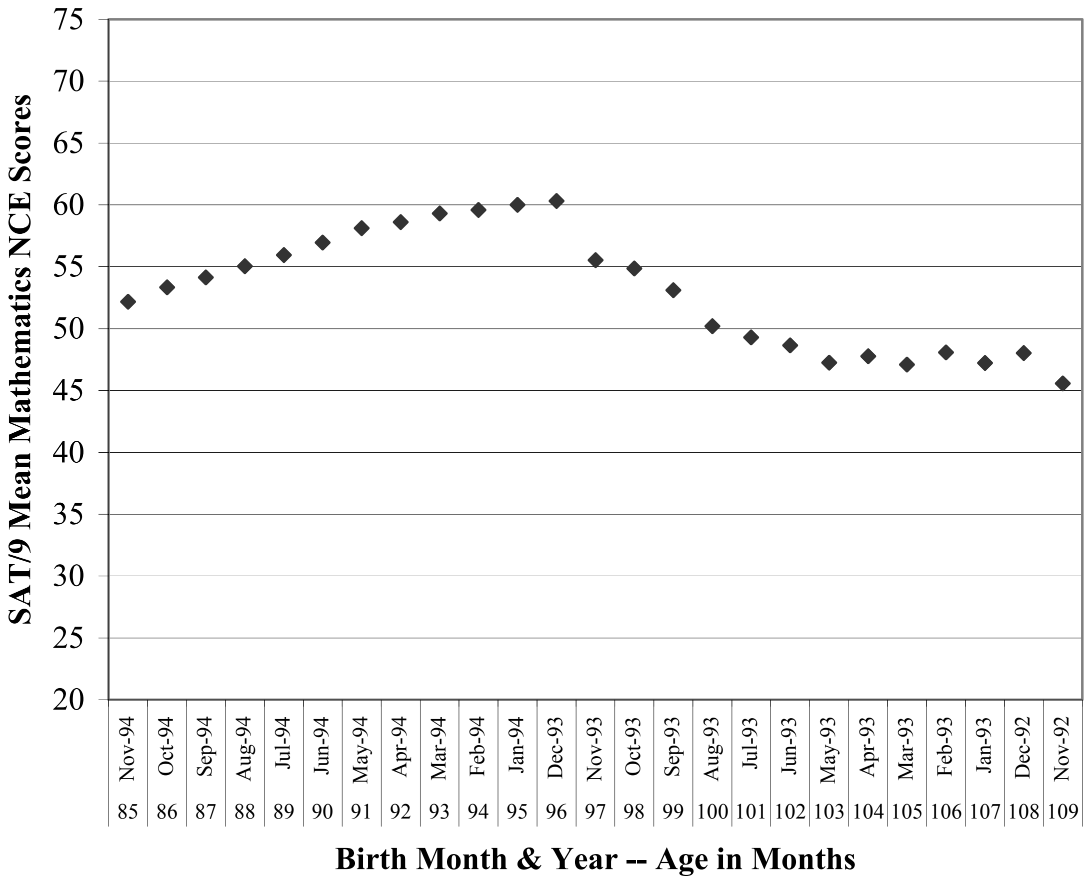
The first age in figure 1 (i.e., 85 months) shows the mean total reading NCE score for students with November 1994 birthdays. These are the youngest students in this particular cohort. The second age (i.e., 86 months) shows the mean total reading NCE score for students with October 1994 birthdays. The age normal peer group for this 2nd grade cohort ranges from 85 months to 96 months.

There is a positive relationship between age and achievement for the age normal peers. As age normal peers get older, their test scores on average get higher.

For students who have been retained (i.e., students who are 97 months and older) there is a negative relationship between age and achievement. As students get older, their test scores on average decline.

Next, data were analyzed to determine whether the relationship between age and achievement is content dependent. Figure 2 shows the SAT/9 average total mathematics NCE score for students in grade 2 by age in months who were tested in spring 2002.

Figure 2. SAT/9 mean total mathematics score by age in months: Grade 2 STAR 2002, n = 469,805



The data pattern in figure 2 is similar to that of figure 1 with a positive relationship between age and achievement for age normal peers and a negative relationship for students who have been retained. The relationship between age and achievement is not content dependent.

To test the linear relationship between age and achievement, SAT/9 mean NCE total reading scores were regressed on age in months. First, the relationship was tested for the age normal peers. Table 1 shows these results.

Table 1

SAT/9 mean total reading NCE score regressed on age in months for grade 2 age normal peers

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	*R-Square*
Regression	1	38.3973	38.3973	293.2372	0.00000	0.9670
Residual	10	1.3094	0.1309			
Total	11	39.7067				

There is a statistically significant relationship between age and mean achievement for the age normal peer group and the relationship is strong (i.e., the $R^2 = .97$). Figure 3 graphically displays the relationship between mean test scores and age for the age normal peer group. Figure 3 also displays the regression equation (i.e., $\hat{y} = 4.7395 + x_{age}(0.5182))$. For each month age increases, the average total reading NCE score increases ½ point.

Figure 3. SAT/9 mean total reading NCE score regressed on age for the age normal peer group: Grade 2 STAR 2002

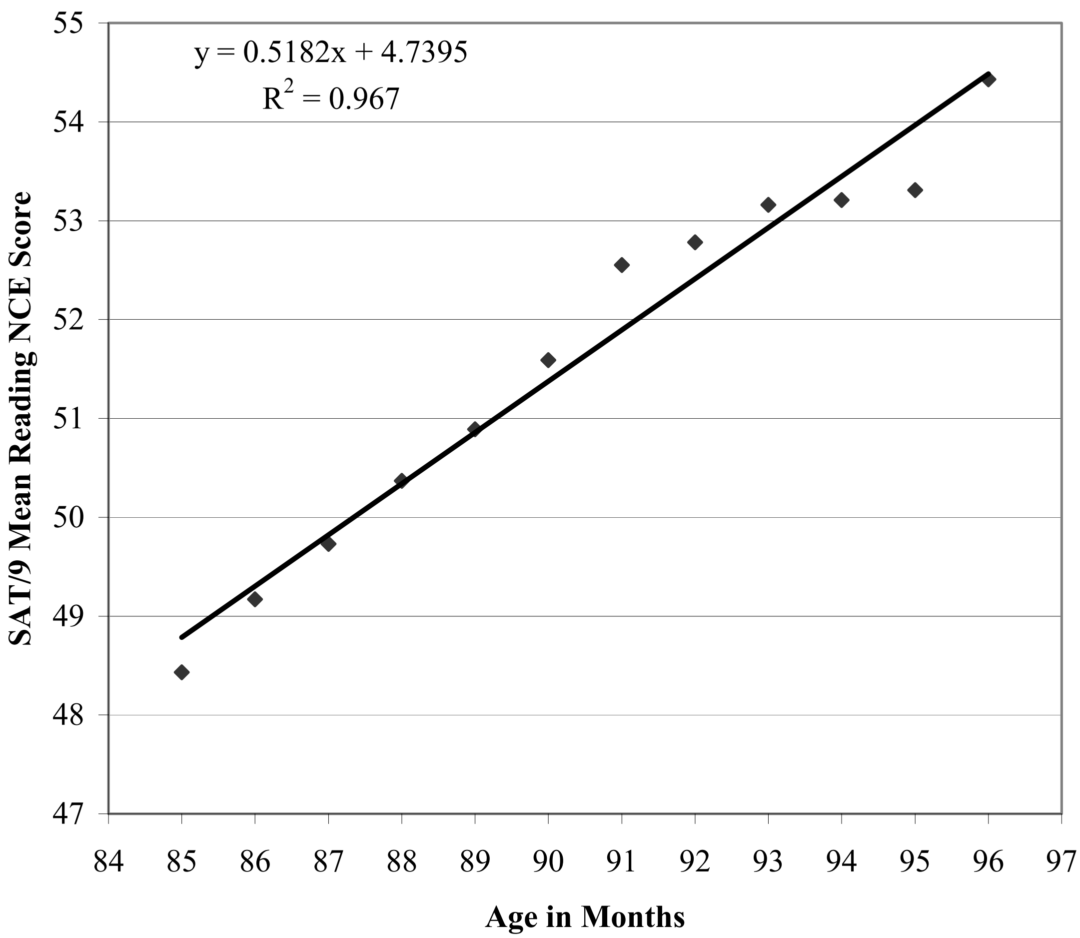


Figure 3 shows a positive linear relationship between mean test score and age for the age normal peer group.

Next, SAT/9 total reading scores were regressed on age in months to test the relationship between age and achievement for retained students. The ages of 97 months and higher represent the students who had been held back. As stated, the student demographic information does not contain information as to why students were held back. Table 2 shows these results.

Table 2
SAT/9 mean total reading NCE score regressed on age in months for grade 2 retained students

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	R-*Square*
Regression	1	190.456	190.456	87.71862	0.0000014	0.888572
Residual	11	23.8834	2.17122			
Total	12	214.339				

There is a statistically significant negative relationship between age and achievement for retained students and the relationship is strong (i.e., $R^2 = .89$).

Figure 4 graphically displays the relationship between mean test scores and age for retained students and the regression equation (i.e., $\hat{y} = 146.84 + x_{age}(-1.023))$. For each month age increases, the average total reading NCE score decreases 1 point.

Figure 4. SAT/9 mean total reading NCE score regressed on age for retained students: Grade 2 STAR 2002

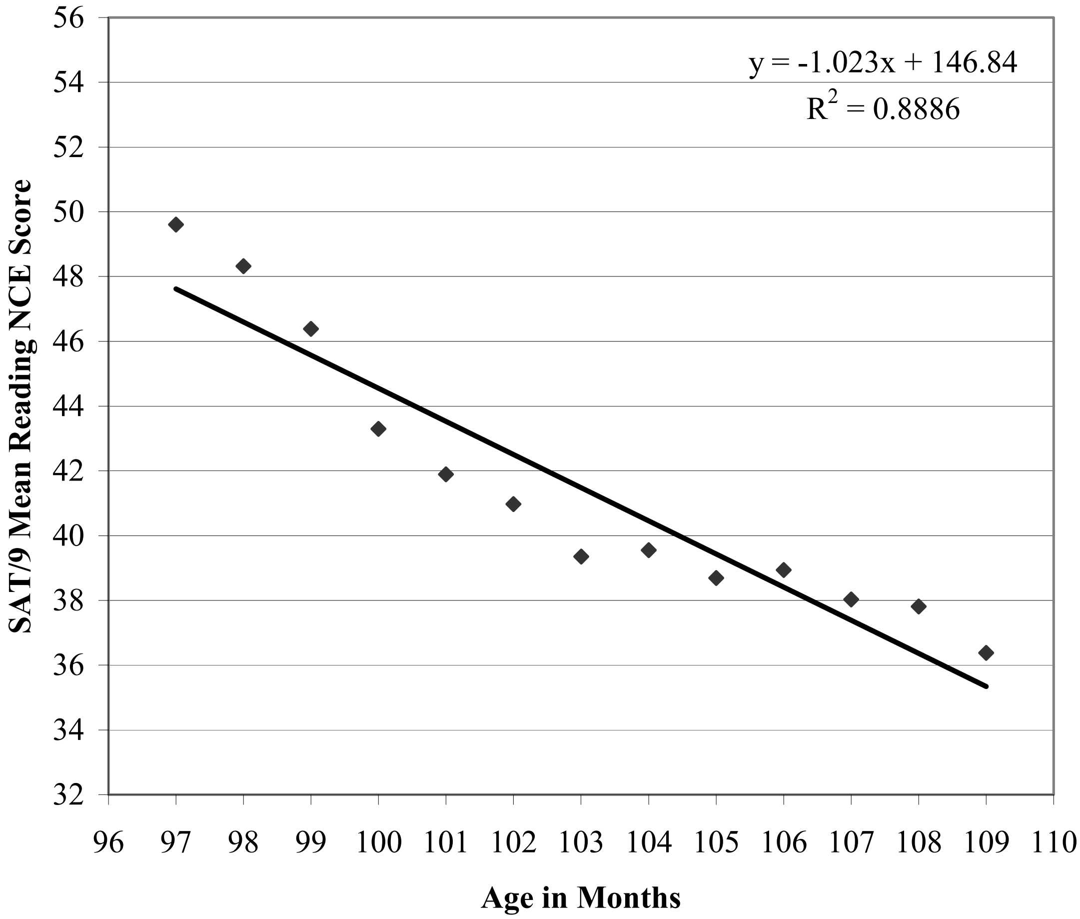


Figure 4 shows that test scores begin to decline for retained students and continue to do so through the age range.

The variance of mean test scores for retained students is approximately 16 points. That is, even thought the R^2 is less for grade 2 retained students than for age normal peers, the difference in test scores for retained students is almost three times the difference in test scores for age normal peers.

Educators and researchers that recommend delayed school entry for students typically mean the students who are one to three months older than the age normal peer group. The "summer birth date" research concerns states where the cut off for school entry is around September 1. Students with summer birth dates are those students with birth dates three months prior to a September 1 cut off. In California, students need to be five years old by December 2 to enroll in kindergarten or six years old by December 2 to enroll in first grade. Therefore, the relationship between age and achievement was tested when students who are one to three months older than the age normal peer group were added to the age normal peer group. Table 3 shows these results.

