Review of Educational Research
Fall 1996, Vol. 66, No. 3, pp. 227–268

The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review

Harris Cooper
University of Missouri–Columbia
Barbara Nye
Tennessee State University
Kelly Charlton
James Lindsay
Scott Greathouse
University of Missouri–Columbia

A review of 39 studies indicated that achievement test scores decline over summer vacation. The results of the 13 most recent studies were combined using meta-analytic procedures. The meta-analysis indicated that the summer loss equaled about one month on a grade-level equivalent scale, or one tenth of a standard deviation relative to spring test scores. The effect of summer break was more detrimental for math than for reading and most detrimental for math computation and spelling. Also, middle-class students appeared to gain on grade-level equivalent reading recognition tests over summer while lower-class students lost on them. There were no moderating effects for student gender or race, but the negative effect of summer did increase with increases in students' grade levels. Suggested explanations for the findings include the differential availability of opportunities to practice different academic material over summer (with reading practice more available than math practice) and differences in the material's susceptibility to memory decay (with fact- and procedure-based knowledge more easily forgotten than conceptual knowledge). The income differences also may be related to differences in opportunities to practice and learn. The results are examined for implications concerning summer school programs and proposals concerning school calendar changes.

In 1994, the National Education Commission on Time and Learning (1994) urged school districts to develop school calendars that acknowledged (a) differences in student learning and (b) the major changes taking place in American society. The report reflected a growing concern about school calendar issues on the part of local school boards, administrators, and teachers, especially as the calendar relates to students at risk for academic failure.

In the early years of formal schooling in America, school calendars were designed to fit the needs of particular communities (Association of California

Support for this project was provided by the Center for Research in Social Behavior, University of Missouri–Columbia, and the Center for Excellence for Research and Policy on Basic Skills, Tennessee State University.

School Administrators, 1988). In agricultural areas it was typical for children to attend school for only 5 or 6 months so that they were free to participate in the farming economy, from planting to harvesting. During the same era, urban schools were operating on 11- or 12-month schedules.

By the turn of the century, the implementation of standardized, grade-leveled curricula created pressures to also standardize the amount of time children spent in school. The present 9-month calendar, under which schools are closed in summer, emerged as the norm when 85% of Americans were involved in agriculture. Today, about 3% of Americans' livelihood is tied to the agricultural cycle, but the school calendar has not changed. While alternatives to the traditional school calendar have been in evidence throughout this century, large-scale adoption of alternatives has never taken place. However, for a variety of reasons, a renewed interest in changing the school calendar emerged in the 1980s.

Suggestions for change in the present calendar can be grouped into two alternatives. Some proponents of calendar change call for an extended school year that increases the number of days children spend in school. Indeed, the number of Americans who think children should be spending more time in school has recently been on the increase (Elam, Rose, & Gallop, 1996). Supporters of this idea point out that the United States ranks near the bottom among industrialized nations in the number of days children attend school (Barrett, 1990). Among the more prominent arguments for increasing the number of school days is the potential to (a) increase the amount that students learn and (b) more closely fit the lifestyles of today's families (Gandara & Fish, 1994). Those skeptical of extending the school year express concern about cost, teacher and student burnout, and whether increasing the quantity of schooling necessarily translates into increased achievement (Karweit, 1985; Mazzarella, 1984).

Other proponents of calendar change call for schedules that use school buildings year-round. Under these arrangements, children might or might not attend school for more days, but the summer vacation is no longer sacrosanct (Ballinger, Kirschenbaum, & Poimbeauf, 1987). Instead, children might go to school for 45 days and then get 15 days of vacation, go 60 days and then get 20 days of vacation, or the like. Year-round scheduling has been especially popular in school districts where the need for space is paramount. For example, under either of the calendars just mentioned, children can be placed in alternate vacation sequences; that is, one quarter of students can be on vacation at any given time while the building is in continuous use, thus increasing the number of students a particular school facility can accommodate. Skeptics question whether money is truly saved by year-round scheduling (Merino, 1983) and worry about disruptions to family life (e.g., siblings on different schedules, child care arrangements; Sardo-Brown & Rooney, 1992).

Proponents of both forms of alternative scheduling raise concerns about the possible negative impact of summer vacations on learning. They suggest that children learn best when instruction is continuous, and a 3-month break is simply too long. The long vacation breaks the rhythm of instruction, leads to forgetting, and requires that a significant amount of time be spent on review of old material when students return to school in the fall.

In addition, the long summer break can have a greater negative effect on the learning of children with special educational needs. For example, children who speak a language at home other than English may have their acquisition of English language skills set back by an extended period without its usage. Many states mandate extended-year programs for students with physical or learning disabilities because they recognize these childrens' need for continuous instruction (Katsiyannis, 1991).

And finally, there is growing concern that whatever negative impact summer vacations have on learning might be uneven across children from different economic groups. Tying summer vacations to equity issues, Jamar (1994) wrote, "Higher [socioeconomic status] students may return to school in the fall with a considerable educational advantage over their less advantaged peers as a result of either additional school-related learning, or lower levels of forgetting, over the summer months" (p. 1).

This article describes a review of the research on the impact of summer vacation on children's retention and acquisition of academic skills and materials. We were able to find no prior comprehensive review of the summer vacation literature. In our review, we will examine the research to determine not only the overall effect of summer vacation but also its differential effect for different subject matters and for students with different personal and familial characteristics. We will begin with a description of some of the methodological and conceptual issues associated with studying the effects of summer vacation on achievement. Then we will describe our literature search and the strategy and techniques we used to synthesize the existent research base. Finally, we will discuss our results and their implications for the debate about alternative school calendars.

Methodological and Conceptual Issues in the Studies of Summer Effects

There are three methodological and conceptual issues that require exposition before the summer vacation literature is described. The first issue relates to the length of the summer vacation interval. Summer vacations can vary in their length, and, more importantly, researchers often choose spring and fall testing dates that include some time spent in school. The second issue concerns the alternative metrics available to researchers to express the change in achievement over time. Absolute change in achievement can be measured, or the researcher can derive a measure that expresses change relative to testing norms. The third issue, related to the second, is how relative measures of change can demonstrate summer losses or gains when the norming data should adjust for these changes.

The Length of the Testing Interval

Different states and school districts have different requirements for the number of days children must attend school. Therefore, the length of summer vacations will vary from study to study. Regrettably, we found few studies that provided the exact number of days between the end and beginning of the school year. More importantly, we found no study that claimed to have tested students on the final day of school in spring and the first day of school in fall. Frankly, we doubt that any did. Instead, studies varied in how much instructional time was contained within the summer vacation interval.

The amount of instructional time contained in the testing interval can be critical to a study's ability to accurately estimate a summer vacation's effect on test scores. Figure 1 illustrates the problem. Assume that the dashed line represents the

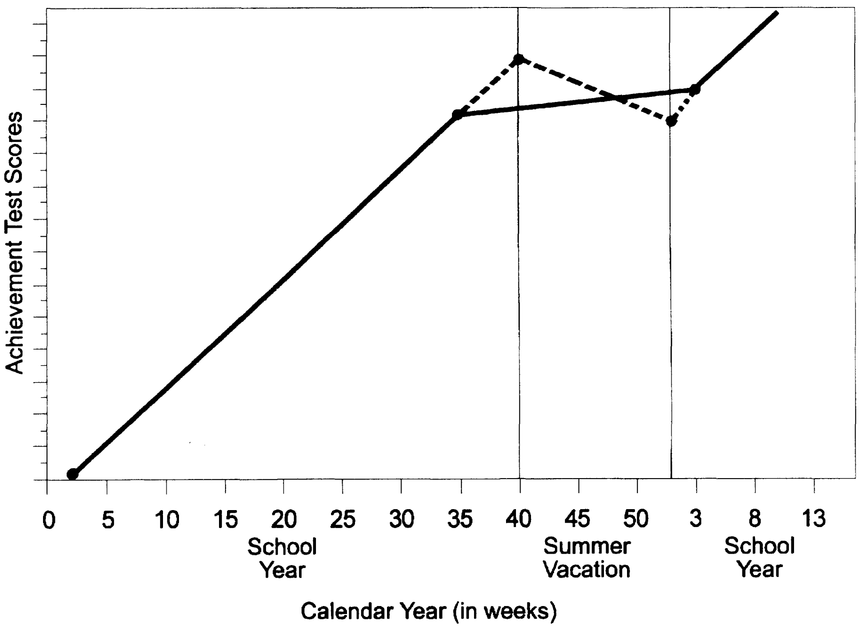

FIGURE 1. *A hypothetical relation between changes in student achievement over summer and spring-to-fall testing intervals*

actual change in students' achievement over summer break and the adjacent weeks; it indicates that a summer loss in achievement has occurred. However, if the spring and fall testings are scheduled such that 3 weeks of instruction take place both before and after vacation, and between the testings, then the test results might seem to suggest a summer gain, as indicated by the solid line. As the testing interval includes more instructional time, the potential for misleading data increases.

We suspect that longer testing intervals in studies of summer vacation correlate more closely with the inclusion of more instructional time than with lengthier vacations. Therefore, in the meta-analysis that follows, we will examine the relation between the length of the testing interval and the effect of summer vacation. If our suspicion is true, and if the trend in achievement change resembles that depicted in Figure 1, we would expect that as the testing interval gets longer the revealed effect of summer vacation on achievement should get more positive.

Alternative Metrics for the Measurement of Achievement Change

The measurement of changes in achievement level is a highly complex topic and one that is vigorously debated in both educational and statistical literature. Here, we will simply describe the most frequently used metrics and assume that interested readers can refer to many other sources for fuller discussions of their properties.

Absolute measures of change. Changes in raw scores are computed by having students take the same test twice and computing the absolute change in the number of items answered correctly. In some instances, researchers interested in absolute change may sample items from the same test or use parallel forms of the same test previously demonstrated to be equivalent. Raw score changes are used when the researcher wishes to gauge the absolute gain or loss in knowledge experienced by the child.