Table 3

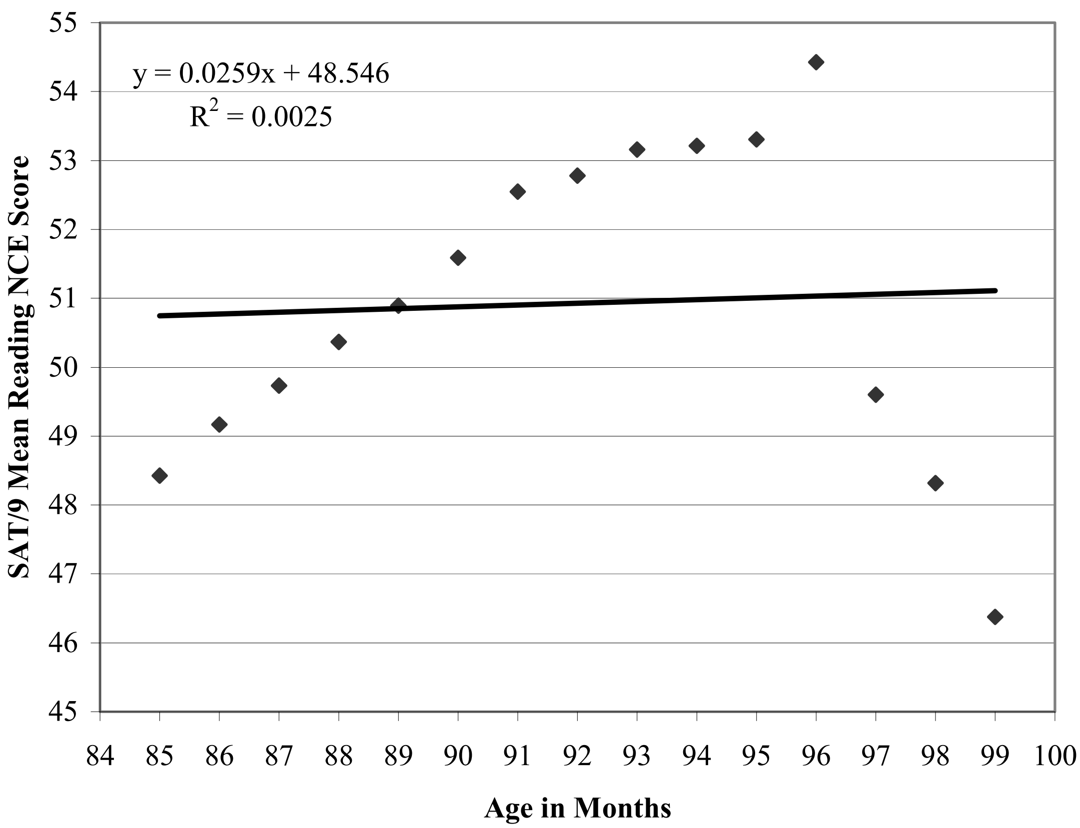
SAT/9 mean total reading NCE score regressed on age in months for grade 2 normal age peers and retained students

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	R *Square*
Regression	1	0.1877	0.1877	0.0326	0.8594	0.0025
Residual	13	74.7667	5.7513			
Total	14	74.9544				

There is no longer a statistically significant linear relationship between age and achievement. The R^2 = .003. When the sample includes retained students, the positive linear relationship between age and achievement disappears. Figure 5 graphically displays these same data.

Figure 5. SAT/9 mean total reading NCE score regressed on age for the age normal peer group plus three months overage: Grade 2 STAR 2002



Being older is better to a point. Beyond that point the effect is negative. Age normal students who are older do better on average. However, students who are older because they have been retained do worse on average. The recommendation to delay school entry or retain students to improve academic achievement is not supported by these data.

Examining *mean* test scores provides a simplified way to examine the relationship between age and achievement. However, using the mean also disguises the actual relationship. Table 4 shows the results of regressing total reading scores for individual students on age in months.

Table 4

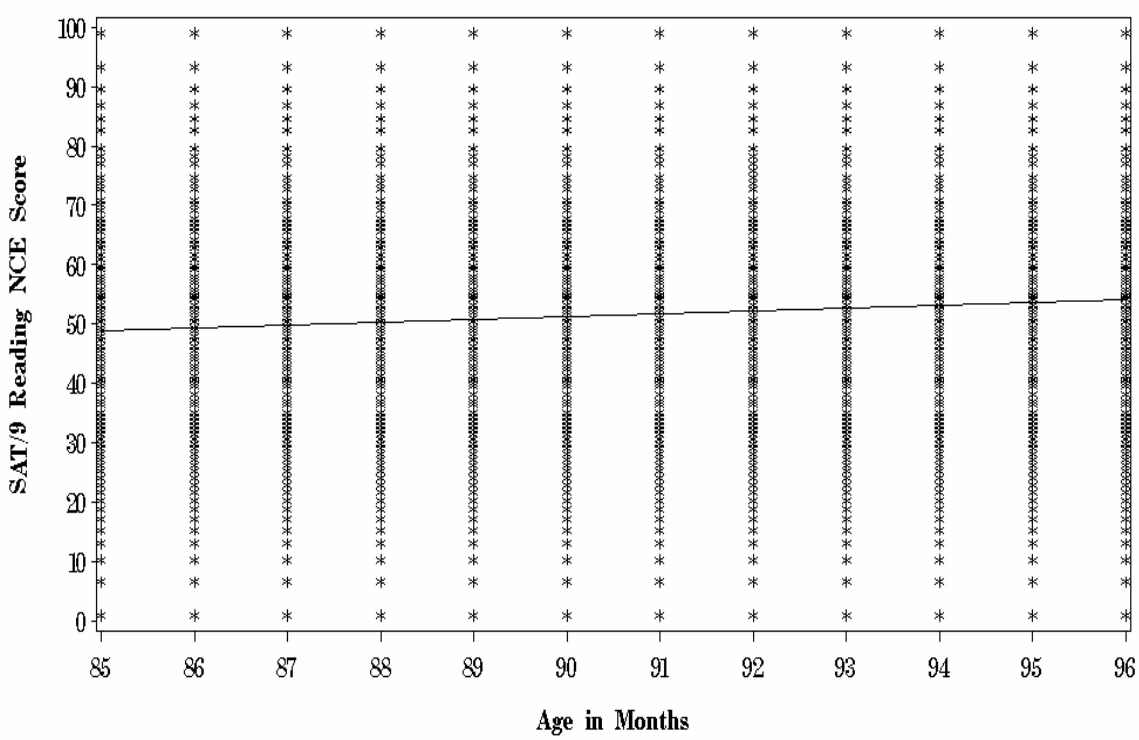
SAT/9 total reading NCE score regressed on age in months for grade 2 age normal peers

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	*R Square*
Regression	1	993813	993813	2695.52	<.0001	0.0069
Residual	390251	143881916	368.69070			
Total	390252	144875728				

The positive relationship between age and test scores is still statistically significant. However, the $R^2 = .0069$. That is, even though there is a statistically significant relationship between age and achievement, age accounts for little of the variance in test scores. The regression equation is: $\hat{y} = 8.743 + x_{age}(.473)$. Figure 6 graphically displays these data.

Figure 6. Stanford 9 reading NCE score regressed on age for the age normal peer group: Grade 2 STAR 2002



These data indicate that making strong inferences about student academic performance if age is known is not sound. Many students with November 1994 birth dates (i.e., the youngest students) performed well and many students with December 1993 birth dates (i.e., the oldest students) performed poorly.

Table 5 and figure 7 show the relationship between age and achievement for retained students when total reading scores for individual students are regressed on age.

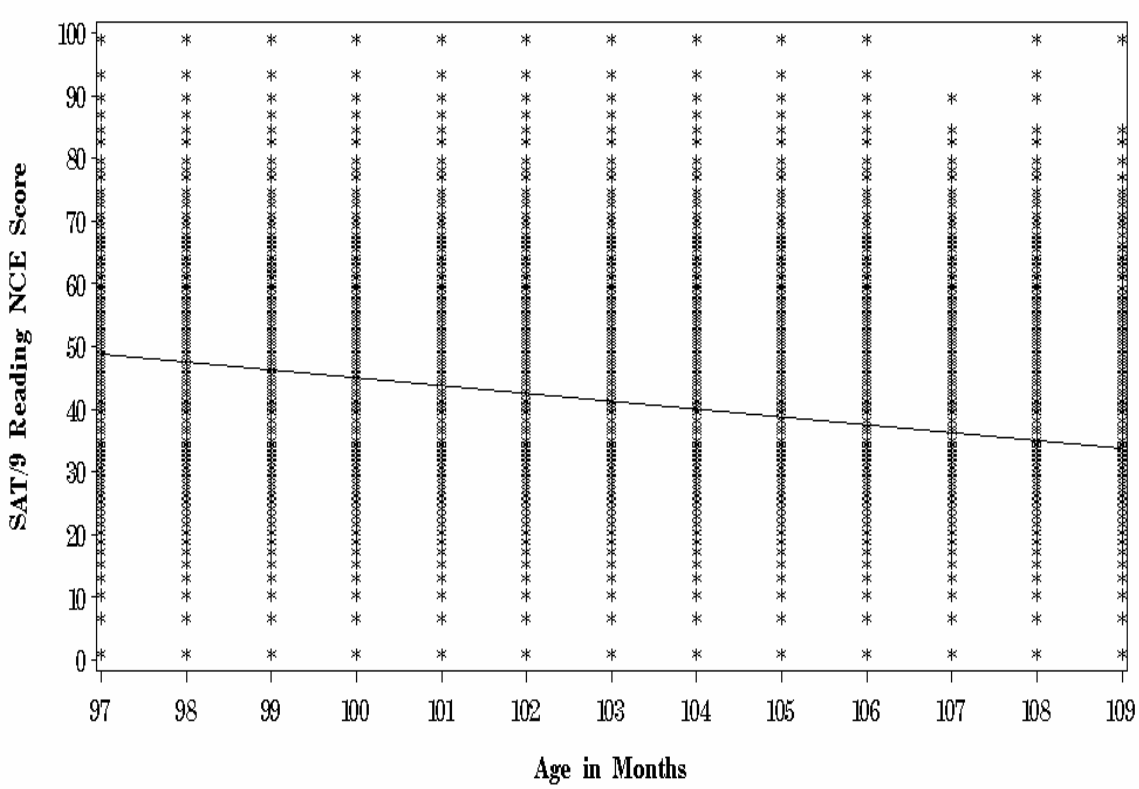
Table 5

SAT/9 total reading NCE score regressed on age in months for grade 2 retained students

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	*R Square*
Regression	1	1065911	1065911	2980.03	<.0001	0.0436
Residual	65383	23386492	357.68460			
Total	65384	24452403				

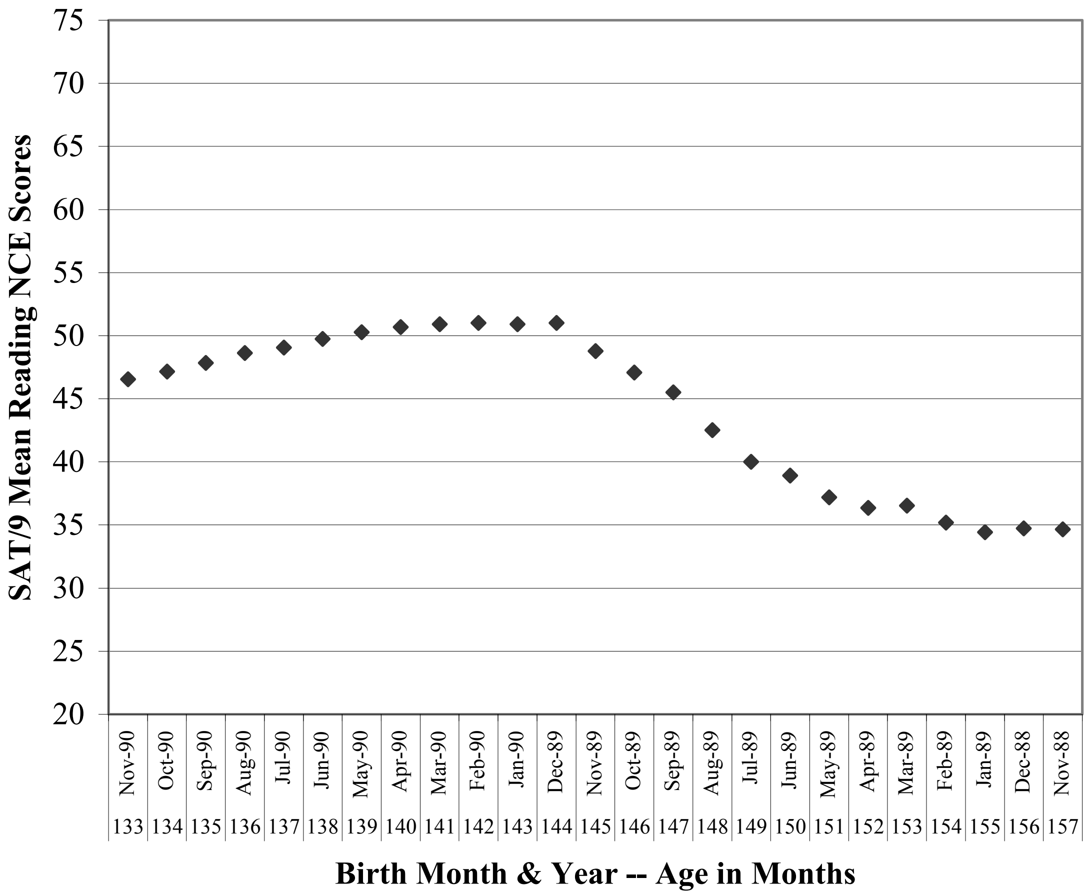
Figure 7. Stanford 9 reading NCE score regressed on age for retained students: Grade 2 STAR 2002



The regression equation is: $\hat{y} = 170.413 + x_{age}(-1.254)$. The negative relationship between age and test scores is still statistically significant. Again, age accounted for little of the variance in test scores (i.e., $R^2 = .0436$) but it accounted for more of the variance than it did for the age normal peer group. This may be due to the fact that there were fewer students in the retained group than in the age normal peer group. Fewer students mean less variance and so age has less variance for which to account. Or, the negative relationship between age and achievement is stronger for retained students than the positive relationship between age and achievement for the age normal peers.

To determine if the relationship between age and achievement is maintained over time, the relationship between age and achievement was tested for older students (i.e., grade 6 students). Figure 8 shows the average total reading NCE scores for students in grade 6 by age in months who were tested in spring 2002.