Another measure of absolute change is the standardized score. Standard scores are first normalized and then scaled to a standard mean and standard deviation. Quite often, the standard mean is 50 and the standard deviation is 10. Thus, a child with a score of 45 is performing one half of a standard deviation below the sample mean. Raw score and standardized score measures of absolute change are ordinally equivalent to one another but are not linear transforms.

Relative measures of change. Alternatively, researchers may be interested in studying not absolute gain or loss but rather change relative to a comparison group. Almost universally, the comparison group is a national sample of children of chronological age or grade level equal to that of the sample of interest. Scores of students in the tested sample, then, are measures of how they compare to the national sample. Thus, a child's score does not describe a fixed level of performance; a negative score or change from a first to a second testing can therefore indicate either an absolute loss in performance or a smaller gain in performance than that achieved by the norming sample.

Relative change has been expressed most frequently in terms of grade-level equivalents. In this metric, for example, students at the beginning of the first month of sixth grade in the tested sample who score exactly at the mean for their grade level would receive a grade-level equivalent score of 6.0. A score of 6.2 would indicate performance at a level equivalent to a student who is in the second month of sixth grade. Interestingly, when most standardized tests are normed to national samples, 10 decimal points are used to express the 12 months of the calendar year. The discrepancy is accommodated by assuming that students make 1 month of academic progress over the 3 months of summer.

A second measure of relative change is a percentile rank. As the name implies, a percentile rank expresses student scores relative to where they would place in the distribution of the norming sample, with a score of 50 usually reflecting a mean score.

Again, the measures of relative change will be ordinally identical but are not linear transforms. Additionally, large studies suggest that all four measures of change are highly correlated with one another (see Heyns, 1978). However, the choice of metric can affect the magnitude of the effect size derived from it.

The (Il)logic of Loss Over Summer on Normed Tests

An interesting question arises concerning how it is possible to demonstrate losses or gains on normed tests over any designated period when the norms themselves should take such changes into account. An example will make the issue clearer. Suppose there are 100 questions on a test that is being normed for grade-level equivalents for children in the last month of fifth grade (5.9) and the first month of sixth grade (6.0). Suppose further that the average score at the end of fifth grade turns out to be 50. Thus, a student with a raw score of 50 would be given a grade-level equivalent score of 5.9. Suppose further that the average score at the beginning of sixth grade is 45, representing an absolute loss of 5 correct answers. However, the assigned grade-level equivalent score given a student scoring 45 in the first month of sixth grade would be 6.0. How, then, can "unusual" losses or gains be demonstrated on normed tests over summer vacation?

There are at least two ways that "unusual" changes in grade-level equivalents can occur over summer. First, the students sampled in the study of the effects of summer vacation might differ in their rate of academic maturation from the sample of students used to norm the achievement test. Thus, because the study sample might be selective, their rate of achievement change might be faster or slower at this particular developmental period than the national averages used to norm the test.

Relatedly, even if the study sample is identical to the norming sample, it is still the case that different subsamples within the study sample can show differential rates of change that average out to be identical to the overall norms. Thus, while a national sample of children in a study of summer vacation effects might not be expected to show "unusual" changes in grade-level equivalents, it would not be surprising to find that when the study sample is systematically disaggregated different degrees of change emerge for different subsamples.

A second reason why "unusual" changes might appear on normed tests is that it is unlikely that the achievement test has been renormed on a monthly basis. Instead, the test might have been normed once, twice, or three times during the school year, and then linear extrapolations might then have been used to estimate other monthly norms. If this is the case, large gains or losses for particular months would be smoothed out. For example, suppose a test is normed twice a year, in March and October. For fifth and sixth graders, the grade-level equivalents for these months would be 5.6 and 6.1, respectively. Suppose further that the average March raw score is 45 and the average October raw score is 55. A linear extrapolation would set a June score of 51 at a grade-level equivalent of 5.9 and a September score of 53 at 6.0. Finally, suppose students actually drop in absolute score over summer, say, from 51 to 49, but then gain dramatically in the first month of school, say from 49 to 53. If this were the case, and given the assumption of a smooth linear ascent when scores were normed, then the summer decrease followed by the September increase would appear as "unusual" changes in grade-level equivalents.

Methods of Research Retrieval and Synthesis

Literature Search Procedures

Three methods of literature search were used in order to identify research reports that addressed the question of the effect of summer vacation on student achievement. First, the ERIC and PsycLIT computerized reference databases were searched; the former contained reports published between 1982 and 1994, and the latter contained reports published between 1987 and 1994. Both databases were entered using the keywords *summer loss, summer vacation, summer break, summer intercession, summer school,* and *seasonal variations* alone and in combination with related terms (e.g., *learning*). Second, the reference sections of recent reports were examined for potentially relevant material. Finally, several researchers known to be active in the field were contacted and asked to provide recent papers.

Search Outcomes and Synthesis Strategy

The literature search uncovered 39 research reports that contained descriptions of empirical studies meant to test the effect of summer vacation on school achievement. However, we found that few studies published more than 20 years ago (prior to 1975) contained enough information to provide data suitable for inclusion in a meta-analysis. Also, because of the age of these studies, it is not clear that their results would be relevant to today's school children. Therefore, we decided to present the results of our review in four sections. First, we will describe each of the 26 summer break studies conducted before 1975. The results of these studies will then be synthesized by employing a vote-count method of the simple direction of findings. Then, the 13 studies conducted after 1975 will be described. This will be followed by a formal meta-analysis of the cumulative findings of these 13 studies.

Early Studies of Summer Effects on Achievement

The earliest known study of summer loss was reported in 1906 by William White, a professor of mathematics at the State Normal School in New Paltz, New York. Seven pupils were tested on math computation in June and then retested in September. Summer loss was found for speed but not accuracy. No statistical tests were conducted.

In 1919, Garfinkel examined the June and September math computation scores of 747 fifth, sixth, and seventh graders. He found a general decrease in both speed and accuracy. Garfinkel also reported loss scores for children categorized by whether their primary summer activity had been play, study (summer school), or work (employment). Garfinkel found comparatively less summer loss for children engaged in work and concluded that "the child who keeps his brain alert by meeting actual problems of life is in a better position to retain what he has learned than the one who plays or studies" (p. 48). No attempt was made to control for other individual differences that might have been confounded with summer activity.

Five years later, Brueckner and Distad (1924) examined the June and September reading scores of 315 first graders. They reported no general summer loss but did suggest that there was greater variation in September scores than in June scores. Further, they noted that "the lowest intelligence group had the greatest loss" (p. 707), although there were only a few students in this group and no statistical inference tests were conducted. Our examination of their data suggested a summer loss for all student groups, ranging from 21% to 67%, with no consistent relation to intelligence. This was the first empirical test of possible differential effects of summer vacation on children with different abilities.

A year later, Patterson and Rensselaer (1925) reported a study of summer loss in a group of 149 fourth through eighth graders. Tests in both reading and math were given. There were losses in math scores across all grades. Reading comprehension scores diminished for fourth and fifth graders but increased for sixth, seventh, and eighth graders. Patterson and Rensselaer examined scores separately for three groups of students categorized by IQ scores. They found that students with the lowest IQ scores showed a gain in reading achievement over the summer, while students of average and above-average IQ showed reading losses. For computation scores, the fourth and sixth graders with the highest IQs showed the greatest loss; however, in an inconsistent finding, the seventh and eighth graders with the highest IQs showed math gains, while other students showed losses. Again, however, no statistical tests were used.

The 1920s produced seven more studies of summer loss. Noonan (1926), in a study of fifth and sixth graders, found no loss in reading and a loss in math computation and spelling that she labeled "so small that it may be ignored for practical purposes" (p. 67). Elder (1927) found silent reading in third through sixth graders improved for better readers and decreased for poorer readers, creating greater group heterogeneity. Kramer (1927) found more forgetting over the summer in arithmetic for fifth graders labeled "bright" or "average" than for those labeled "dull." Nelson (1928) reported summer loss for third, fifth, and seventh graders in math computation and spelling. Several fall retests revealed that the losses took from 2 to 6 weeks to recoup depending on the subject matter and grade. Bruene (1928) found summer gains in reading and losses in math for fourth, fifth and sixth graders. Much of the reading gain was attributed to students with IQ scores above 110. Bruene also found summer losses in language usage and history, but gains in science, for fifth and sixth graders. Irmina (1928) reported inconsistent summer effects on achievement, and Morgan (1929) reported summer losses in math computation, math problem solving, and reading comprehension.

Interest in summer loss as a research topic diminished during the 1930s and 1940s, and only six new empirical investigations appeared in an 18-year period. The first two (Kolberg, 1934, with seventh graders; Schrepel & Laslett, 1936, with eighth and ninth graders) showed summer to have differential effects based on intelligence or mental age, such that students of lower intelligence showed more detrimental effects. Keys and Lawson (1937) found summer losses in mathematics and gains in reading among fourth, fifth, and sixth graders. Lahey (1941) found losses in math fundamentals but gains in math problem solving, with retention positively correlated with intelligence. Cook (1942) found a summer loss in first and second graders across most subjects but found that the amount of loss was negatively associated with the amount of studying students did and that students of higher intelligence studied more. Bender (1944, reporting a study conducted in 1937) found no relation to intelligence and a mixed pattern of general gains and losses depending on the subject matter.

Twenty years passed before the next summer loss study appeared. When summer loss did reemerge as a topic of study, measurement instruments were dramatically improved, sample sizes were considerably larger, and the use of inferential statistics was commonplace. However, the central themes of research remained the same: Did summer loss occur, and, if so, did it have different impacts for different grade levels, different subject areas, or students with different abilities or backgrounds?

Parsley and Powell (1962) examined the effects of summer vacation on the achievement of 1,080 second through seventh graders chosen because they were of average intelligence. The six subtests of the California Achievement Test showed average grade-equivalent summer losses in math fundamentals and spelling and gains in math reasoning, reading comprehension, vocabulary, and English mechanics. The researchers also tested for gender and grade level effects but found no consistent patterns, although their analytic strategy (involving the calculation of 66 t tests) was less than optimal.