Figure 8. SAT/9 mean total reading score by age in months: Grade 6 STAR 2002, n= 465,633



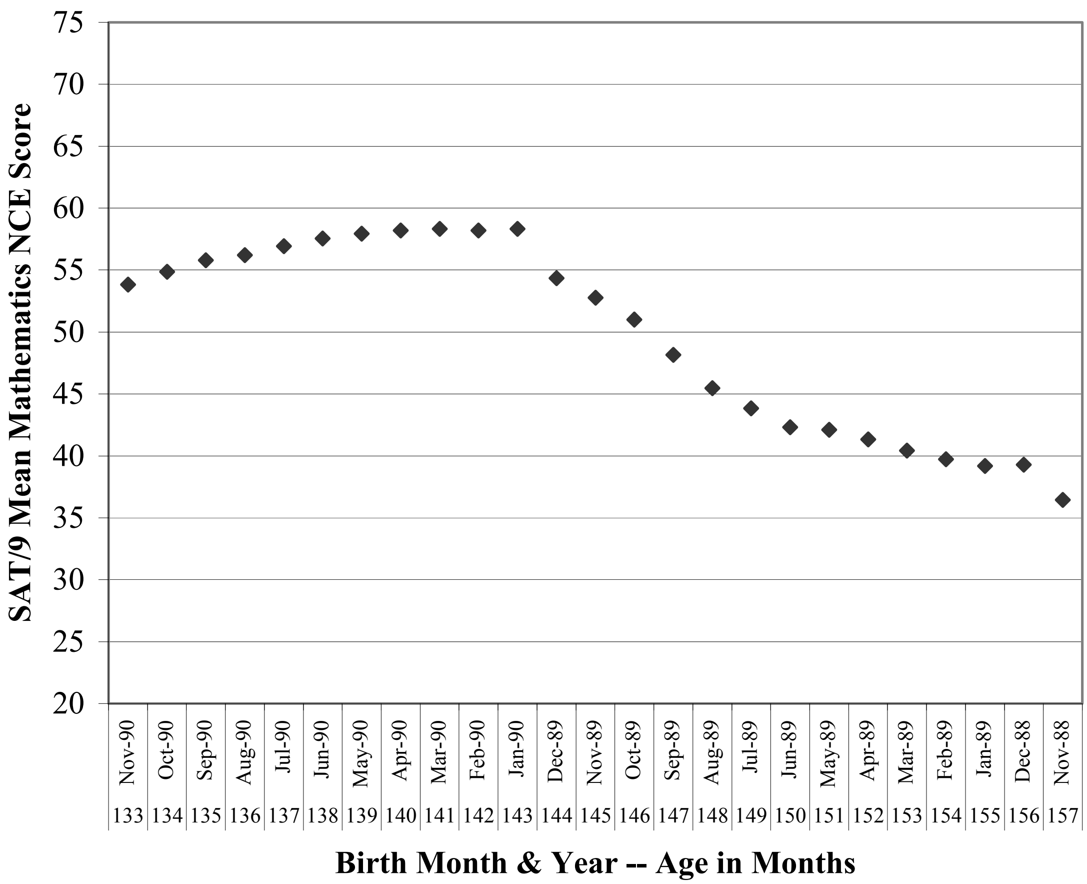
The first age in figure 8 (i.e., 133 months) shows the mean NCE total reading score for students with November 1990 birth dates. These are the youngest students in this particular cohort. The second age (i.e., 134 months) shows the mean NCE total reading score for students with October 1990 birth dates. The age normal peer group for this cohort ranges from 133 to 144 months. Retained students are 145 months and older.

As with grade 2 students there is a positive relationship between age and achievement for age normal peers (i.e., students 133 to 144 months old). As students get older their test scores get higher.

For retained students (i.e., students 145 to 157 months old) there is a negative relationship between age and achievement. As students get older their test scores get lower. The relationship between age and achievement is consistent across two grades.

Again data were analyzed to determine whether the relationship between age and achievement is content dependent. Figure 9 shows the SAT/9 average total mathematics NCE scores for students in grade 6 by age in months who were tested in spring 2002.

Figure 9. SAT/9 mean total mathematics score by age in months: Grade 6 STAR 2002, n = 469,486



The data in Figure 9 are similar to that of figure 8. That is, there is a positive relationship between age and achievement for age normal peers and a negative relationship for students who have been retained. The relationship between age and achievement is not content dependent.

To test the significance of the relationship between age and achievement, mean total reading NCE score were regressed on age in months. Table 6 shows these results for age normal peers.

Table 6

SAT/9 mean total reading NCE score regressed on age in months for grade 6 age normal peers

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	R *Square*
Regression	1	26.0252	26.0252	101.8848	0.00000	0.9106
Residual	10	2.5544	0.2554			
Total	11	28.5796				

Results indicate that there is a statistically significant positive relationship between age and achievement. The R^2 for grade six students (i.e., .91) is lower than the R^2 for grade 2 students (i.e., .97). The positive relationship between age and achievement for age normal peers decreased in strength as students aged.

Figure 10 graphically displays the mean reading NCE scores for normal peers regressed on age in months.

Figure 10. SAT/9 mean total reading NCE score regressed on age for the age normal peer group: Grade 6 STAR 2002

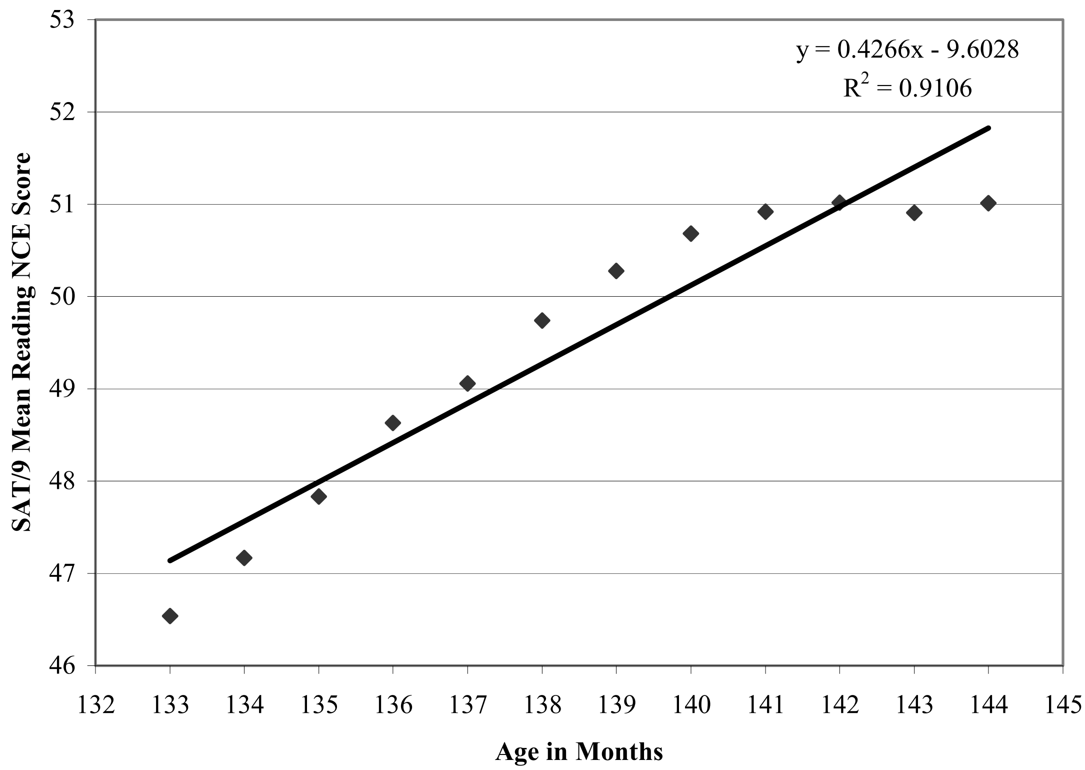


Figure 10 displays the positive linear relationship between mean test score and age for the age normal peer group.

Next, mean total reading NCE scores were regressed on age in months for retained students. Table 7 shows these results.

Table 7
SAT/9 mean total reading NCE score regressed on age in months for grade 6 retained students

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	R *Square*
Regression	1	271.9146	271.9146	92.6208	0.00000	0.8938
Residual	11	32.2936	2.9358			
Total	12	304.2082				

As with grade 2, there is a statistically significant negative relationship between age and achievement. The R^2 for grade six students (i.e., .89) is the same as the R^2 for grade 2 students (i.e., .89). Through grade six the strength of the relationship between age and achievement for retained students has remained constant.

Figure 11 shows these same results.

Figure 11. SAT/9 mean total reading NCE score regressed on age for retained students: Grade 6 STAR 2002

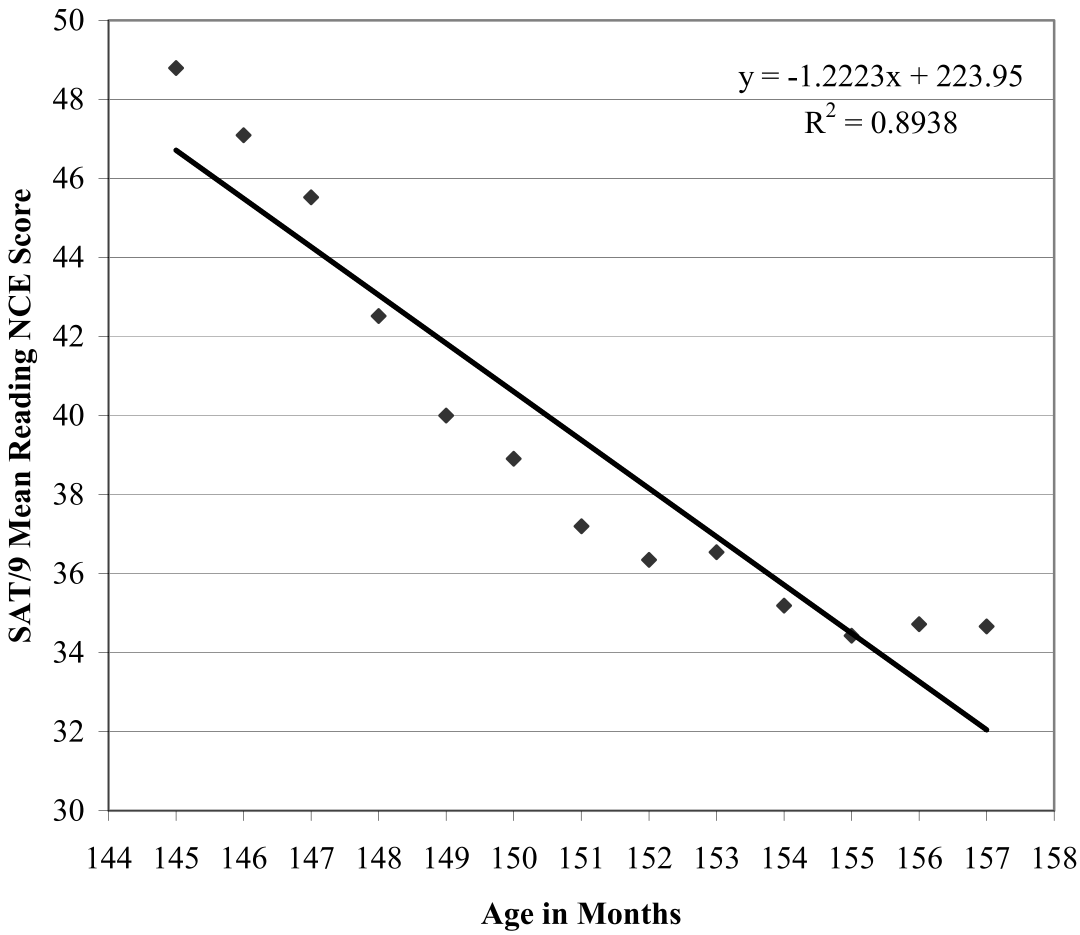
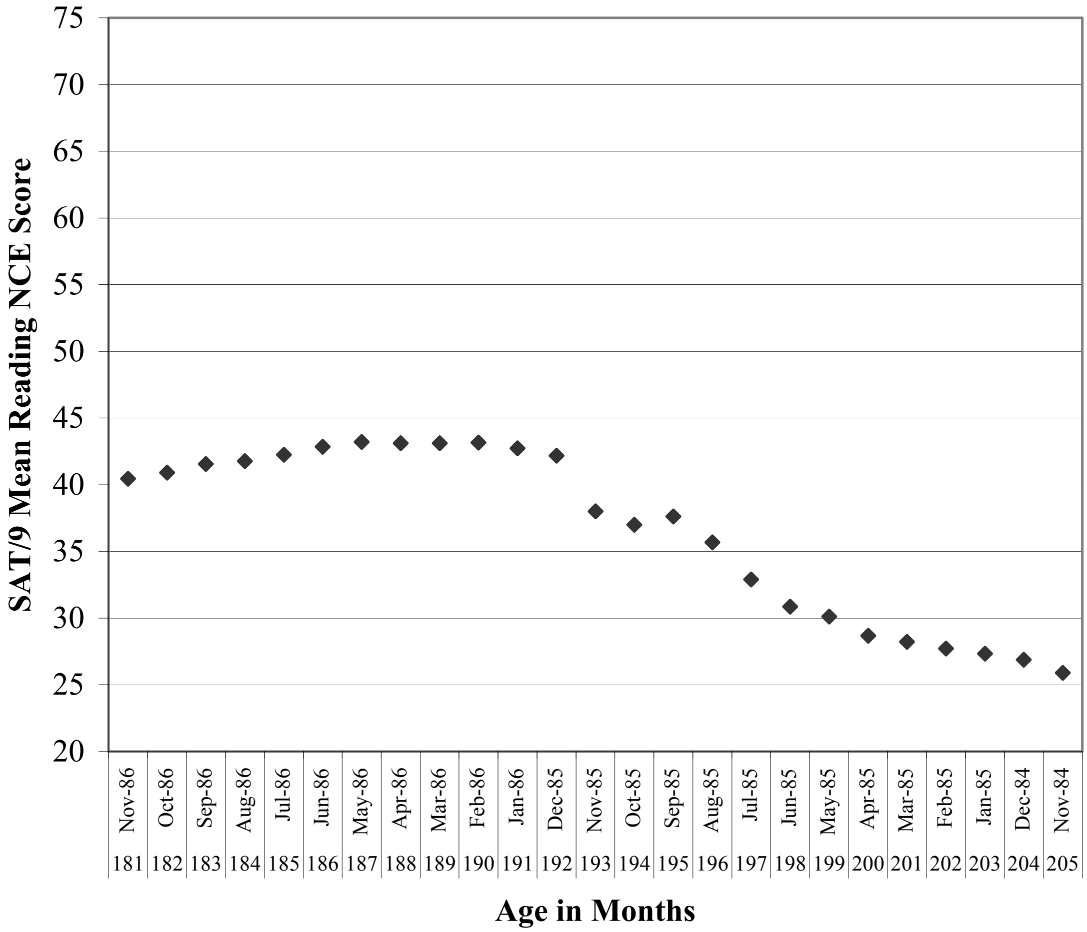


Figure 11 shows that test scores begin to decline for retained students and continue to do so through the age range.

Next, the relationship between age and achievement was tested for grade 10 students. Figure 12 shows the average total reading scores for students in grade 10 by age in months who were tested in spring 2002.

Figure 12. SAT/9 mean total reading score by age in months: Grade 10 STAR 2002, n=386,910



The first age in figure 12 (i.e., 181 months) shows the mean NCE total reading score for students with November 1986 birth dates. These are the youngest students in this particular cohort. The age normal peer group for this cohort ranges from 181 to 192 months. Retained students are 193 months and older.

Again, mean total reading NCE score were regressed on age in months to test the significance of the relationship between age and achievement. Table 8 shows these results for age normal peers.

Table 8

SAT/9 mean total reading NCE score regressed on age in months for grade 10 age normal peers

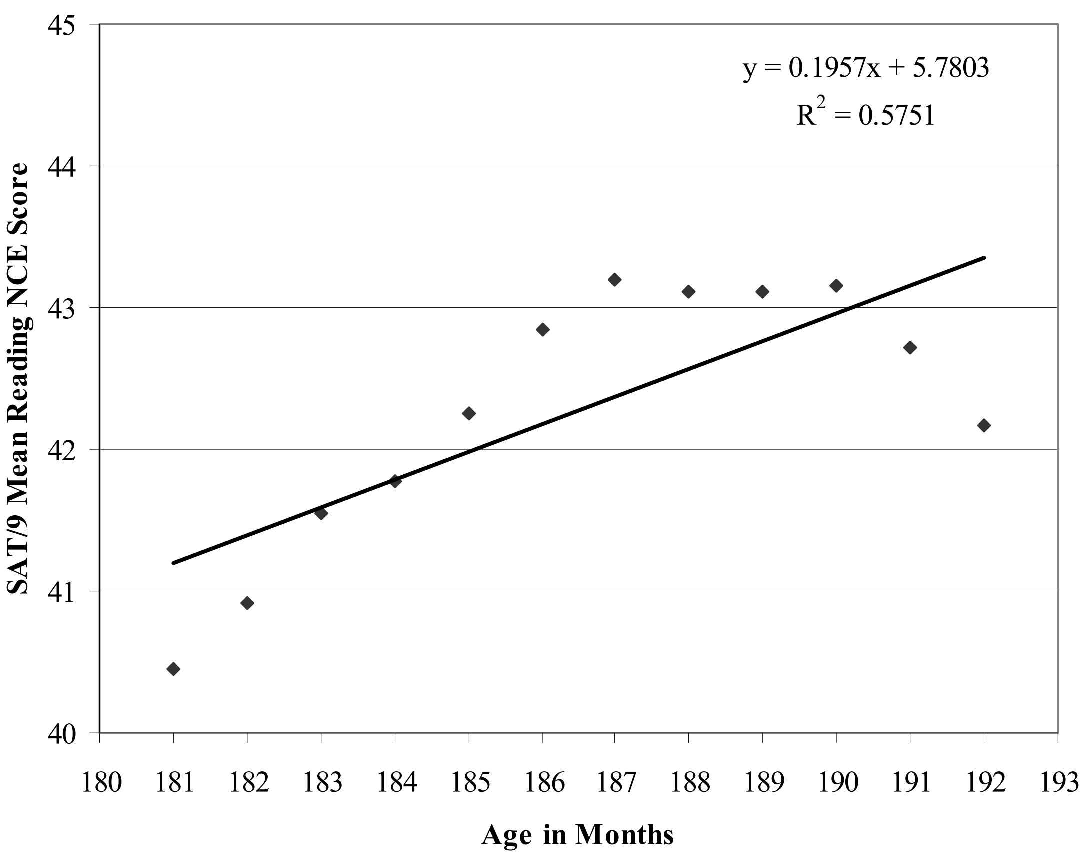
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	*R Square*
Regression	1	5.4747	5.4747	13.5335	0.00425	0.5751
Residual	10	4.0453	0.4045			
Total	11	9.5200				

Results indicate that there is a statistically significant positive linear relationship between age and achievement. However, the R^2 for grade ten students (i.e., .58) is relatively low compared to students in grades two and six.

Figure 13 shows these same results.

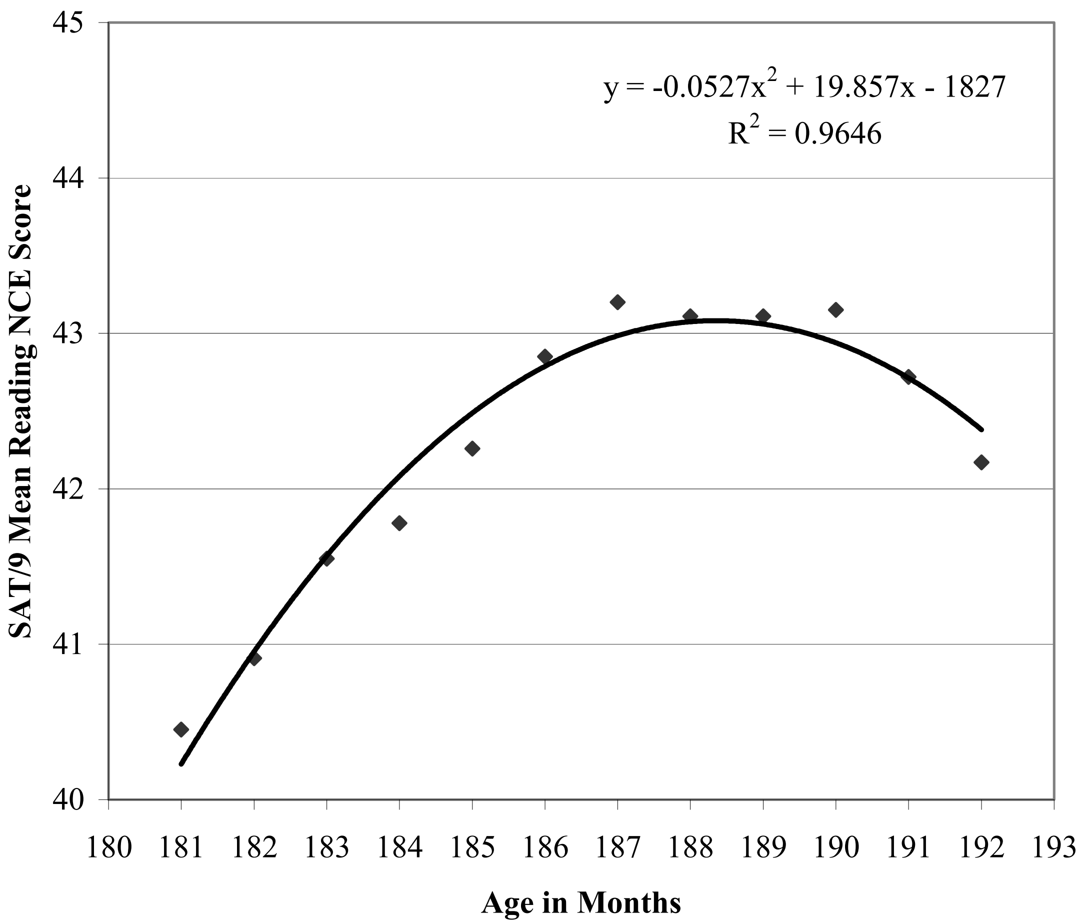
Figure 13. SAT/9 mean total reading NCE score regressed on age for the age normal peer group: Grade 10 STAR 2002



When students are 187 months, achievement levels off until 190 months. At 191 months, achievement begins to decline. The variance in mean test scores is only a couple of points. Despite statistical significance, it is safe to say that there is no longer a positive linear relationship between age and achievement for age normal peers. I make this statement for a couple of reason reasons. The first is that the oldest age normal peers do not have the highest average test scores. Second, the variance in test scores for the age normal peers is very small. That is, there is statistical significance but no practical difference in test scores. Whatever academic advantage being older had for younger students is gone by grade ten.

Figure 14 shows the results of fitting a 2nd order polynomial through the data.

Figure 14. SAT/9 mean total reading NCE score regressed on age for the age normal peer group: Grade 10 STAR 2002



These results emphasize that there is no longer a positive linear relationship between age and achievement for age normal peers.

Mean total reading NCE scores were regressed on age in months for grade 10 retained students. Table 9 shows these results.

Table 9
SAT/9 mean total reading NCE score regressed on age in months for grade 10 retained students

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*	R *Square*
Regression	1	219.4507	219.4507	161.3951	0.00000	0.9362
Residual	11	14.9568	1.3597			
Total	12	234.4075				

As with students in grades 2 and 6 there is a statistically significant negative linear relationship between age and achievement. The R^2 for grade 10 students (i.e., .94) is higher than the R^2 for grade 2 and grade 6 students (i.e., .89). As students get older the strength of the negative linear relationship between age and achievement for retained students increases.

Figure 15 shows these same results.

Figure 15. SAT/9 mean total reading NCE score regressed on age for retained students: Grade 10 STAR 2002

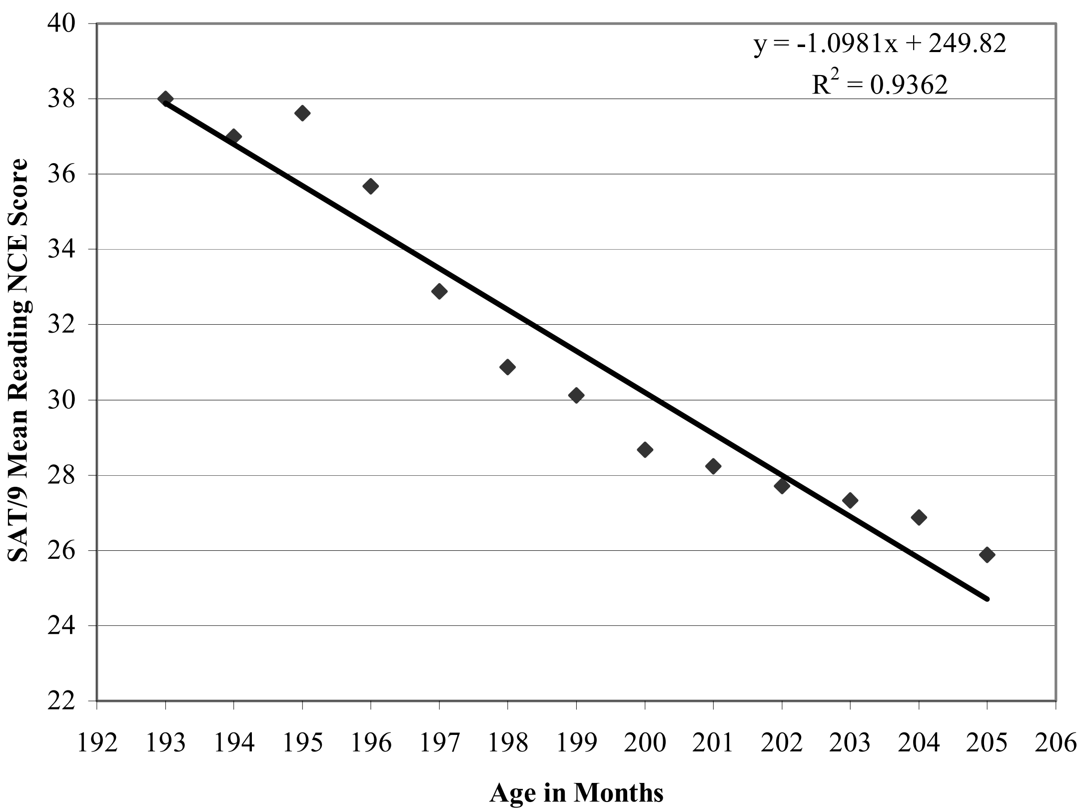


Figure 15 shows that test scores decline for retained students and continue to do so through the age range.

Figure 16 shows that as students get older the difference in mean test score for the oldest and youngest students in the age normal peer group decreases.