Botwin (1965) conducted a dissertation on summer vacation that examined its effect on both math and reading achievement for second through seventh graders. However, separate analyses of variance were conducted for each of six subtests, and grade level was included in the analyses as a categorical rather than a continuous variable, making interpretation of the statistical results problematic. Our use of the consistency in direction of comparisons across five of six grade levels as the criterion for "significance" suggested summer loss in math fundamentals and spelling, gains in math reasoning and English mechanics, and no consistent findings for vocabulary or reading comprehension. Botwin included intelligence as a factor in the analyses but in a manner that largely defied interpretation. For English mechanics, however, an inspection of means suggested that the summer gain increased with higher intelligence scores and that the lowest intelligence group actually lost English mechanics skill.

Scott (1967) reported summer vacation retention scores as part of a study meant to compare two math curricula. With samples ranging in size between 100 and 200 students at each elementary grade level, Scott found summer loss at every grade on a measure combining both computation and concepts. Regrettably, because summer loss was not the focus of the study, statistical tests of the reliability of the losses was not presented, nor could they be calculated from the reported information.

Arnold (1968) examined the reading and vocabulary summer retention scores of "disadvantaged" Mexican American third graders taught using two experimental programs and one traditional program. For the traditional program, statistically significant losses in reading comprehension were found on each of three tests. No significant differences were found in vocabulary, and whether gain or loss was evidenced was inconsistent across tests. Arnold's report was the first that permitted a rough calculation of the effect of summer vacation (based on the sample size and significance level). The d-index was .44, which indicates that students lost about 4/10 of a standard deviation in reading comprehension scores between spring and fall.

Beggs and Hieronymus (1968) compared spring and fall scores on the Iowa Test of Basic Skills for a large sample of fifth and sixth graders. Consistent losses were found in math concepts and problem solving, reading comprehension, spelling, and English usage. Greater losses were found in reading for students scoring in the lower percentiles on the Iowa test. For vocabulary, losses were found for students in the lower percentiles on the test, while gains were found for students scoring in the higher percentiles. Again, no estimate of effect size could be derived from the reported data.

Hayes and Grether (1969) conducted an analysis of reading achievement and word knowledge scores for close to 370,000 second through sixth graders attending New York City schools. These authors used the school as the unit of analysis (there were over 600 schools) and divided schools into six groups dependent upon the number of minority students and students receiving free lunch. There were overall gains in both reading and vocabulary scores. However, the summer interval between testings did include about 1 month of instructional time in both spring and fall. More importantly, there was a strong linear relation between the free-lunch and minority student composition of schools and the results of testing. Poorer schools and/or schools serving larger minority populations showed average losses in reading and vocabulary, while richer, majority schools showed gains.

Fitzsimmons's (1969) dissertation reported summer losses in reading but gains in other language skills. She also reported greater losses in reading among students with higher achievement scores but not among students with higher IQs.

In the last study conducted in the 1960s, Soar and Soar (1969) reported gains on all subtests of the Iowa Test of Basic Skills. However, these authors cautioned that the summer interval actually included the last and first months of school, including fall review, "which presumably results in an advantage for the fall testing" (p. 581).

Only one additional study of the effects of summer vacation was conducted during the first half of the 1970s. Mousley (1973) conducted a small study of summer effects on reading and found no significant differences.

A Vote-Count Synthesis of Early Studies of Summer Effects on Achievement

Table 1 displays information on the 26 studies conducted prior to 1975. For purposes of summarization, the results are grouped into six subject matter categories: math computation; math concepts, problem solving, or reasoning; reading comprehension; spelling; all other language (e.g., vocabulary, literature, grammar); all other subjects (e.g., history, science). Using these six distinctions and a few based on different grade levels within studies, we were able to generate a total of 86 discrete comparisons from the 26 research reports. These are listed in the column labeled "Findings" in Table 1.

Table 2 contains a summary count of comparisons based on whether gains or losses in achievement were found over the summer months, with gain or loss based on the simple direction (not statistical significance) of a finding. In this count, we have ignored the important issues of the independence of data points and whether the comparison involved an absolute or relative metric. Further, we did not weight each spring-versus-fall comparison by its sample size. Thus, the conclusions we draw from Table 2 need to be viewed as imprecise and suggestive only.

Table 2 reveals that 48 of the 80 comparisons in which gain or loss occurred indicated a loss in skill over the summer. This overall result is a statistical trend, but not significant, using a simple two-tailed sign test, $p < .08$. However, results differ dramatically for different subject areas. All 17 comparisons of math computation before and after summer break indicated a loss in math skill, $p < .0001$. Tests involving math concepts, reasoning, or applications showed relatively equal instances of summer gains and losses. Reading comprehension results were split, with 10 of 17 comparisons showing gains in test scores. All 11 comparisons of spelling revealed summer losses, $p < .001$. The set of comparisons involving language skills other than reading or spelling revealed 9 of 14 comparisons showing gains. Finally, examination of other subject areas indicated that summer vacation led to gains in 6 of 8 subject areas.

We would interpret the early studies as indicating a clear loss of mathematical computation and spelling skills over the summer months but no summer loss in other subjects. We would qualify this assertion by noting that the math studies

TABLE 1

Early studies of the effects of summer vacation on academic achievement

First author	Year	Sample size	Grade	Metric	Findings	Content area	Moderators
White	1906	7	2, 7	Abs	Loss	Math computation	
Garfinkel	1919	747	5–7	Abs	Loss	Math computation	Less loss among employed students
Brueckner	1924	315	1	Abs	Loss	Reading comprehension	No relation to IQ
Patterson	1925	149	4–6	Abs	Loss	Math computation	More loss in math comp for higher IQ
			7–8		Loss	Math computation	Gain in math comp for higher IQ only
			4–5		Loss	Reading comprehension	Gain in read comp for lower IQ only
			6–8		Gain	Reading comprehension	
Noonan	1926	124	5–6	Abs	Loss	Math computation	
					Gain	Reading comprehension	
					Loss	Spelling	
Elder	1927	203	3–6	Abs	Gain	Reading comprehension	Loss for poorer readers Gain for better readers
Kramer	1927	150	5	Abs	Loss	Math computation	More loss for higher IQ
Nelson	1928	123	5	Abs	Loss	Math computation	
			7		Loss		
			3		Loss	Spelling	
			5		Loss		
			7		Loss		

TABLE 1 *(continued)*

First author	Year	Sample size	Grade	Metric	Findings	Content area	Moderators
Bruene	1928	69	4–6	Rel	Loss	Math computation	Gain in reading occurred for higher IQ only
					Loss	Math reasoning	No relation in math to IQ
					Gain	Reading comprehension	
					Loss	Spelling	
			5–6		Loss	Language usage	
					Gain	Science	
					Loss	History	
Irmina	1928	841	2–7	Abs	Loss	Language	
					Gain	Literature	
					Loss	Science	
					Gain	History	
Morgan	1929	38	6	Rel	Loss	Math computation	
					Loss	Math problem solving	
					Loss	Reading comprehension	
Kolberg	1934	163	7	Abs	Gain	History	Gain for higher IQ Loss for lower IQ
Schrepel	1936	121	8–9	Abs	Loss	Math computation	More gain or less loss for students with higher mental ages on 16 of 22 subtests
					Gain	Reading comprehension	
					Gain	Language	
					Gain	History, geography	
			8		Gain	Math reasoning	
			9		Loss	Math reasoning	

Keys	1937	164	4–6	Rel	Loss	Math computation	
					Loss	Math problem solving	
					Gain	Reading comprehension	
					Loss	Spelling	
					Gain	English, literature	
					Gain	History, science	
Lahey	1941	229	9	Abs	Loss	Math fundamentals	More retention for higher IQ
					Gain	Math problem solving	
Cook	1942	52	1–2	Abs	Loss	Reading comprehension	Less loss for higher IQ More loss for less study
Bender	1944	1,592	3–8	Abs	Loss	Math fundamentals	No relation to IQ
					Gain	Math reasoning	
					Gain	Reading comprehension	
					Gain	Word knowledge	
					Loss	Composition	
					Loss	Spelling	
Parsley	1962	1,080	2–7	Rel	Loss	Math fundamentals	No relation to student gender or grade level
					Gain	Math reasoning	
					Gain	Reading comprehension	
					Loss	Spelling	
					Gain	Vocabulary	
					Gain	English mechanics	
Botwin	1965	768	2–7	Abs	Loss	Math fundamentals	More gain in English mechanics for higher IQ
					Gain	Math reasoning	
					No pattern	Reading comprehension	
					Loss	Spelling	
					No pattern	Vocabulary	
					Gain	English mechanics	

TABLE 1 *(continued)*

First author	Year	Sample size	Grade	Metric	Findings	Content area	Moderators
Scott	1967	1,306	1–6	Abs	Loss	Math computation	
					Loss	Math concepts	
Arnold	1968	115	3	Abs	Loss	Reading comprehension	
					No effect	Vocabulary	
Beggs	1968	2,160	5–6	Rel	Loss	Math concepts	More loss in reading for lower ITBS
					Loss	Reading comprehension	Loss in vocabulary for lower ITBS
					Loss	Spelling	Gain in vocabulary for higher ITBS
					No pattern	Vocabulary	
					Loss	English usage	
Hayes	1969	370,000	2–6	Rel	Gain	Reading comprehension	Loss in both reading and vocabulary
					Gain	Vocabulary	for lower SES, gain for higher SES
Fitzsimmons	1969	440	4–5	Abs	Loss	Reading comprehension	Generally greater losses for higher
					Loss	Spelling	achievement
					Gain	Vocabulary	No difference in loss based on IQ
					Loss	English usage	No difference between sexes
					Gain	Maps, graphs, tables	
Soar	1969	189	6	Abs	Gain	Math concepts and problems	
					Gain	Reading comprehension	
					Gain	Vocabulary	
Mousley	1973	3 classes	3	Abs	No effect	Reading comprehension	
					No effect	Vocabulary	

Note. Determination of loss or gain is based on direction of means, not statistical significance. Abs = absolute, Rel= relative. ITBS = Iowa Test of Basic Skills.

TABLE 2
Vote-count summary of directional findings of early studies (1906–1973) of summer vacation effects on academic achievement

	Math computation	Math concepts	Reading comprehension	Spelling	Other language	Other subjects	Total
Gain	0	6	10	0	10	6	32
Loss	17	6	7	11	5	2	48
Sign test p-level	.0001	ns	ns	.001	ns	ns	.08

Note. There were six tests of summer vacation effects that revealed no effect or pattern.

were clustered in Grades 4 through 9, so inferences about early primary grades and about high school should not be made.

In addition, the early studies tested several potential moderators of the effects of summer vacation. These are described in the last column of Table 1. The most frequently tested moderator of summer vacation effects was students' intelligence; this was tested in 11 of the 26 studies. We could identify no consistent pattern of results across the studies relating a measure of intelligence to achievement gain or loss over the summer. Roughly equal numbers of studies showed positive relation, negative relation, and no relation to summer change.

Three studies sought to determine if the students' initial (spring) scores on an achievement test were related to summer gain or loss on the same test. Of course, such comparisons are vulnerable to regression effects. However, two of the three comparisons indicated that students who scored lower on the test in the spring lost more, and/or that students with higher spring scores gained more over summer. These findings are counter to those predicted by the regression phenomena.

In sum, then, a reasonable interpretation of studies conducted prior to 1975 would suggest that summer vacation had a detrimental effect on the math computation and spelling skills of students in middle grades. Also, there was no consistent effect of intelligence on the impact of summer vacation. A single test of gender as a moderating factor revealed no difference, while a single test of students' socioeconomic status suggested that summer vacation led to reading and vocabulary gains for students of higher socioeconomic status but losses for those of lower socioeconomic status.

To reiterate, however, it should be kept in mind that these early studies were of uneven quality. Many did not use statistical tests, and our conclusions were based on only the crudest synthesis procedures. Thus, we would place little weight on these conclusions unless they are corroborated by more recent, better conducted studies.

Studies of Summer Effects on Achievement Reported After 1975

We were able to identify and retrieve 14 research reports appearing in the past 20 years that contained data from 13 studies assessing the impact of summer vacation on student achievement. Below, each report will be described briefly in its historical context, with one exception. The Sustaining Effects Study, a major government-sponsored national survey, was the subject of multiple reports and interpretations, so it will be treated separately. After the historical review, the results of the meta-analysis will be presented. Important characteristics of each report are given in Table 3.

The first report on the effects of summer vacation to appear in the past two decades described a master's thesis project conducted by Grenier (1975). The study examined summer effects on the math skills of 763 seventh graders. One important and unique feature of Grenier's study was that she administered the fall testing to four separate groups of students at 2-week intervals beginning in late August. This allowed her not only to test for changes over summer but also to examine how long it took for students to recoup any summer losses. She found an initial summer loss in math computation but gains in concepts and problem solving. The loss in computation disappeared by the second testing. Grenier's study contributed one independent sample of 154 students (those taking the fall achievement test at the earliest date) to the meta-analysis.

Pelavin and David (1977) reported a set of studies that proved highly controversial. They examined five evaluations of compensatory education (CE) programs to test the hypothesis that the time period on which an evaluation is based can influence its outcome. Noting the possibility that CE students might be especially prone to experience achievement losses over summer, Pelavin and David were concerned that evaluations that measured achievement change from fall to spring might give very different (and larger) estimates of the effect of the compensatory program than estimates based on fall-to-fall intervals. To test their hypothesis, Pelavin and David obtained data from program evaluations conducted in Omaha, Nebraska, and four junior high schools in California. Each evaluation contained data collected over two summer vacations. They concluded that "large increases in school-year achievement are not sustained even until the next fall" (p. iii). The finding, based on grade-level equivalent scores for over 3,800 students (in 16 independent cohorts), was "remarkably consistent across achievement tests, grade level, subject area, and program" (p. iii). Included in Pelavin and David's samples were students from both public and private schools. Both groups showed summer loss. For our meta-analysis, we could identify 16 independent samples of students in Pelavin and David's reports, ranging in size from 36 to 980 students.

In a report to Congress, the National Institute of Education (NIE, 1978) also addressed the issue of summer loss in the context of CE and took issue with Pelavin and David (1977). Employing a subsample of approximately 3,000 students drawn from the more comprehensive Instructional Dimensions Study (IDS), the NIE examined changes over summer in reading and math total scores on the Comprehensive Test of Basic Skills (CTBS) for first and third graders who were either in CE programs or not. Non-CE first graders showed 1-month grade-level equivalent gains over summer in both math and reading. CE first graders showed a 1-month gain in math over summer but no change in reading. Non-CE third graders showed a 2-month gain in math and a 6-month gain in reading. CE third graders showed a 1-month loss in math and a 3-month gain in reading. Thus, non-CE students showed either expected or greater than expected relative progress over summer. CE students showed relative loss in two comparisons to national norms, and absolute loss in one. In all instances, comparisons between CE and non-CE students indicated that changes over summer favored non-CE students. Regrettably, the NIE did not report standard deviations, which meant that effect sizes could not be calculated for this study, and the study could not be included in the meta-analysis. However, this study did contribute eight independent samples to our calculation of changes in grade-level equivalent scores.

The NIE (1978) suggested three possible reasons for the difference between their findings and those of Pelavin and David (1977), and for why the IDS findings were more trustworthy. First, they noted that the IDS was larger and methodologically sounder than the earlier studies (e.g., IDS tests were administered by study employees rather than school personnel). Second, the earlier studies used various achievement tests that differed in content, format, and assumptions about growth over the summer period. Finally, the NIE noted that the IDS data were more recent than those of Pelavin and David, and that an improvement of programs in the intervening time may have removed the conditions that led to summer loss.

Probably the most frequently cited study of summer effects is Heyns's (1978) research on Atlanta school children. She undertook her study with the dual purposes of showing that "schooling has a substantial independent effect on the achievement of children and that the outcomes resulting from schooling are far more equal than those that would be expected based on the social class and racial origins of sample children" (p. 9). To demonstrate her point, she compared test score changes in word recognition when school was in session with changes when school was out. Word recognition was chosen as the index of achievement from the Metropolitan Achievement Test because it was the most reliable subtest and most highly correlated with the principal component extracted from a factor analysis of all nine subtests. Heyns found not only that going to school did indeed improve achievement but also that "summer learning is considerably more dependent on parental status than is learning during the school year" (p. 93). In her data, summer vacation widened the gap in achievement between rich and poor, and between White and Black. Finally, Heyns tested for the effects of IQ and found that IQ was positively related to summer gain. However, including IQ in multiple regression equations did not remove the effects of students' socioeconomic status or race. Heyns's study contributed eight independent samples to the meta-analysis (two grades by two racial groups by two income levels). In addition, for fifth graders, effect sizes could be calculated for both raw scores and grade-level equivalent scores.

Gastright (1979) estimated the effect of summer vacation on the achievement of Title 1 students in the Cincinnati school district. Two independent samples of students were followed for two summers. Three of the four estimated effects found a summer loss on a relative measure of achievement; the fourth estimate revealed no effect. Absolute changes showed summer gains in three instances and a loss in one instance. However, *d*-indexes could not be calculated for the absolute measure because standard deviations were not reported.

In a study meant to examine gender differences in achievement, Hawn, Ellett, and Des Jardines (1981) examined the mathematics scores of first and second graders from Georgia, South Carolina, Mississippi, Virginia, and Idaho. The school districts were participating in a Follow Through program and were drawn from districts described as "economically deprived." While the authors were not interested in the effects of summer vacation per se, their data indicated that three of four independent cohorts of students experienced a loss in relative achievement over summer and that the loss did not differ for boys and girls.

TABLE 3

Studies of the effects of summer vacation published since 1975

First author	Year	State or country	N	SES	Grade(s)	Other	Content area	Test	Metric	+/–	No. *d*s	No. DGLEs
Grenier	1975	GA	154		7 to 8		Math computation	CTBS	Rel	–	4	4
							Math concepts	CTBS	Rel	+		
							Math application	CTBS	Rel	+		
							Math application	LSMA	Rel	+		
Pelavin	1977	NE	2,627	Low	3 to 8	Pub	Reading comp	GMRT	Rel	–	29	29
			515		4 to 8	Pri				–		
		CA	432		7 to 8	Pub	Math total	CTBS	Rel	–		
			248				Reading total	CTBS	Rel	–		
NIE	1978	USA	296	Mid	1 to 2		Math total	CTBS	Rel	+	0	8
			435				Reading total			+		
			305		3 to 4		Math total			+		
			178				Reading total			+		
			344	Low	1 to 2		Math total	CTBS	Rel	+		
			97				Reading total			=		
			512		3 to 4		Math total			–		
			306				Reading total			+		
Heyns	1978	GA	1,127	All	5 to 6		Word knowledge	MAT	Abs	–	12	8
									Rel	+		
			513		6 to 7				Abs	+		
Gastright	1979	OH	295	Low	2 to 3		Reading total	MAT	Rel	–	4	0
					3 to 4					=		
			175		3 to 4					–		
					4 to 5					–		
Hawn	1981		324	Low	1 to 2	Males	Math total	MAT	Rel	–	4	0
			272			Females				+		
			324		2 to 3	Males				–		
			301			Females				–		

Klibanoff	1981	USA	≈39,000	Mix	1, 2, 3, 4, 5		Math total	CTBS	Rel	+	10	0
							Reading total			+		
McCormick	1981	Canada	59	Mix	K to 1		Reading total	LWRT	Abs	+	1	0
Shaw	1982	CA	128		Mixed	Reg ed	Math total	WRAT	Abs	–	6	0
			108			Res sp				–		
			58			Spec cl				–		
						Reg ed	Reading recog			+		
						Res sp				–		
						Spec cl				–		
Johns	1984		322	Mid	1 to 7	Males	Reading comp	GMRT	Rel	–	28	28
			318			Females				–		
						Males	Vocabulary			+		
						Females				+		
Wintre	1986	Canada	182	Mid	1, 3, 5		Math computation	MAT	Abs	–	0	0
							Math concepts			+		
							Reading comp			+		
							Vocabulary			+		
Allinder	1992		48		2 to 3		Math computation	CBM	Abs	–	4	0
			80				Spelling			–		
			47		4 to 5		Math computation			–		
			100				Spelling			–		
Entwistle	1992, 1994	MD	313	Mix	1 to 2	Black	Math concepts	CAT	Rel	–	16	0
					2 to 3					–		
					1 to 2	White				–		
					2 to 3					–		

Note. CTBS = Comprehensive Test of Basic Skills, LSMA = National Longitudinal Study of Mathematical Ability, MAT = Metropolitan Achievement Test, LWRT = Letter and Word Reading Test, WRAT = Wide Range Achievement Test, GMRT = Gates-MacGinitie Reading Test, CBM = Curriculum-Based Measurement, CAT = California Achievement Test.