Figure 16. Difference between mean total reading NCE scores for the oldest and youngest students in the age normal peer group: STAR 2002

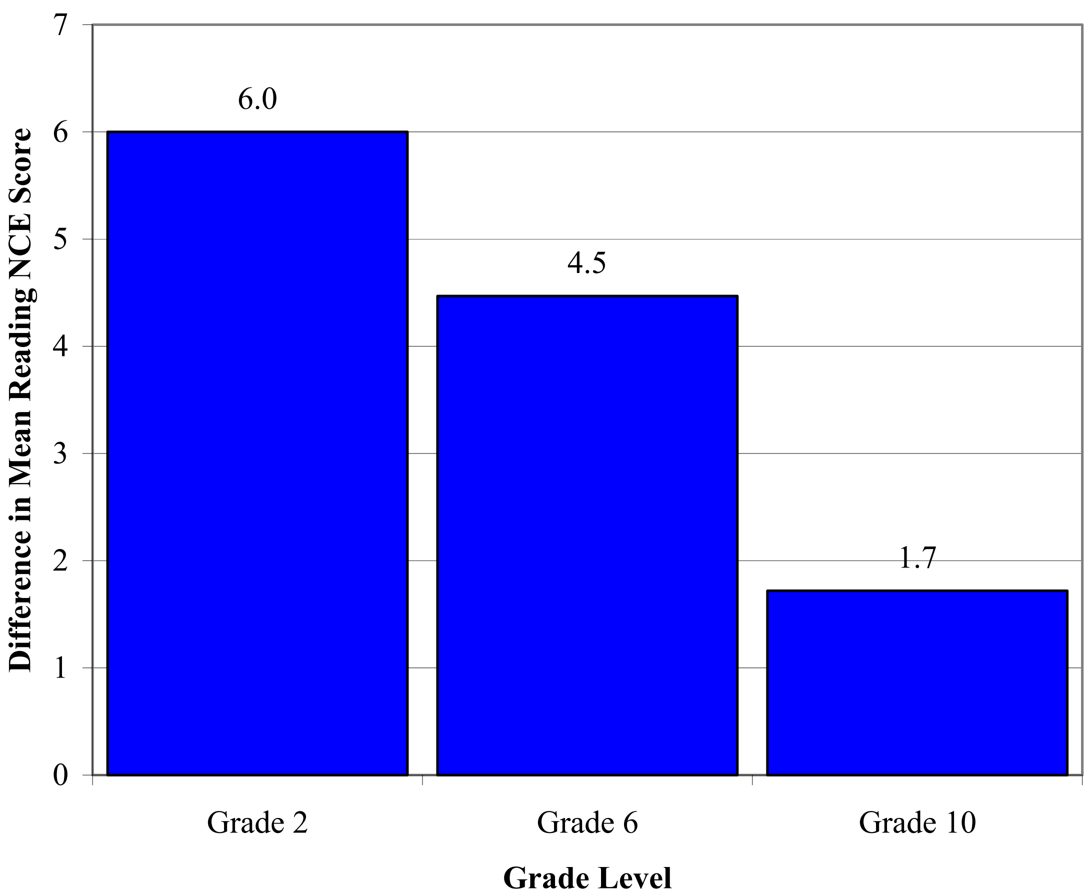
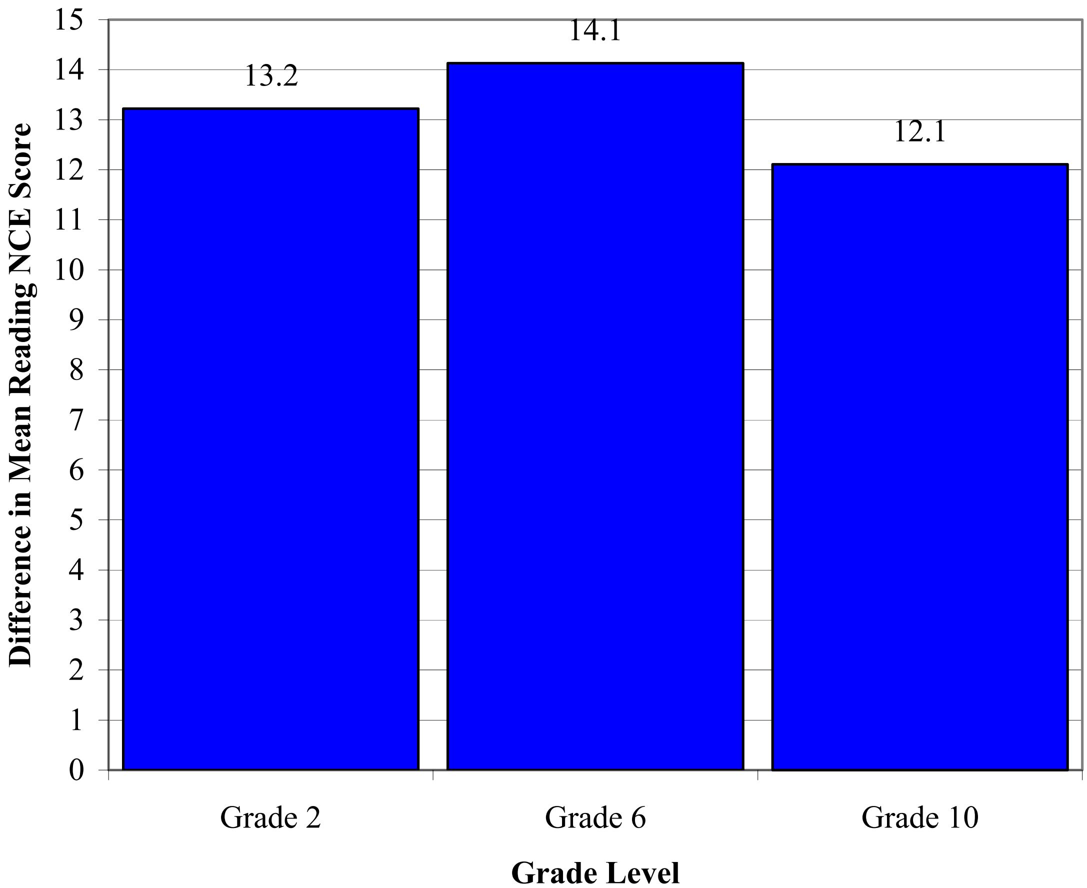


Figure 16 shows that the variance in test scores for age normal peers decreases over time. By grade 10 the variance is so small that there is no practical difference in the test scores.

Figure 17 shows the difference in mean test scores for the oldest and youngest students for retained students.

Figure 17. Difference between mean total reading NCE score for the oldest and youngest students in the retained group: STAR 2002



First, it can be seen that the difference between the oldest and youngest students is greater for retained students than for age normal peers. Second, as retained students get older the difference remains somewhat constant. For retained students the variance in test scores does not decrease over time.

Another way to evaluate the advantage of retention is to compare test scores for birth dates close to the entrance cut off date. For example, students in 2nd grade with November 1984 birth dates are the youngest students. These are the students who are believed to be most at risk of academic failure by Uphoff & Gilmore (1985). Students with November 1983 birth dates are the retained students who have been given time to mature in ways that lead to academic success according to Uphoff & Gilmore (1985). Older students with November birth dates should demonstrate higher academic performance than younger students with November birth dates. Table 10 compares test scores for older and younger students with November, October, and September birth dates for three grade levels.

Table 10
Mean SAT/9 total reading NCE score for retained students compared to age normal students by birth month

Grade 2

Nov-93	49.6	Oct-93	48.3	Sep-93	46.4
Nov-94	48.4	Oct-94	49.2	Sep-94	49.7
Difference	1.2		-0.9		-3.3

Grade 6

Nov-89	48.8	Oct-89	47.1	Sep-89	45.5
Nov-90	46.5	Oct-90	47.2	Sep-90	47.8
Difference	2.3		-0.1		-2.3

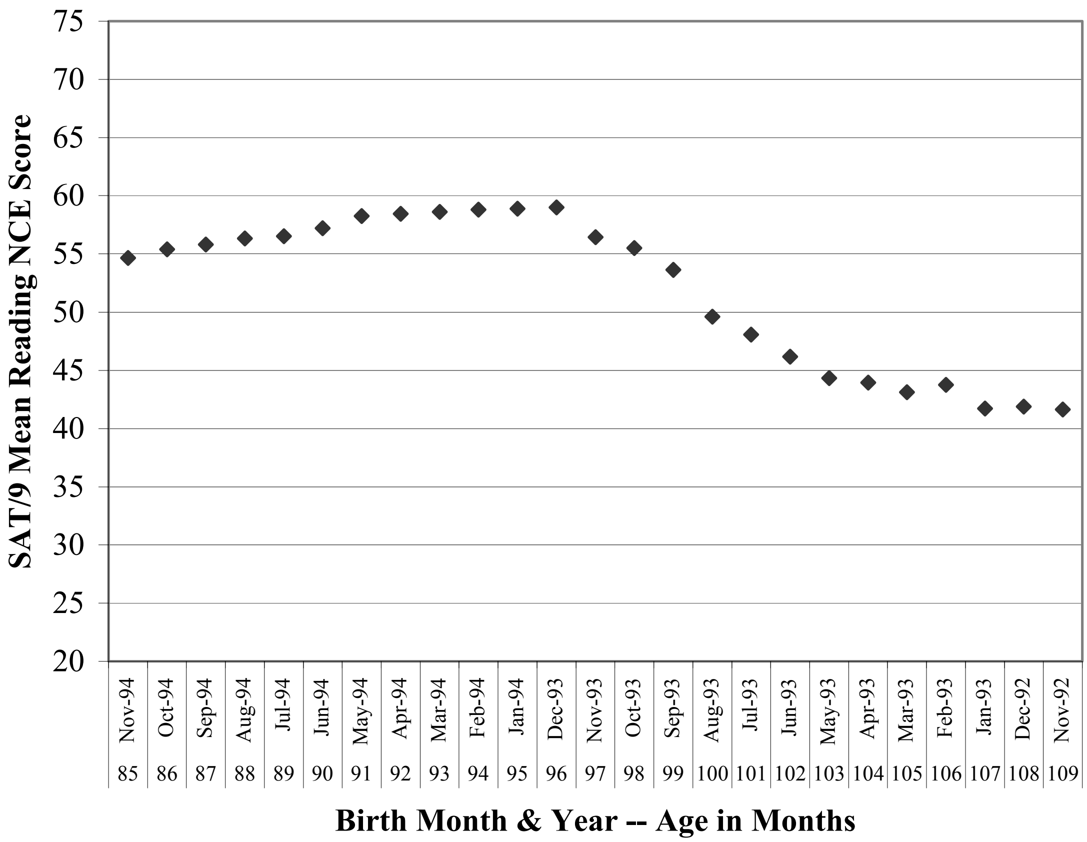
Grade 10

Nov-85	38.0	Oct-85	37.0	Sep-85	37.6
Nov-86	40.5	Oct-86	40.9	Sep-86	41.6
Difference	-2.5		-3.9		-3.9

Retaining students with November birth dates does not on average provide a large academic advantage over younger classmates. In grade 2 retained students score on average 1 NCE point higher than their younger classmates and in grade 6 they score 2 points higher. By grade 10 older students score 2 and a half points lower than their younger classmates. Older students with October and September birth dates score lower than their younger classmates. These analyses continue to undermine the contention that retention provides students an academic advantage over their younger classmates.

As stated, even though most students will be older than their age normal peers for three reasons: they started school late, spent two years in a transitional kindergarten or first grade program, or were "flunked" and forced to repeat a grade, there may be other reasons students are older than their classmates. For example, maybe there are special education students who are older because they participate in un-graded programs but are forced to indicate a grade level for testing purposes. Maybe there are English learners (EL) who enter the system older than their classmates. Maybe special education and EL students artificially depress the test scores of older students. To examine this possibility, special education and EL students were removed from the analysis. Figure 18 shows average reading test scores by age in months for 2nd graders with special education and EL students were removed from the analysis.

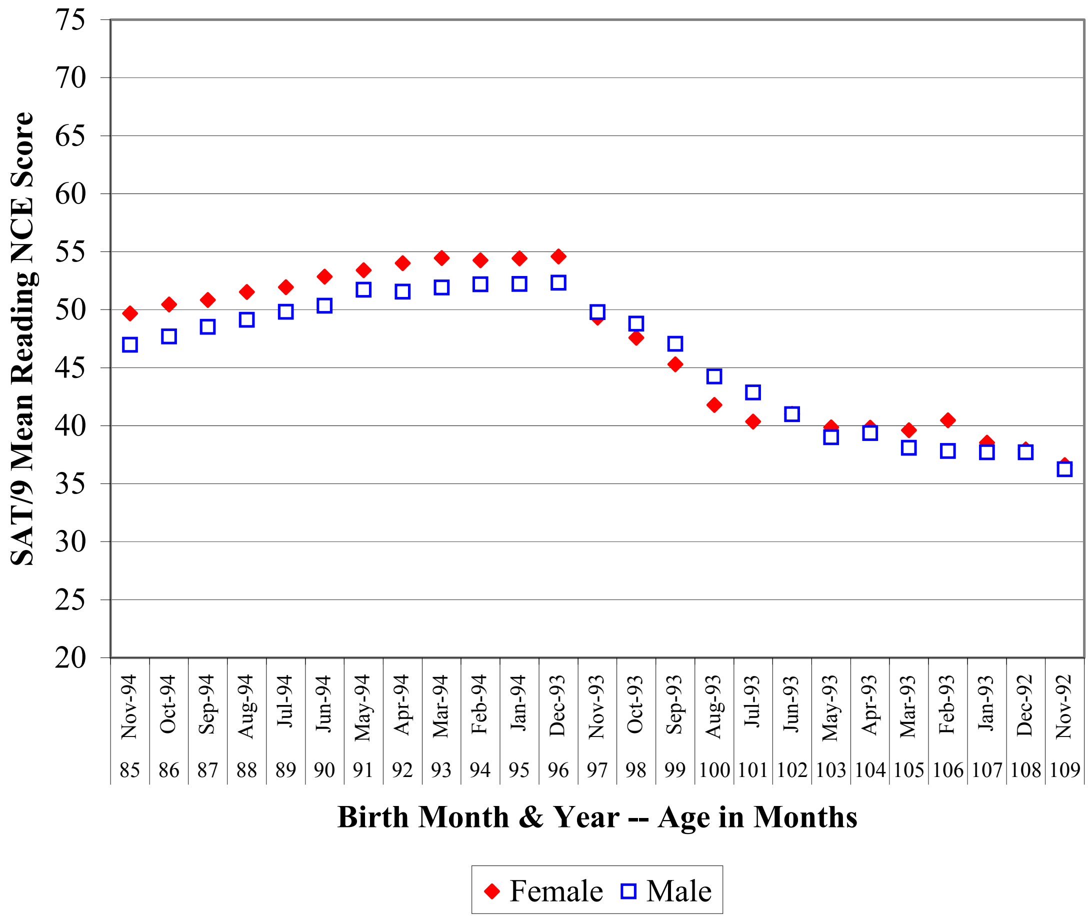
Figure 18. Mean SAT/9 total reading score by age in months: Grade 2 STAR 2002: No special education or EL students, n = 271,255



Overall test scores are higher when special education and EL students are not included. However, the pattern of scores by age in months is very similar to earlier analyses. That is, for age normal peers scores go up as age goes up. For retained students scores go down as age goes up. Removing special education and EL students does not modify earlier conclusions.