Rel = relative, Abs = absolute.

No. *d*s = number of *d*-indexes that could be computed from a given report. No. DGLEs = number of DGLEs that could be computed from a given report.

A small study by McCormick and Mason (1981), conducted in a Canadian city, examined the effect of summer vacation on the reading scores of kindergartners. The students in the study were described as coming from lower-middle-class homes. Gains in raw score reading achievement were evident on each of nine subtests focusing primarily on letter identification.

As a master's project, Shaw (1982) compared the effects of summer vacation on regular education students and special education students who were placed either in regular or resource classrooms. Students were sampled from 28 schools throughout Stanislaus County, California. Overall, Shaw found a slight gain in both raw and grade-level equivalent reading recognition scores and a loss in total math scores over the summer months (effect sizes could be calculated for raw scores only). When analyzed by subgroup, the reading gain was confined to regular education students only; both groups of special education students had lower reading scores in fall than in spring. All three groups experienced loss in math achievement.

Johns and Vacca (1984) compared the spring and fall reading comprehension and vocabulary scores separately for 640 boys and girls in Grades 1 through 7. Thus, two test scores were compared for each of 14 independent samples of students. Overall, vocabulary scores improved while reading comprehension scores generally declined. Johns and Vacca claimed to have found a significantly greater gain in vocabulary for girls, though this conclusion seems hard to justify based on their data.

Wintre (1986) examined the effect of summer vacation on Canadian students in Grades 1, 3, and 5. Four sets of raw scores from subtests of the Metropolitan Achievement Test (MAT) were analyzed in a mixed-model analysis of variance that included as factors grade level, occasion (spring and fall), and subject area. All sources of variance were significant except for the three-way interaction. However, Wintre did not report standard deviations, and multiple-degree-of-freedom tests were not followed by single-degree-of-freedom contrasts. Therefore, it was not possible to calculate any effect sizes from this study. The main effect for summer vacation did indicate that students showed gains over the summer in raw achievement scores in word knowledge, reading, and math concepts but showed losses in math computation.

Allinder, Fuchs, Fuchs, and Hamlett (1992) gave a curriculum-based achievement test in math computation and spelling to separate samples of Canadian second and fourth graders. Examination of raw scores revealed summer loss for both content areas at both grade levels. Perhaps most noteworthy about this study was the size of the reported effects. It contained three of the five largest reported effects, including a summer loss in spelling in excess of one standard deviation. Of course, the two subject areas examined were those that the early studies indicated were most likely to be affected by summer vacation.

Finally, Entwistle and Alexander (1992, 1994) examined "setback" among an economically varied sample of Baltimore school children over two consecutive summers. These researchers divided their sample according to racial group (White and Black) and whether children attended segregated or integrated schools. Results for math concepts and reading comprehension were presented in two separate research reports. The analysis largely paralleled that of Heyns (1978) and replicated her earlier result. For math concepts, the authors reported a difference in achievement between Whites and Blacks that increased over time and was largely due to differences in summer change. Summer effects were also strongly related to the economic status of children's families. Lower-income children showed greater summer losses. Also, White children in segregated schools showed the greatest achievement gains, while Black children in segregated schools showed the least gain. For reading, children whose parents had dropped out of high school lost ground during the summer relative to students whose parents had not. Black children in integrated schools showed more reading gain over summer than Black children in segregated schools. For our meta-analysis, Entwistle and Alexander's study provided four effect sizes (one for math and one for reading for the summer between first and second grades, and one for math and one for reading for the summer between second and third grades) for each of four samples of students (race by type of school).

The Sustaining Effects Study

The Sustaining Effects Study (SES) was undertaken primarily to examine the effects of CE. As part of the effort, data were collected on as many as 120,000 students in a nationally representative sample of elementary schools for three successive school years, beginning in the fall of 1976 (Carter, 1984). Both spring and fall assessments of achievement were available for some students. Thus, the nature of the data collection permitted an appraisal of the impact of summer vacation on achievement. The SES deserves special attention here because of the size and nature of its sample and because of the disagreement that emerged concerning the interpretation of its results.

Ginsburg, Baker, Sweet, and Rosenthal (1981) were the first to use the SES data to examine the effects of summer vacation. They employed a nonrepresentative subsample of about 2,500 students in their analyses. Their purpose was to retest the results reported by Heyns (1978) using a sample drawn from a broader geographical area (Heyns's database included children from the Atlanta area only). The SES and Heyns's study differed in two other important ways. First, Heyns's conclusions were based on the word knowledge subtest of the MAT, whereas the SES used the total reading and total math scores from the CTBS. Second, whereas Heyns had used grade-level equivalent scores in her study, the SES used a metric called *vertical scale scores*. Vertical scale scores were developed for each child by using the projected national score distributions from the SES in the first year of the study. Otherwise, Ginsburg et al. attempted to use statistical analysis techniques that were identical to those of Heyns.

First, using partial correlations that controlled for prior achievement, Ginsburg et al. (1981) retested Heyns's finding that family income is more closely tied to student achievement during the summer than during the school year. The SES data did not replicate this finding; instead, they revealed partial correlations between income and achievement change that were half as large as Heyns's and equal for both spring and fall testings.

Second, using the ratio of average monthly achievement gains during the school year to average gains over the entire calendar year, Heyns found that achievement gains were larger during the school year. Also, larger ratios were found for Black and lower-income students, indicating that relatively more of their learning occurred during the school year. Heyns interpreted these findings to mean both that schooling affected achievement and that the effect of being in school was greatest for low-income students. Ginsburg et al. (1981) found an overall effect of schooling similar to that found by Heyns, but they did not replicate the relation of schooling to either race or income.

In subsequent analyses, the SES data did partially replicate one of Heyns's income- and race-related summer effects. Concerning simple changes in achievement levels from one testing to the next, the SES data revealed that "high income whites show[ed] a larger reading increase during the summer than [did] the other groups" (Ginsburg et al., 1981, p. 15), which did not differ from one another.

Finally, Ginsburg et al. (1981) retested several more complex regression models built by Heyns that included income and race. They concluded that although a few of the race and income coefficients were significant, it appeared that "any relationship between achievement change and socioeconomic status is, at best, tenuous" (p. 21). Ginsburg et al. summarized their findings by suggesting that the SES data revealed greater achievement gains during the school year than during summer, especially in math, but no relation between summer learning and a child's economic background.

In the same year, Klibanoff and Haggart (1981) came to a somewhat different set of conclusions based on a larger sample drawn from the SES data. Their analysis of over 39,000 student scores led them to this conclusion:

> With the exception of the declines in math [vertical scale scores] for [Grades] 3–4 and 4–5, the results . . . do not provide much support for the notion of an absolute loss over the summer. . . . It is also clear from the data, however, that the gains made during the summer by CE students are somewhat smaller than those made by non-CE students. . . . This relative loss is clearly more evident in reading than in math. (pp. xxiv–xxvi)

The average effect size estimate of summer change was approximately $d = +.08$ for reading and $d = +.02$ for math. For math, only fourth graders showed a gain.

Klibanoff and Haggart (1981) also reported no evidence that achievement changes over summer were related to initial achievement level. Evidence of summer loss was concentrated among higher achievers, where regression effects could be at work. With regard to both student income and race, Klibanoff and Haggart reported that students participating in free lunch programs and Black students tended to show smaller gains in reading than other students over the summer but that the opposite tended to be true for math. In both cases, however, the differences were described as not large. We could not estimate effect sizes for these different groups of students because the needed standard deviations were not reported.

As part of their critique of the work of Pelavin and David (1977, see above), Klibanoff and Haggart (1981) were especially skeptical of the use of grade-level equivalents as the metric for assessing summer loss. They claimed that grade-level equivalent scores tend "to overestimate the amount of gain made during the school year, caused by problems in interpolating fall norms from spring norms" (p. 10). Thus, what appeared to be a summer loss might actually have been an overestimation of spring scores. Klibanoff and Haggart, however, did acknowledge that grade-level equivalent scores had validity for comparing relative gain or loss for different groups of students over equal amounts of time.

Carter (1984) was next to interpret the SES. He did not present any reanalysis of data but instead summarized the various SES reports in a widely cited article that appeared in *Educational Researcher.* His language was somewhat stronger than that of the original authors:

> There are large reading gains over the summer. There are math gains and losses over the summer, and particularly in the higher grades there may be math losses. In comparing CE students with regular students, there may be a slight relative summer gain for CE students in reading, but not in math. The differences are so small that they have no practical significance. (p. 8)

Carter's penultimate statement appears to directly contradict the earlier quotation from Klibanoff and Haggart (1981).

Heyns (1987) took issue with both interpretations of the SES data. First, she pointed out that the summer vacation interval used in the SES actually contained 8 or 9 weeks of school instruction. This limitation was acknowledged by Klibanoff and Haggart (1981), who reported that spring testing occurred about 5 weeks prior to summer vacation, that fall testing occurred 3 weeks into the new school year, and that these dates would "err in favor of finding positive growth during the summer" (p. 6). We need to point out, however, that the summer interval used in the SES is not inconsistent with the interval (reported as May to October) employed by Heyns (1978) in her own study of summer effects.

With the SES overall estimate of summer effects thus suspect, Heyns (1987) next turned her attention to results comparing different groups over similar periods of time. First, she pointed out that the SES data unequivocally indicated a slower rate of learning during the summer period than during the school year. This was true in both reading and math and at every grade level.