The pattern of mean test scores by age might differ by subgroup. To evaluate gender differences, the average total reading NCE scores for students in grade 2 was calculated by age in months and by gender. Figure 19 shows these results.

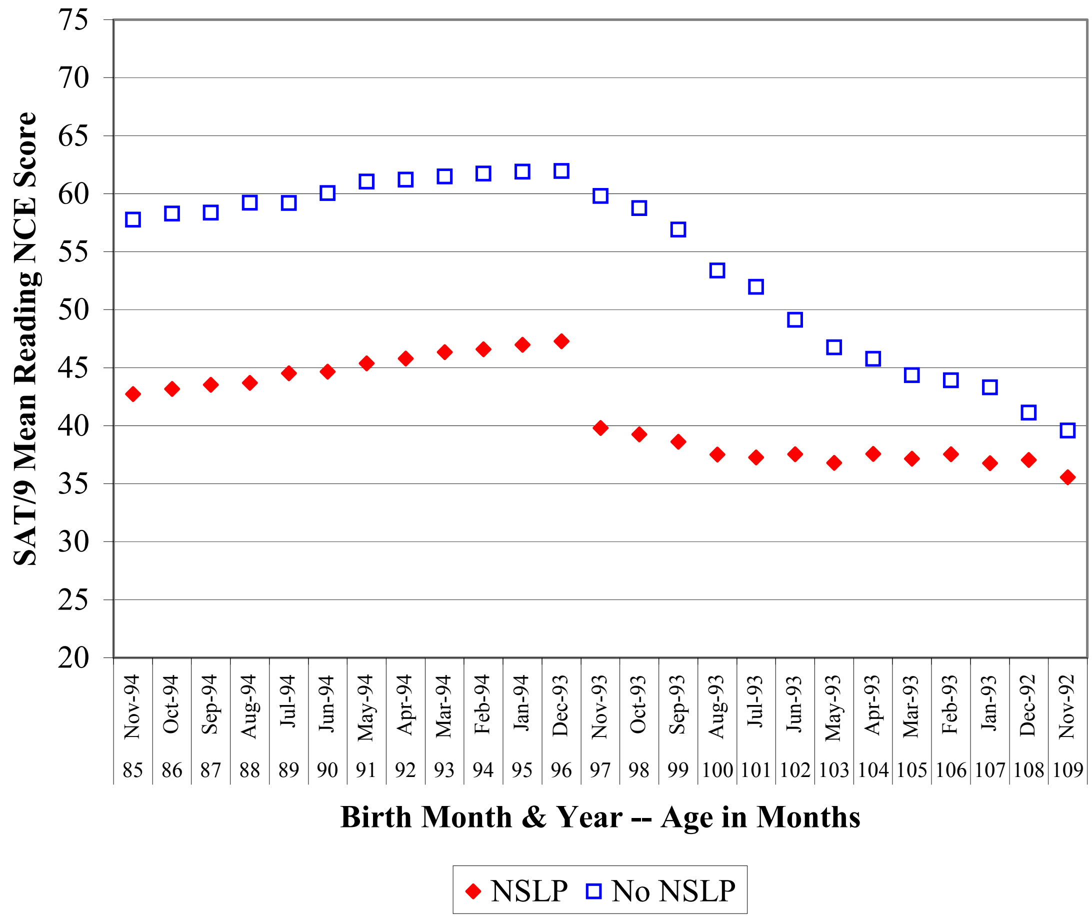
Figure 19. SAT/9 mean total reading score by gender and age in months: Grade 2 STAR 2002: Female n = 222,969, male n = 232,439



The pattern of mean test scores by age is consistent for females and males. There are no gender differences. For both females and males there is a positive relationship between age and achievement for the age normal peers. For students who have been retained there is a negative relationship between age and achievement.

To evaluate SES differences, the average total reading NCE scores for students in grade 2 was calculated by age in months and by the national school lunch program (NSLP). NSLP indicates whether or not students receive free or reduced lunch. Participation in NSLP is an indicator of lower SES. No NSLP is an indicator of higher SES. Figure 20 shows these results.

Figure 20. SAT/9 mean total reading score by NSLP and age in months: Grade 2 STAR 2002: NSLP n = 263,290, no NSLP n = 192,348



The pattern of mean test scores by age is somewhat different for NSLP and no NSLP students. No NSLP students have the familiar patter of a positive relationship between age and achievement for the age normal peers and a negative relationship between age and achievement for retained students. Although NSLP students have lower average test scores, there is a positive relationship between age and achievement for the age normal peers. For retained students there is a negative relationship between age and achievement as compared to age normal students. Average test scores for retained are lower than those for age normal students. However, as NSLP students get older their average test scores don't show the decline seen in other figures. Average test scores tend to flatten out. It may be that the average test scores have gotten so low that it is difficult for average scores to get any lower.

Entrance Age Patterns in California Public Schools

Even though evidence from this study and others indicates that being older on average does not provide an academic advantage, there may be interest in the proportion of delayed entry students that make up the overage population. There may also be interest in the extent of academic "red-shirting" in California public schools. Available data does not distinguish students who started school late from those who were retained for other reasons but it may be possible to make some inferences by looking at age frequency distributions.

Figure 21 shows the percent of students in grade 2 by age in months who were tested in spring 2002.

Figure 21. Age of second grade students on December 1, 2001 for the STAR 2002 test, n = 485,796

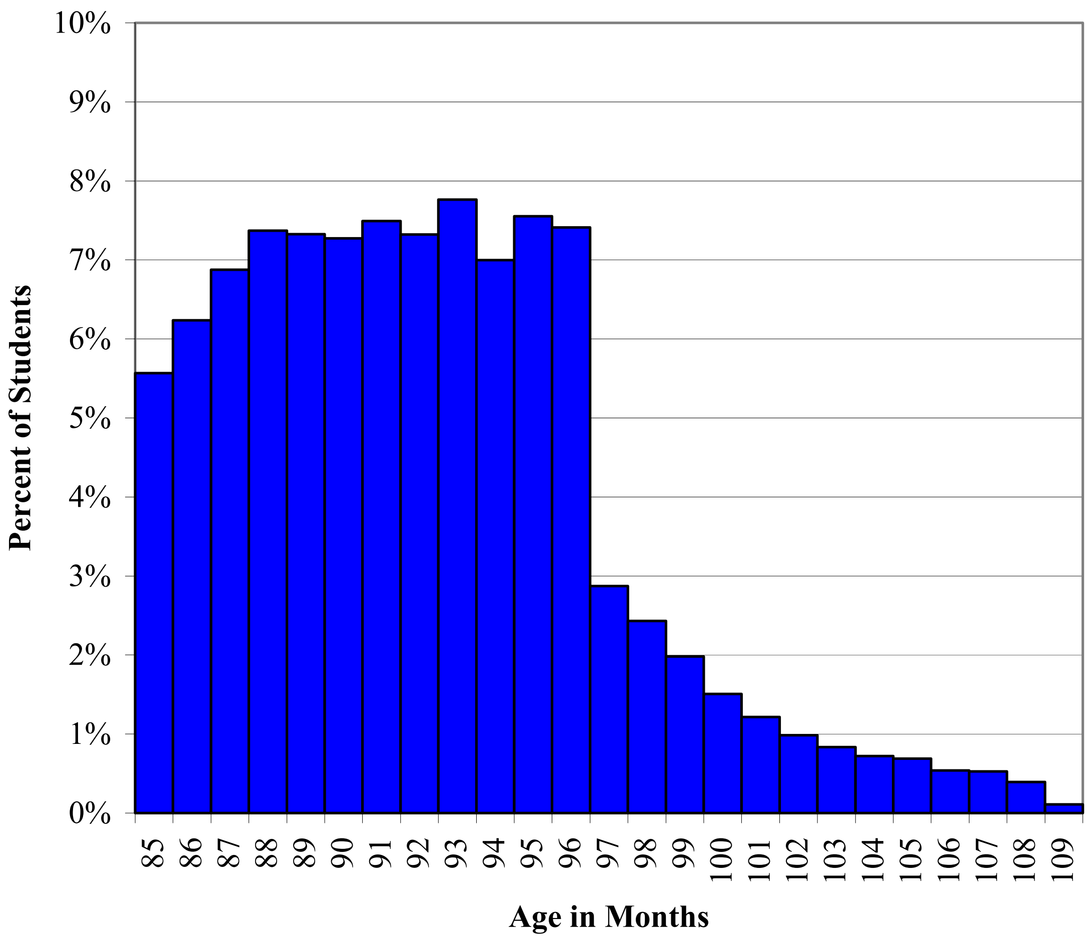


Figure 21 shows that there is a smaller percent of students who are 85 months old relative to the other age normal months (i.e., 86 to 96 months). There is a larger percentage of students with October birthdays (i.e., 86 months) but it is still less, relative to the other age normal months. The percentage of students with September birthdays increases but is still less than the other age normal months. Students with August birthdays (i.e., 88 months) are the first group of students to have a percent that is consistent with the other months in this age normal cohort. There is a smaller percent of students who are 85 to 87 months old relative to their age normal peers (i.e., 88 to 96 months) because some proportion of these "late birthday" students have experienced delayed school entry.

The ages of 97 months and higher represent the students who have been held back. The retained students who have experienced delayed school entry are most likely those closest to the entry age cut (i.e., 97 to 99 months). Student demographic information does not contain information as to why students were held back. However, the number and proportion of "red-shirted" students can be conservatively estimated.

One way is to work with the age normal cohort (i.e., 85 to 96 months) and ignore ages 97 months and higher. First, find the average number of students who are in the age range from 88 to 96 months. For students in figure 21 this value is 35,899. Assume that this is the number of students who should be in each of the months 85 to 87. Next subtract the number of students who are actually in each of the months 85 to 87 from the average of 35,899 (i.e., 85 months: 35,899 – 27,049 = 8,850; 86 months: 35,899 – 30,298 = 5,601; 87 months: 35,899 – 33,410 = 2,489). Sum these values (i.e., 8,850 + 5,601 + 2,489 = 16,941). Finally, divide this sum by the total number of students who should be in the months 85 to 96 if there were no delayed school entry (i.e., $\frac{16{,}941}{430{,}792} = 3.9\%$)[3]. Four percent of the second grade cohort in figure 21 has been academically "red-shirted." This translates into 25 percent of the students with November birthdays (i.e., $\frac{8{,}850}{35{,}899} = 24.7\%$), 16 percent of the students with October birthdays (i.e., $\frac{5{,}601}{35{,}899} = 15.6\%$), and 7 percent (i.e., $\frac{2{,}489}{35{,}899} = 6.9\%$) of the students with September birthdays.

The estimated percents can be interpreted as probabilities. For example, age normal students with November birthdays have a 25 percent probability of experiencing delayed school entry. Age normal students with October birthdays have a 16 percent probability of experiencing delayed school entry.

Another way to estimate the proportion of "red-shirted" students is to look at the whole grade level cohort. There are 485,796 students in this cohort and 71,945 of these second graders have been retained. The retained students represent 15 percent of the total cohort (i.e., $\frac{71{,}945}{485{,}796} = 14.81\%$). It is unlikely that more than ten to twelve percent of kindergarten and first grade students were "flunked" statewide. That means the percent of students who experienced delayed entry would be around three to five percent. This is consistent with the first estimate[4]. Despite controversy in the research literature about academic "red-shirting," a healthy number of students (i.e., $.03 \times 485{,}796 = 14{,}574$ to $.05 \times 485{,}796 = 24{,}290$) in California experience delayed school entry.

In the second estimate it is unlikely that the three to five percent is spread evenly through the whole overage cohort. It is most likely concentrated around the three months closest to the entrance age cut (i.e., 97 to 99 months). These months represent 7.7 percent of the

---

[3] The average value of 35,899 is substituted in the months 85 to 87.

[4] There may be other ways to guesstimate the percent of students who have been "red-shirted." The reader is invited to do so.

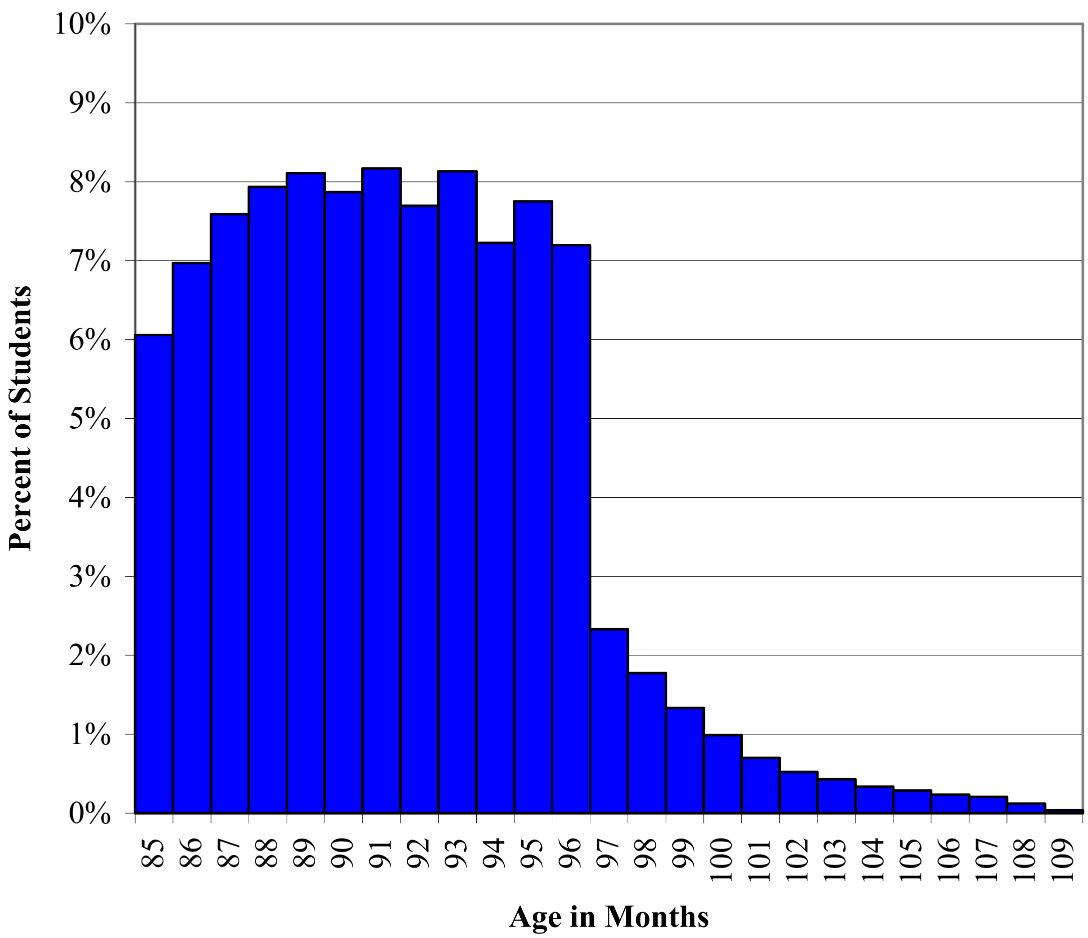
second grade cohort or 37,406 students. This means that delayed entry students make up 39 to 65 percent of the retained students closest to the cut (i.e., $\frac{14,574}{37,406} = 39\%$,

$\frac{24,290}{37,406} = 64.9\%$).