Next, Heyns (1987) examined differential rates of achievement change in the SES data based on students' family income and race. Heyns did not disagree with Klibanoff and Haggarts's (1981) interpretation of the SES data but did make the case for evidence of differential rates of change more forcefully:

> For reading scores, the less advantaged group, whether defined by free lunch or minority status, gained the least during the summer months in all five cohorts. Three out of five cohorts declined in math scores during the summer; the relative declines in math, however, favored the less advantaged groups. (p. 1154)

Further, Heyns stated that "the entire racial gap in reading achievement is due to these 'small differences' in summer learning" (p. 1154).

Bryk and Raudenbush (1988) were next to analyze SES data. They did so in the context of a demonstration of the use of hierarchical linear models for statistical analysis. These models are meant to solve problems engendered by the fact that education data sets often measure student change over time while students are nested within classrooms, within schools, and within districts.

Bryk and Raudenbush (1988) worked with a small subsample of 618 students from 86 schools who were followed from spring of first grade to spring of third grade (five testings). The analyses focused on the relationships of both student poverty and school poverty concentration to both reading and math scores. Their results indicated significant drops in summer rates of learning for both math and reading. In fact, the summer rate of math learning was "virtually zero" (p. 95). Bryk and Raudenbush further claimed that due to the characteristics of the hierarchical linear model, their findings were "consistent with the contention that the observed summer drop-off is really an educational effect rather than a measurement artifact" (p. 95). To explain their finding, they suggested that reading and verbal skills are learned both in school and at home, whereas math learning occurs primarily in school.

Due to the size of their sample, Bryk and Raudenbush (1988) felt they could not reliably model the student poverty variable. However, the school poverty concentration variable indicated that summer drop-off rates in reading did not differ among schools, whereas mathematics drop-offs were smaller in high-poverty schools than in low-poverty schools. The authors attributed this finding to the existence of summer programs in high-poverty schools. However, they were missing data on the number of students in the sample who attended summer programs, so they could not model the summer school effect.

In many ways, the debate surrounding the SES data resembles an argument over whether an 8-ounce glass filled with 4 ounces of water is half empty or half full. We would suggest that regardless of the perspective of the viewer, the SES data reveal (a) no overall evidence of either absolute or relative summer loss, although the length of the summer interval seriously compromises the trustworthiness of this conclusion; (b) little or no gain over summer in math, and certainly less gain in math than in reading; and (c) slower rates of gain over summer for students from poorer or minority backgrounds.

For the purpose of meta-analysis, the SES contributed 10 effect sizes (calculated from the Klibanoff & Haggart, 1981, article) representing total math and reading summer effects for five independent samples of students in different grades. However, the SES is so large that its impact would swamp the rest of our data set. Therefore, we will report selected results with the SES both included and excluded, and with an adjusted weighting.

A Meta-Analysis of Recent Studies of Summer Effects on Achievement

Methods of Meta-Analysis

Metrics for expressing effect sizes. Two different metrics were used to estimate and describe the effects of summer vacation on achievement. First, for each sample we calculated a standardized mean difference, or *d*-index (Cohen, 1988). This was done by subtracting the sample's average achievement score in the spring from its average score in the fall and dividing this difference by the average of the two associated standard deviations. This metric permits expression of the change in achievement scores relative to the sample's own performance in the spring, regardless of the metric of the test itself. Thus, a *d* index of +.20 means that the average fall achievement score in the sample is two tenths of a standard deviation higher than the average spring score.

Also, when possible, we calculated the simple difference between a sample's average fall and spring grade-level equivalent scores, or what we will refer to as the *DGLE* (difference in grade-level equivalents). This metric expresses the change in achievement scores relative to national norms. Thus, a DGLE of +.20 means that the average fall achievement score in the sample increased by the equivalent of 2 months over summer break, when a 1-month increase was expected based on the calibration of the national norms.

Calculation of average effect sizes. Both weighted and unweighted procedures were used to calculate average effect sizes across samples. In the unweighted procedure, each independent effect size estimate was given equal weight in calculating the average effect. In the weighted procedure, each independent effect size was first multiplied by the inverse of its variance, and the sum of these products was then divided by the sum of the inverses. In this procedure, effect sizes based on larger samples were given greater weight. Because the effect sizes were based on change scores derived from a single sample, their variance was estimated by the formula $\sigma = (1 + d^2/4)/n$ (L. V. Hedges, personal communication, June 10, 1993).

Different sets of samples contributed to the calculation of average *d*-indexes and DGLEs. This occurred because the descriptions of some samples contained information that permitted the calculation of effect sizes using both metrics, while other samples permitted calculation with one metric but not the other. The final two columns in Table 3 indicate how many of each type of effect metric could be calculated from each research report.

One problem that arises in calculating average effect sizes is deciding what constitutes an independent estimate of effect. This meta-analysis used a shifting unit of analysis (H. Cooper, 1989). In this procedure, each effect size is first coded as if it were an independent event. So, for example, if a single sample description permitted calculation of separate *d*-indexes for the effects of summer vacation on achievement in math concepts and reading comprehension, the two effect sizes would be coded separately. However, for the estimate of summer vacation's overall effect on achievement, these two *d*-indexes would be averaged prior to entry into the analysis. In the weighted procedure, this single averaged effect would be weighted by the inverse of its variance. Thus, the sample would contribute only one effect size weighted proportionately to its sample size. However, in an analysis that examined the average effect of summer vacation on math and reading separately, this sample would contribute one effect estimate to each of the two calculations. Thus, the shifting unit approach retains as much data as possible while holding to a minimum any violations of the assumption that data points are independent.

Average *d*-indexes were tested for significance by calculating 95% confidence intervals for weighted estimates. If the confidence interval did not contain $d = .00$, the *d*-index was considered significant. DGLEs were not tested for significance.

Tests for moderators of effects. Possible moderators of summer vacation's effect on achievement were tested using homogeneity analysis (H. Cooper & Hedges, 1994; Hedges & Olkin, 1985). Homogeneity analysis compares the amount of variance exhibited by a set of effect sizes with the amount of variance expected if only sampling error is operating. The analysis can be carried out for groups of individual effect sizes or for group averages. Hedges and Olkin (1985) describe how the homogeneity analysis can be carried out using the general linear model program of the Statistical Analysis System (SAS Institute, 1985). The WEIGHTED statement is used to minimize the weighted residual sums of squares. In such an analysis, the sum of squares due to the tested moderator variable is treated as a chi-square statistic.

Study coding and intercoder reliability. Eighteen different characteristics of each effect size were coded. Most of these are described below. Each sample was independently coded by two coders, and any discrepancies were resolved in conference.

Characteristics of the Studied Samples

In all, 66 independent samples, described in 12 of the 13 research reports, contributed *d*-indexes and/or DGLEs. There were 11 research reports from which *d*-indexes could be obtained. These reports described 62 independent samples for which we could calculate a total of 118 separate *d*-index estimates of the effect of summer vacation on achievement test scores. Of these 62 samples, 5 contributed four *d*-indexes each to the data set, while the remaining samples each contributed either one or two *d*-indexes. For 43 samples contained in five reports, results were described in terms of DGLEs.

The total number of students in all the samples was 47,994. Of these, 38,384 students participated in the five samples contributed by the SES. A total of 9,610 students participated in the other 61 samples, with sample sizes ranging from 24 to 683 and averaging 153.

Of the 66 samples, 31 were described in reports that appeared in the 1970s, 27 appeared in the 1980s, and 8 appeared in the 1990s. Twenty-three of the samples were reported in four journal articles, 25 in two technical reports, 6 in two convention presentations, 8 in one book, and 4 in two master's theses or doctoral dissertations.

Students in 28 samples were described as coming from low-income families, and students in 20 samples were described as coming from middle-income families. Generally, this assessment was based on the community served by a participating school or on the percentage of students in a sample who were eligible for free or reduced-price lunch. Eleven samples were drawn from families of mixed income, and students' economic backgrounds were not specified in 7 samples.

Ten samples were described as containing only White students, and six as containing only Blacks; the rest contained either students of both races or students of unspecified race. Nine samples were all male, and 9 were all female. Thirty-one samples were drawn from urban areas, 4 from suburban areas, and 1 from rural areas; the remaining samples were drawn either from mixed or unspecified areas. Because so few samples were identified as anything but urban, the community location variable was dropped from further analysis.

The average number of days between the spring and fall testings, in the 34 samples for which this information was available, was 131, and the range of days was from 92 to 153. For 12 of the samples, the effect of summer vacation was measured by absolute measures of change. For 58 samples, relative measures of change were used, and for 4 samples both types of measures were reported.

In the overall analyses of *d*-indexes, each independent sample contributed a single effect size calculated to be the average of its separate effects. In the analyses that examined moderator variables that might influence the effect of summer vacation, each independent sample could contribute one average *d*-index to each category contained in a particular analysis. So, for example, if a sample had two *d*-indexes, one for a total math test and one for a total reading test, then this sample contributed one effect size to each subject area when subject area was examined as a moderator of the summer vacation effect (see H. M. Cooper, 1989, for a full explanation). DGLEs were calculated in a similar manner.

Overall Effect of Summer Vacation

The overall average effect sizes were calculated using both unweighted and weighted procedures. The data set containing the 62 independent samples providing *d*-indexes was examined first. This analysis estimated the unweighted overall effect of summer vacation to be $d = -.09$. Thus, the average student's fall score was one tenth of a standard deviation below where it had been in the spring. The corresponding unweighted DGLE (for all 43 contributing samples) was –.09, or a loss of about one month.

The weighted DGLE was comparable, at about –.13. However, when *d*-indexes were weighted by sample size, the overall average *d*-index equaled +.02, indicating a slight gain in achievement scores from spring to fall. The reason for the difference between the unweighted and weighted average *d*-indexes is the large size of the SES sample and the fact that its results were more positive than those of other studies. To illustrate the effect of the SES, a second data set excluding the SES sample was used to recalculate the overall average *d*-index. This data set produced an unweighted average overall *d*-index of –.10 and a weighted overall average *d*-index of –.13.