Given the proportion of students who are "red-shirted", the drop in test scores for second grade students who are 97 to 99 months seems particularly shocking. Test scores begin to decline even though 39 to 65 percent of these students have experienced delayed school entry to provide them a cognitive and emotional advantage over their grade level peers.

To evaluate whether data in figure 21 represent a recent phenomenon, the age frequency distribution for students in grade 2 who were tested in spring 1998 was calculated. Figure 22 shows these results.

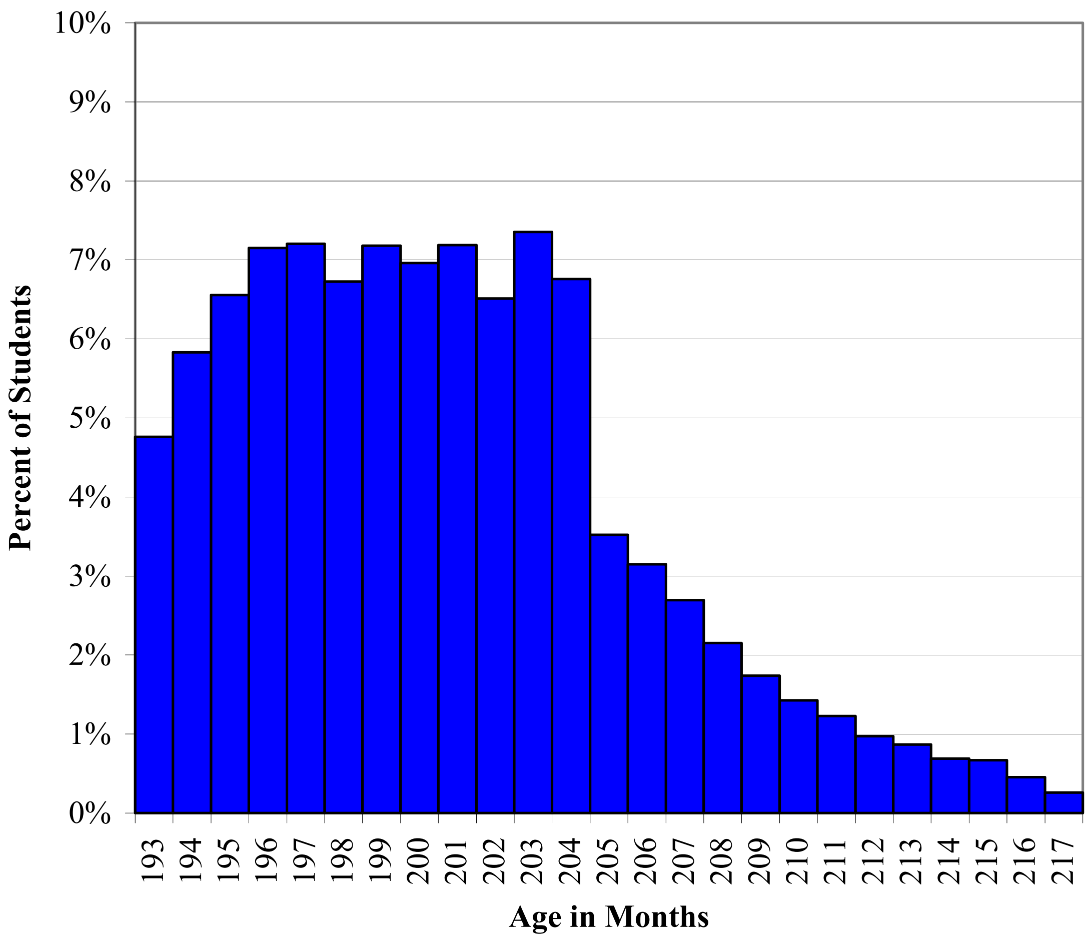
Figure 22. Age of second grade students on December 1, 1997 for the STAR 1998 test, n = 445,707



The pattern is similar to that of second grade students tested in who were tested in spring 2002. It provides some evidence that the practice of delayed entry for students whose birthday is close to the cut-off date has been a practice in California for several years.

To further evaluate whether data in figure 21 represent a recent phenomenon, the age frequency distribution was calculated for grade 11 students who were tested in spring 2002. Eleventh grade students would have been second graders during the 1992-93 school year. Figure 23 shows these results.

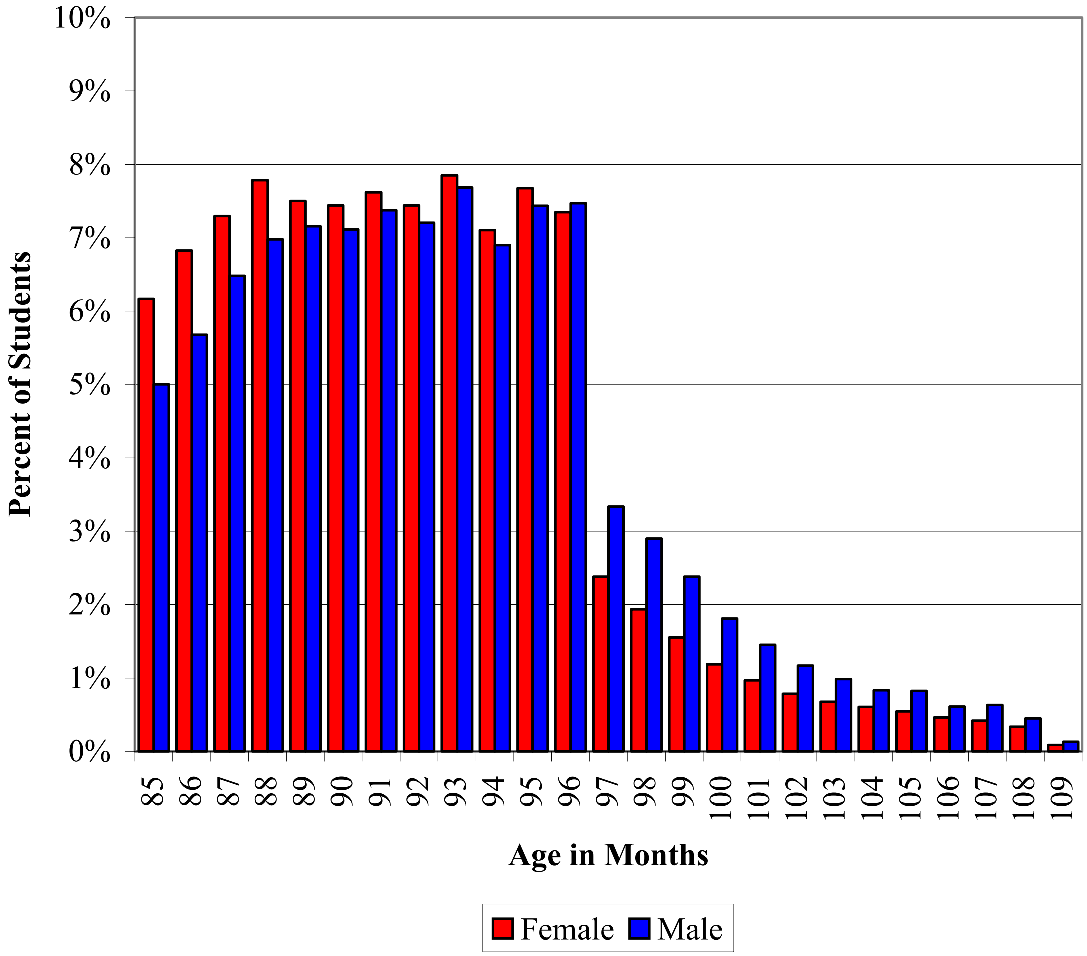
Figure 23. Age of eleventh grade students on December 1, 2000 for the STAR 2002 test, n = 377,434



The pattern is similar to that of students tested in second grade. This is additional evidence that the practice of delayed entry for students whose birthday is close to the cut-off date has been a practice in California for several years.

To evaluate gender differences in delayed school entry, the percent of students in grade 2 was calculated by age in months and by gender. Figure 24 shows these results.

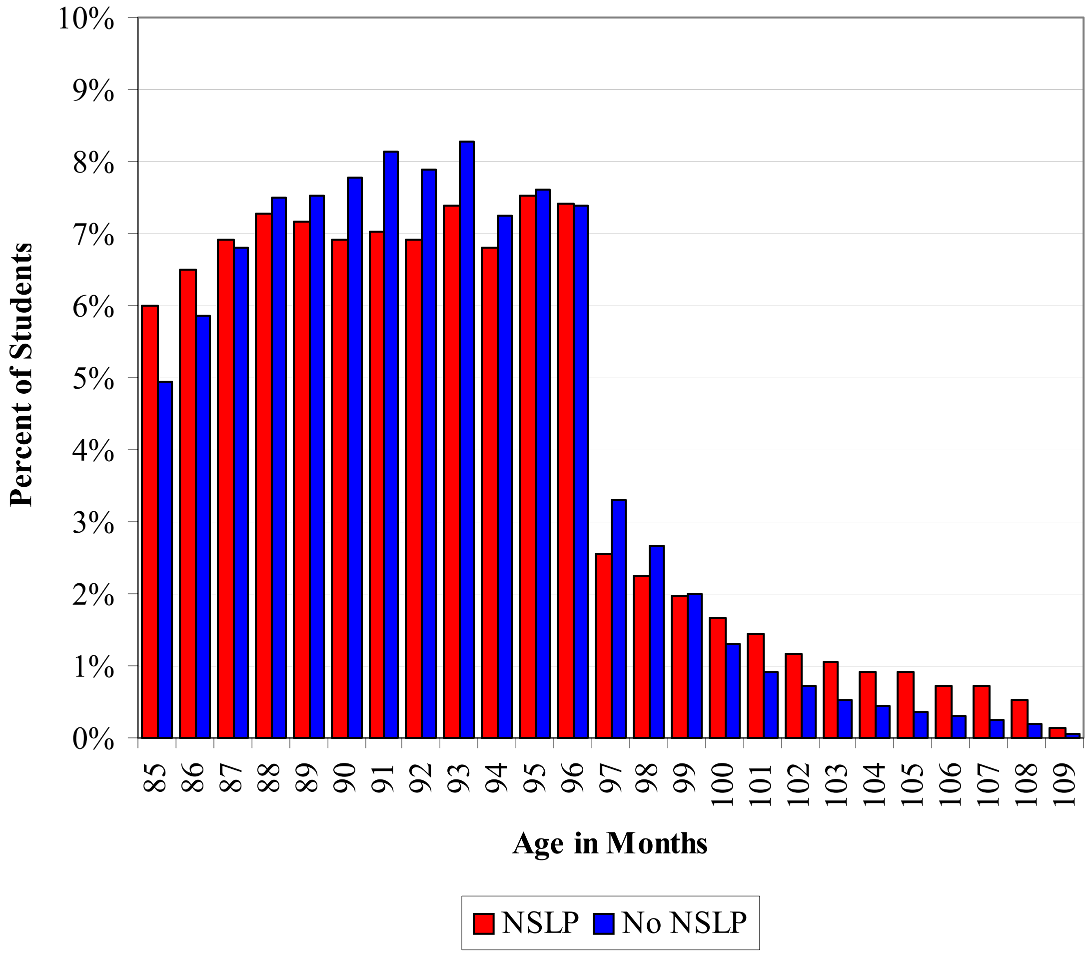
Figure 24. Age of second grade students by gender on December 1, 2001 for the STAR 2002 test, n = 485,556



The research literature indicates that males are more likely to be academically "red-shirted" than females. Using estimation procedure #1, approximately 5 percent of the males and 3 percent of the females experienced delayed school entry. For males this represents 31 percent of the students with November birthdays, 22 percent of the students with October birthdays, and 11 percent of the students with September birthdays. For females this is 18 percent of the students with November birthdays, 9 percent of the students with October birthdays, and 3 percent of the students with September birthdays. These data are consistent with the research literature. For example, age normal males with November birthdays have a 31 percent probability of experiencing delayed school entry while age normal females have a 22 percent probability.

The research literature also indicates that parents identified as higher SES are more likely to delay school entry than those parents identified as lower SES. Figure 25 shows the percent of students in grade 2 by age in months by NSLP.

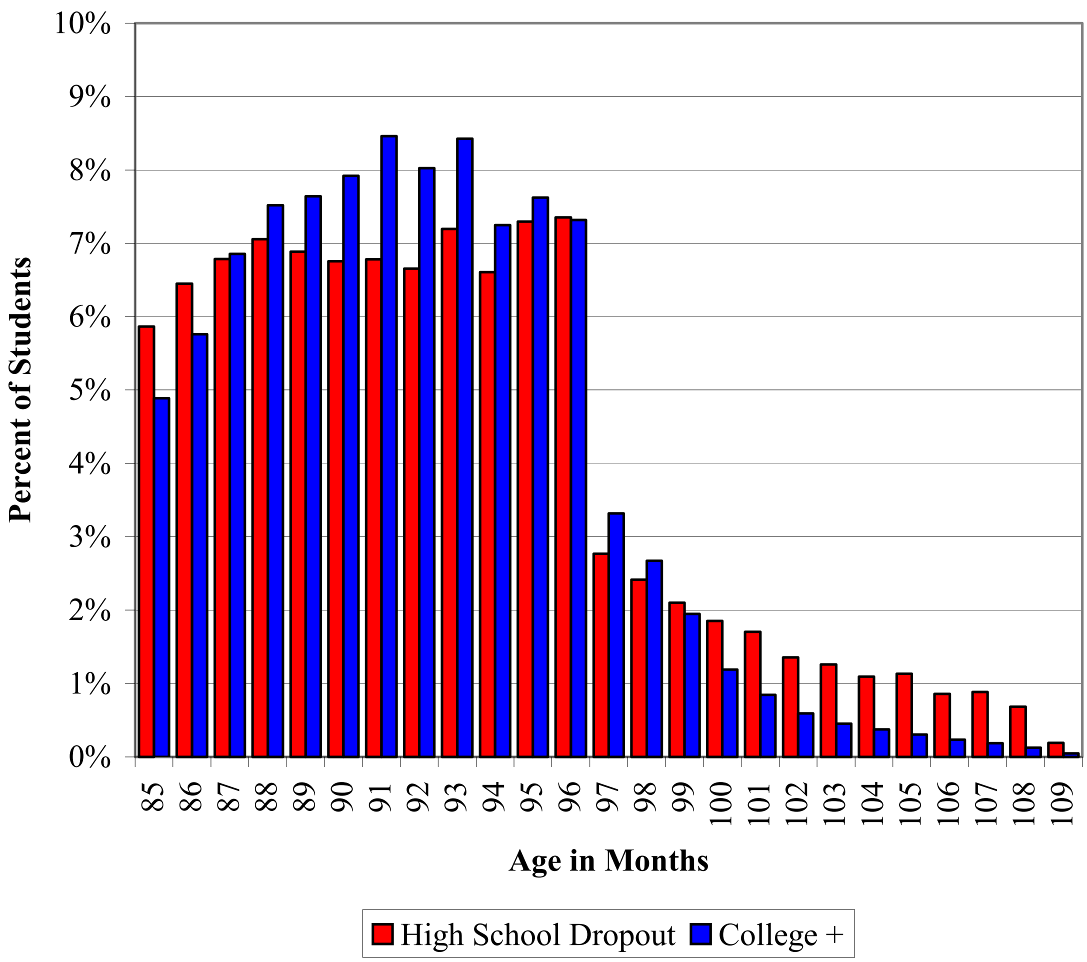
Figure 25. Age of second grade students by NSLP on December 1, 2000 for the STAR 2002 test, n = 485,796



Using estimation procedure #1, these data indicate that approximately 5 percent of the no NSLP and 3 percent of the NSLP students experienced delayed school entry. For no NSLP students this represents 36 percent of the students with November birthdays, 24 percent of the students with October birthdays, and 12 percent of the students with September birthdays. For NSLP students this is 16 percent of the students with November birthdays, 9 percent of the students with October birthdays, and 3 percent of the students with September birthdays. These data are consistent with the research literature. Students who do not participate in NSLP (i.e., higher SES) are more likely to experience delayed school entry than students who do participate (i.e., lower SES).

Parent education is another proxy for SES. Figure 26 shows the percent of students in grade 2 by age in months by parent education. Parent education in this case is represented by "extremes." Low education means that the parent with the highest level of education in the household is a high school dropout. This is an indicator of lower SES. High education means that the parent with the highest level of education in the household has a college degree and has some post college education. This is an indicator of higher SES.

Figure 26. Age of second grade students by parent education on December 1, 2000 for the STAR 2002 test, n = 166,078



These data indicate that approximately 6 percent of the students with the college + parent and 2 percent of the students with the high school dropout parent experienced delayed school entry. For students with the college + parent this represents 37 percent of the students with November birthdays, 26 percent of the students with October birthdays, and 12 percent of the students with September birthdays. For the students with the high school dropout parent this represents 16 percent of the students with November birthdays, 7 percent of the students with October birthdays, and 2 percent of the students with September birthdays. Again, these data are consistent with previous research and the data in figure 25. Students with a parent who has a college degree and some post college education (i.e., higher SES) are more likely to experience delayed entry than students whose most educated parent is a high school drop out (i.e., lower SES).

Discussion

Findings indicate that for students in elementary school on average there is positive linear relationship between age and achievement for age normal peers. That is, on average, older age normal peers perform better academically than their younger classmates. Even though there is positive linear relationship, the difference in average test scores between the oldest and youngest students is not great. The difference in average test scores between the oldest and youngest students gets smaller as grade level increases. By the time students reach 10^{th} grade this difference is negligible. By 10^{th} grade the positive linear relationship between age and achievement has disappeared. There is no academic advantage to being older, even for age normal peers, by the time students reach high school.

For overage students there is on average a negative linear relationship between age and achievement for all grade levels. On average, overage students do less well than their classmates and the older they get the less well they perform. The negative relationship between age and achievement remains constant over time.

The relationship between age and achievement is not content dependent. The relationship is consistent across reading and mathematics test scores. The relationship between age and achievement is also consistent across gender and SES differences.

Even though on average there is a positive relationship between age and achievement for age normal peers and a negative relationship for retained students, there is large variability in the individual test scores. Making strong inferences about student academic performance if age is known is not sound. That is, many of the youngest age normal students perform high and many of the oldest age normal students perform low.

The variability of individual test scores should not be interpreted to mean that it makes no difference as to whether or not students are retained. On average, retained students score lower than age normal peers. Since there are no valid or reliable instruments to identify whom would benefit from retention, retention of any kind is not a sound remediation strategy.

The results also indicate that academic "red-shirting" is practiced and has been in practice in California for several years. This is especially true for students with November birthdays. Consistent with previous research males are more likely than females to experience delayed school entry. Also, students from higher SES households are more likely to experience delayed entry than students from lower SES households.
Even though research evidence does not support delayed school entry, a certain percentage of parents have decided that it is in their children's best interest to hold them back. Whether delayed entry is due to parental beliefs about the increasing curricular demands of kindergarten and first grade or other reasons, the pattern of behavior seems clear.

Contrary to parental beliefs and educational policy, the results from this study argue against modifying entrance age policies, delaying school entry, implementing transitional kindergarten or first grade programs, or retaining students to improve educational achievement. Proponents of these practices argue that they will improve educational achievement. These results do not support that argument.

In addition, although it may be difficult to argue that being overage causes lower achievement, policies and practices that make students older than their age normal peers seem to inversely affect their educational achievement. When students are one year older than their classmates, their average academic performance declines and continues to decline the older they get. Maybe making students different (i.e., older than their grade level peers) lowers their motivation to achieve. The research literature suggests that older students also are more likely to drop out of school.

Learning exists along a continuum. With any group of people and with any content area people will learn at greater and lesser degrees. This occurs for a number of reasons. Three of the most obvious are ability, motivation, and opportunity. Ability, motivation, and opportunity vary with individuals and thus learning also varies.

In public education "lip service" is given to the notion that "all" students are expected to achieve at a certain level at every grade level. Even when grade level expectations are defined, students will master content at different rates. It makes no difference whether there are legislative decrees that include positive and/or negative consequences. There will be differences in student achievement. Student achievement exists along a continuum and making determinations about what it means to achieve mastery or grade level standards is arbitrary. Emrick (1971) stated:

> It is not difficult to show that traditional measurement procedures are inadequate, or at best arbitrary, as a method of identify student skill mastery.

Glass (1978) went even further when he wrote:

> To my knowledge, every attempt to derive a criterion score is either blatantly arbitrary of derives from a set of arbitrary premises.

Delaying school entry or retaining students in other ways to ensure some arbitrary level of achievement is a futile exercise. It cannot be over emphasized that attaining a certain test score is not the same thing as achieving mastery, even if mastery could be defined. At best, schools can identify where students are and move them further along the continuum.

Some students can achieve at greater levels if given additional instruction and time. This is different than the notion of mastery. It simply means that movement along the achievement continuum can be accelerated with additional support. However, if additional time means making students older than their classmates by more than a year, the additional time begins to have a negative effect.

References

Abidin, R.R., Golladay, W.M., Howerton, A.L. (1971). Elementary school retention: an unjustifiable, discriminatory, and noxious educational policy. *Journal of School Psychology*, 9 (4) 410-417.

Angrist, J.D. & Krueger, A.B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87 (418), 328-336.

Bellisimo, Y., Sacks, C.H. & Mergendoller, J.R. (1995). Changes over time in kindergarten holding out: Parent and school contexts. *Early Childhood Research Quarterly*, 10 (2), 205-222.

Brent, D., May, D.C., & Kundert, D.K. (1996). The incidence of delayed school entry: A twelve-year review. *Early Education and Development*, 7 (2), 1996.

Bracy, G.W. (1989). Age and achievement. *Phi Delta Kappan*, 70 (9), 732.

Byrd, R.S., Weitzman, M.L. (1994). Predictors of early grade retention among children in the United States. *Pediatrics*, 93 (3) 481-487.

Byrnes, D.A. (1989). Attitudes of students, parents, and educators toward repeating a grade. In L.A. Shepard & M.L. Smith (Eds.) *Flunking grades* (pp. 34-63). London: Falmer.

Carlson, E., Egeland, B., Jimerson, S., Rotert, M., & Sroufr, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology*, 35 (1), 3-25.

Combs, F.E. & Tanner, C.K. (1993). Student retention policy: The gap between research and practice. *Journal of Research in Childhood Education*, 8 (1), 69-77.

Crosser, S.L. (1991). Summer birth date children: Kindergarten entrance age and academic achievement. *Journal of Educational Research*, 84 (3), 140-146.

DeMeis, J.L. & Stearns, E.S. (1992). Relationship of school entrance age to academic achievement. *Journal of Educational Research*, 86 (1), 20-27.

Dietz, C. & Wilson, B.J. (1985). Beginning school age and academic achievement. *Psychology in the Schools*, 22 (1), 93-94.

Ellwein, M.C., Walsh D.J., Eads, G.M., & Miller, A. (1991). Using readiness tests to route kindergarten students: The snarled intersection of psychometrics, policy, and practice. *Educational Evaluation and Policy Analysis*, 13, 159-175.

Emrick, J.A. (1971). An evaluation for mastery testing. *Journal of Educational Measurement*, 8, 321-326.

Glass, G.V. (1978). Standards and criteria. *Journal of Educational Measurement*, 15 (4), 237-261.

Graue, M.E. & DiPerna, J. (2000). Redshirting and early retention: Who gets the "gift of time" and what are its outcomes? *American Educational Research Journal*, 37 (2), 509-534.

Grissom, J.B. (1988). *Structural equation modeling of retention and overage effects on dropping out of school*. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Grissom, J.B. & Shepard, L.A. (1989). Repeating and dropping out of school. In L.A. Shepard & M.L. Smith (Eds.) *Flunking grades* (pp. 34-63). London: Falmer.

Holmes, C.T. & Matthews, K.M. (1984). The effects of nonpromotion on elementary and junior high school pupils. *Review of Educational Research*, 54, 225-236.

Holmes, C.T. (1989). Grade level retention effects: A meta-analysis of research studies. In L.A. Shepard & M.L. Smith (Eds.) *Flunking grades* (pp. 16-33). London: Falmer.

Kinard, E.M. & Reinherz, H. (1986). Birth date effects on school performance and adjustment: A longitudinal study. *Journal of Educational Research*, 79 (6), 366-372.

Langer, P., Kalk, J.M., & Searls, D.T. (1984). Age of admission and trends in achievement: A comparison of blacks and Caucasians. *American Educational Research Journal*, 21 (1), 61-78.

La Paro, K.M. & Pianta R.C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70 (4), 443-484.

May, D.C. & Kundert, D.K. (1995). Does delayed school entry reduce later grade retentions and use of special education services? *Remedial and Special Education*, 16 (5), 288-94.

Meisels, S.J. (1992). Doing harm by doing good: Iatrogenic effects of early childhood enrollment and promotion policies. *Early Childhood Research Quarterly*, 7, 155-174.

Narahara, M. (1998). Kindergarten entrance age and academic achievement. 21 p. ED 421 218.

Niklason, L.B. (1984). Nonpromotion: A pseudoscientific solution. Psychology in the Schools, 21, 485-499.

Shepard, L. & Smith, M.L. (1988). Escalating academic demand in kindergarten: Counterproductive policies. *Elementary School Journal*, 89 (2), 135-145.

Shepard, L. & Smith, M.L. (1985). *Boulder valley kindergarten study: Retention practices and retention effects*. Boulder, CO, Boulder Valley Public Schools.

Shepard, L. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools*, 34 (2), 85-97.

Smith, M.L. & Shepard, L.A. (1987). What doesn't work: Explaining policies of retention in the early grades. *Phi Delta Kappan*, 6 (2), 129-134.

Smith, M.L. & Shepard, L.A. (1987). Teachers' beliefs about retention. In L.A. Shepard & M.L. Smith (Eds.) *Flunking grades* (pp. 34-63). London: Falmer.

Uphoff, J.K. & Gilmore, J. (1985). Pupil age at school entrance – How many are ready for success? *Educational Leadership*, 43, 86-90.

Uphoff, J.K. (Eds.). (1995). *Real facts from real schools*. Rosemont, NJ: Modern Learning Press.

About the Author

James B. Grissom
Standards and Assessment Division
California Department Education
1430 N Street
Sacramento CA 95814
(916) 319-0361
E-mail: jgrissom@cde.ca.gov
B.A. California State University, San Bernardino, 1970
M.A. University of Colorado, Boulder, 1984
Ph.D. University of Colorado, Boulder, 1988