A homogeneity analysis of the data set including the SES revealed that there was considerably more variance in the *d*-indexes than would be expected from sample error alone, $\chi^2(61, n = 62) = 574.8, p < .0001$. When the SES was excluded from the data set, the homogeneity analysis was still significant, $\chi^2(56, n = 57) = 297.4, p < .0001$. Tables 4 and 5 display the overall average effect sizes.

Year of report appearance. The *d*-indexes were examined to determine whether their magnitudes were systematically associated with the years in which they were reported. This analysis indicated that in both data sets—the one including the SES and the one excluding it—the least negative effect of summer vacation was found in studies conducted in the 1980s, and the most negative effect was found in the 1990s. Whether or not the size of a *d*-index was linearly associated with the year in which the sample was collected depended on whether the SES samples were included in the analysis. With the SES included in the data set, more recent samples revealed less negative *d*-indexes, $\chi^2(1, n = 62) = 11.27, p < .001$; without the SES, however, no linear effect was shown, $\chi^2(1, n = 57) = 0.91$, n.s.

Effect of Summer Vacation on Separate Subject Areas

First, we examined whether the effects of summer vacation on achievement were different depending on whether a subject area was related to mathematics or reading and language. Figure 2 presents a stem-and-leaf display of the 74 *d*-indexes that went into this analysis. The average unweighted *d*-index for math-related subject areas was $d = -.14$. For reading- and language-related subject areas it was $d = -.05$.

The homogeneity analysis that included the SES samples revealed significantly greater summer gains for subject areas related to reading and language (weighted

TABLE 4
Effects of summer vacation on achievement in separate subject areas using the d*-index as the measure of change*

			Average d-index	
Subject area	No. of samples	Sample size	Unweighted	Weighted
All subjects				
With SES	62	46,421	−.09	+.02*
Without SES	57	8,037	−.10	−.13*
Math				
Computation	3	249	−.44	−.32*
Concepts	5	657	+.01	+.01
Applications	1	154	+.17	+.17*
Total with SES	15	40,584	−.13	.00
Total no SES	10	1,907	−.20	−.18*
Reading				
Comprehension	28	3,117	−.10	−.19*
Recognition	11	2,105	+.06	+.03
Vocabulary	14	640	+.14	+.12*
Total with SES	11	39,284	−.04	+.08*
Total no SES	6	671	−.14	−.15*
Spelling	2	180	−.53	−.41*

Note. Unweighted d-indexes were not tested for significance. Size of sample equals the total number of students across all independent samples.
*95% confidence interval does not contain d = .00.

d = +.05) than for math-related subject areas (weighted d = .00), $\chi^2(1, n = 74) = 58.58$, $p < .0001$. The reading- and language-related d-index was significantly different from d = .00. The homogeneity analysis that excluded the SES samples still found a significant difference between the two subject areas, $\chi^2(1, n = 64) = 4.49$, $p < .04$. However, now students showed significant losses in both subject areas, with greater losses evidenced in math (weighted d = −.16) than in reading and language (weighted d = −.11). Table 4 also displays the average d-indexes calculated for three math and four reading or language subject areas, as well as for total math and reading achievement scores. Because the SES reported only total math and reading scores, average d-indexes for these tests were calculated both including and excluding the SES.

The homogeneity analyses revealed that there was greater variance in average d-indexes across subject areas than would be expected by chance alone. This was true whether the sample included the SES, $\chi^2(8, n = 90) = 340.6$, $p < .0001$, or excluded the SES, $\chi^2(8, n = 80) = 160.2$, $p < .0001$. The average d-indexes revealed that summer vacation had a significant negative effect on students' achievement in math computation (d = −.32), total math without the SES in the analysis (d = −.18), reading comprehension (d = −.19), total reading without the SES included (d = −.15), and spelling (d = −.41). Summer vacation had a significant positive effect on math application (d = +.17), vocabulary (d = +.12), and total reading with the SES included (d = +.08). Table 5 contains the DGLEs that correspond to the d-indexes.

TABLE 5
Effects of summer vacation on achievement in separate subject areas using grade-level equivalent scores as the measure of change

Subject area	No. of samples	Sample size	Average DGLE	
			Unweighted	Weighted
All subjects				
Without SES	43	6,364	−.09	−.13
Math				
Computation	1	154	−.22	−.22
Concepts	1	154	+.59	+.59
Applications	1	154	+.56	+.56
Total no SES	7	1,458	−.16	−.08
Reading				
Comprehension	24	2,620	−.20	−.32
Recognition	8	1,645	+.06	−.07
Vocabulary	14	640	+.18	+.17
Total no SES	7	1,802	−.09	−.12
Spelling	0	0		

Note. DGLEs (differences in grade-level equivalent scores) are expressed in years. Differences were not tested for significance, but corresponding *d*-indexes were. Size of sample equals the total number of students across all independent samples.

Other Influences on the Effect of Summer Vacation

In order to test other factors that might moderate the effect of summer vacation, we created an edited data set. First, the impact of the SES on the overall results left us with the dilemma of how to treat this study in the moderator analyses. Specifically, it is preferable to use effect sizes weighted by sample size in these analyses, but weighting the SES *d*-indexes by their full sample sizes (ranging from 7,190 to 8,412) would leave all the other studies with little influence on the analyses. Therefore, in the edited data set we included the SES *d*-indexes, but we set their weights to equal 1,366, or twice the size of the next largest sample. Also, in the edited data set we excluded the three *d*-indexes for math computation and the two *d*-indexes for spelling. These five *d*-indexes were all negative and contained three of the four largest effect sizes. The average overall *d*-index for the edited data set was $d = -.04$.

For purposes of clarity of interpretation, each moderator was tested for its influence on the effect of summer vacation separately for (a) absolute and relative measures of change crossed with (b) math and reading subject areas. Within math and reading subject areas, no further specification of content was made. Thus, each moderator was tested on four separate edited data sets.

Type of report. Homogeneity analyses comparing the average *d*-indexes reported in published sources (journals and books) to those reported in unpublished sources (technical reports, papers presented at meetings, theses, and dissertations) found no significant differences in any of the four data sets.

Length of testing interval. The length of the interval between the spring and fall testings was not linearly associated with the magnitude of *d*-indexes based on absolute measures of change. However, the homogeneity analyses revealed that

Mathematics		Reading and Language
	+.3	5
	+.2	0124457
2	+.1	2246779
9500	+.0	4666889
882210	-.0	11222444557
7754	-.1	122224499
955	-.2	3357
3	-.3	5
0	-.4	156
5	-.5	6
0	-.6	
	-1.0	4

FIGURE 2. *Stem-and-leaf display of* d*-indexes for effect of summer vacation on achievement test scores in (a) math and (b) reading and language*

longer summer intervals were associated with greater gains or lesser losses in relative-metric achievement test scores for both math, $\chi^2(1, n = 6) = 6.31, p < .02$, and reading $\chi^2(1, n = 29) = 14.40, p < .001$.

Family income. There were no samples of middle-class students for whom the absolute change over summer vacation was tested on math achievement. The homogeneity analysis for relative changes in math achievement revealed no difference in average d-indexes.

Table 6 presents the average effects of summer vacation on reading achievement for different income groups and different metrics for measuring change. There was a significant difference between average d-indexes for middle- and low-income students when absolute measures of change were employed, $\chi^2(1, n = 4) = 4.50, p < .04$. Middle-class children showed a significantly greater gain in reading achievement over summer ($d = +.23$) than did lower-income students ($d = +.07$). Both the middle- and lower-income student gains were significantly different from $d = .00$.

The relative measures of reading achievement also revealed a significant difference between middle- and lower-income students, $\chi^2(1, n = 38) = 57.78, p < .001$.

TABLE 6
Average effect of summer vacation on reading achievement for different income groups and different metrics of change

Metric		Income level: Middle	Income level: Low
Absolute	d-index	+.23*	+.07*
	No. of samples	2	2
	Size of sample	225	907
Relative	d-index	+.06	−.21*
	No. of samples	18	19
	Size of sample	990	3,888
Relative	DGLE	+.16	−.19
	No. of samples	20	19
	Size of sample	1,591	4,477

Note. DGLE = difference in grade-level equivalent scores between fall and spring. DGLEs were not tested for significance. Average d-indexes are weighted by sample size.
*95% confidence interval does not contain $d = .00$.

On this measure, lower-income students showed an average loss in reading achievement over summer ($d = -.21$), while middle-income students showed an average gain ($d = +.06$). The loss by lower-income students was significantly different from $d = .00$, but the middle-income student gain fell just short of significance. The DGLEs, also reported in Table 6, revealed a difference between the two income groups of about 3.5 months.

Because there were a sufficiently large number of samples that tested for summer effects on reading using relative measures of achievement, we decided to examine individually the influence of student income level on measures of reading comprehension and recognition. The mean d-indexes and DGLEs underlying these analyses are presented in Tables 7 and 8. The analyses revealed a significant difference in the effect of summer break on the reading comprehension test scores

TABLE 7
Average d-*index for effect of summer vacation on reading comprehension and recognition for different income groups using relative metrics of change*

Reading area		Income level: Middle	Income level: Low
Comprehension	d-index	−.14*	−.27*
	No. of samples	14	10
	Size of sample	640	1,980
Recognition	d-index	+.13*	−.12*
	No. of samples	4	4
	Size of sample	350	1,295

Note. Average d-indexes are weighted by sample size.
*95% confidence interval does not contain $d = .00$.

TABLE 8
Average DGLEs for effect of summer vacation on reading comprehension and recognition for different income groups using relative metrics of change

		Income level	
Reading area		Middle	Low
Comprehension	Average DGLE	−.27*	−.34*
	No. of samples	14	10
	Size of sample	640	1,980
Recognition	Average DGLE	+.23*	−.15*
	No. of samples	4	10
	Size of sample	350	1,980

Note. DGLE = difference in grade-level equivalent scores. Average DGLEs are weighted by sample size.

of students based on their income level, $\chi^2(1, n = 24) = 19.09, p < .001$. Summer vacation had a significantly greater negative effect on the reading comprehension of students coming from low-income families ($d = -.27$) than on students coming from middle-income families ($d = -.14$). The 95% confidence intervals for students at both income levels did not contain $d = .00$, which indicates that the effect of summer was negative for both groups of students. DGLEs indicated that low-income students lost about 0.7 months more than did middle-income students.

A significant difference was also found for income level on the effect of summer vacation on reading recognition scores, $\chi^2(1, n = 8) = 16.74, p < .001$. Low-income students showed a significant loss in reading recognition over summer ($d = -.12$), while middle-income students showed a significant gain ($d = +.13$). DGLEs indicated that middle-income students gained about 2.3 months in reading recognition over summer, while lower-income students lost about 1.5 months. Comparisons could not be made for the other two reading-related areas because only low-income students had total reading scores and only middle-income students had vocabulary test scores.

Student gender and race. Homogeneity analyses that compared all-male samples to all-female ones, and all-White samples to all-Black ones, found no systematic relation of average d-indexes to these student individual differences.

Grade level. The grade level of students revealed no relation to the effect of summer vacation on math achievement test scores. The grade level of students showed a linear trend in relation to absolute changes in reading achievement, $\chi^2(1, n = 5) = 3.25, p < .07$, but this finding was based on a comparison of one sample of kindergartners to four samples of fifth graders. Kindergartners had higher absolute reading achievement in the fall than in the spring, while fifth graders had lower absolute reading achievement in the fall.

More trustworthy was the finding of a significant relation of grade level to relative changes in reading achievement, $\chi^2(1, n = 62) = 64.64, p < .001$. Table 9 presents the average effects of summer vacation on relative reading achievement for different grade levels. The table indicates that as grade level goes up, the effect of summer vacation changes from positive to negative and grows more detrimental.

TABLE 9
Average effect of summer vacation on relative reading achievement for different grade levels

	d-index			DGLE		
Grade	No. of samples	Sample size	*M*	No. of samples	Sample size	*M*
1	7	1,967	+.04	4	764	+.06
2	8	2,189	+.04	2	74	+.14
3	6	2,169	–.02	5	1,203	–.12
4	8	2,778	–.12*	6	1,237	–.34
5	11	4,056	–.09*	10	2,690	–.20
6	10	1,212	–.12*	10	1,212	–.18
7	9	642	–.17*	9	642	–.36
8	3	235	–.21*	3	235	–.46

Discussion

The results of the meta-analysis indicate that when the overall effect of summer vacation on standardized test scores is at issue, students appear at best to demonstrate no academic growth over summer. At worst, students appear to lose 1 month of grade-level equivalent skills relative to national norms. When performance change is gauged relative to the student's own fall scores, the worst-case scenario seems to be that the average student score in the fall is about one tenth of a standard deviation below the spring average.

There is also evidence to suggest that these estimates of the effect of summer vacation are conservative, or optimistic. This conclusion is based on the finding that the average number of days in the spring-to-fall testing interval was 131, equivalent to the number of days in the months of June, July, August, and September, plus the first 10 days of October. While the length of summer vacations varies from school district to district, it seems reasonable to assume that the typical study of summer effects has included at least 5 weeks of instructional time. Further, the meta-analysis found that as the length of the summer interval increased, the amount of loss in test scores decreased. Most likely, the longer summer intervals covary more with the inclusion of more instructional time rather than longer summer breaks. Thus, greater amounts of instructional time in summer intervals probably serve to mitigate the estimated negative impact of the summer break. Therefore, the effect of summer vacation would likely be more detrimental, perhaps dramatically so, if it were measured from the day school is dismissed to the day students return.

In fact, the regression output that accompanies a homogeneity analysis can be used to estimate the effects of summer vacations of varying lengths. For example, using the edited data set in an unweighted regression analysis, we found the estimated overall effect of summer vacation on achievement to be $d = -.03$. This estimate is associated with a summer break lasting 131 days. If we use the intercept and regression coefficient associated with this analysis to estimate the effect of a 102-day summer break—that is, one lasting from the Friday prior to Memorial Day until the Tuesday following Labor Day—we find the estimated *d*-

index for this interval to be $d = -.07$.

In addition, the best-case scenario, suggesting that students demonstrate no loss but also no growth over summer, is based on the weighted estimate of the vacation's impact including the Sustaining Effects Study (SES). The SES sample contained four times as many participants as the rest of the evidential base combined and revealed generally positive changes in academic skills over summer. Thus, the SES is solely responsible for the most optimistic assessment. The question then arises, Why not just believe the results of the SES? First, the SES included a summer interval that was above average in length, lasting 140 days, and included 8 weeks of instruction. Therefore, it could be argued that if the SES results were adjusted to account for its relatively long summer interval, its findings would be consistent with the more pessimistic result of the remaining body of evidence. Second, and perhaps most revealing, is the rationale that the SES researchers gave for the decision to include 5 weeks prior to summer vacation and 3 weeks after school began within their summer interval. Klibanoff and Haggart (1981) wrote, "Aside from allowing this time for students to 'settle down,' many teachers agree that significant achievement gains are unlikely to be observed during these periods" (p. 5). The researchers did not discuss how they reconciled this perception on the part of teachers, and the SES's apparent legitimation of it in the choice of testing dates, with the study's finding of gain over summer.

Effects of Summer Vacation on Separate Subject Areas

The overall results of the meta-analysis mask dramatic differences in the effect of summer vacation on different skill areas. First, the results indicate that summer loss was more dramatic for math-related subject areas than for reading or language. Murname (1975) provided a possible explanation for this differential effect when he suggested that reading and language skills are learned both at home and in school, while mathematics learning may be more restricted to formal school settings. Put differently, children's home and community environments may provide more opportunity to practice reading skills and to learn new words than to practice and learn mathematics. If this is so, we would expect the differential effect of the long vacation on the two skill areas.

Second, the skill areas of math computation and spelling showed strikingly larger effects than any other subject areas. The strong negative impact of summer break on these subject areas was evident not only in terms of relative losses but in terms of absolute losses, as well. In addition, the losses were apparent even in the review of studies conducted prior to 1975, where only crude measures of effect could be obtained.

A possible explanation for this differential effect is that both math computation and spelling skills involve the acquisition of factual and procedural knowledge, whereas other skill areas, especially math concepts and problem solving and reading comprehension, are more conceptually based. Cognitive psychologists suggest that factual and procedural learning requires extensive practice, while conceptual understanding requires a lot of experience but not necessarily practice (G. Cooper & Sweller, 1987; Geary, 1995). Thus, the relative lack of opportunity to practice computation and spelling over summer vacation may mean that these facts and procedural skills are most susceptible to decay.

The two explanations for the differential effects of summer break on different subject matters rely on differences in opportunities to practice and the susceptibility of certain knowledge to be dissipated without practice. From these principles we can generate some predictions about how other subject areas, not tested in recent investigations, might be affected by summer break. For example, we could speculate that over summer students will tend to forget science facts but retain knowledge of scientific concepts. Foreign language vocabulary can be expected to be lost. Relatedly, the constellation of meta-analytic findings lends credence to the argument that summer break would be especially detrimental to students who speak a language at home that is different from the language of school instruction.

Student Individual Differences

Intelligence. In addition to the influence of subject area, researchers have examined numerous student individual differences as possible moderators of the effect of summer vacation on achievement. Among the earliest and most frequently studied influences has been the students' intelligence test scores. Overall, these studies revealed little evidence to suggest that student intelligence has an impact on the effect of summer break. Roughly equal numbers of studies showed positive, negative, and no relationships between the variables. However, these investigations sampled students whose intelligence fell within the normal range. Therefore, the conclusion that no relation exists should not be generalized to students with abnormally high or low IQ scores or to students with disabilities in particular learning domains.

Only one study categorized students based on whether or not they were eligible for special education (Shaw, 1982). This study examined absolute changes in total math and reading recognition scores. For math, all students experienced a summer loss, but special education students in special classes showed the largest loss. For reading, regular education students experienced a gain over summer, while special education students, whether attending special classes or resource rooms, showed a summer loss. Thus, the results of the meta-analysis should not be construed as providing evidence that counters arguments that children with learning disabilities or low intelligence have a special need for extended-year schooling (cf. Sargent & Fidler, 1987).

Family income. As interest in intelligence as a moderator of summer effects diminished in recent years, interest in the family income of students became the central focus of research. Supplementing the literature that looked at income level directly is a large group of studies that reported the income background of the students who were sampled, even though income was not an object of study. In the meta-analysis, the results of these studies could be used to help estimate summer's effects for different income groups.

The meta-analysis revealed no differential effect of summer on the mathematics skills of middle- and lower-class students. All students lost math skills over summer. However, substantial differences were found for reading and language. The results indicated that middle-class children showed significantly greater absolute summer gains in reading and language achievement than lower-income students. Middle-class students showed a nonsignificant gain in grade-level equivalent reading scores, while lower-class students showed a significant loss. On average, summer vacations created a gap of about 3 months between middle- and lower-class students. For specific reading areas, comprehension scores for both income groups declined over summer, but declined more for lower-class students. Reading recognition scores showed a significant gain for middle-class students and a significant loss for lower-class students.

The results of the meta-analysis support Heyns's (1978) and Entwistle and Alexander's (1992, 1994) contention that socioeconomic inequities are heightened by summer break. These researchers explained the effect of income by suggesting that low socioeconomic status translated directly into fewer learning opportunities and/or less support for learning-related activities during the summer vacation.

However, Entwistle and Alexander (1992) contended that the learning inequity revealed itself in the area of math skills. The meta-analysis found that the effect of income was actually manifested in reading, not math. The differential summer effect for math and reading can be understood if we assume that across homes, income levels are roughly equally deficient in opportunities to practice and learn math, as suggested by the general results concerning loss of math skills. Middle- and lower-income homes, on the other hand, might differ in their access to reading materials and language learning opportunities.

It would also be interesting to examine the findings of reading gains by middle-class children over summer in light of classic learning and memory theories regarding spontaneous recovery and reminiscence. *Spontaneous recovery* refers to the fact that "an extinguished response will, with rest, recover some of its strength" (Houston, 1991, p. 78). *Reminiscence* has been defined as "the opposite of forgetting, that is, as an improvement in the memory of a target over time" (Glass & Holyoak, 1986). Interestingly, reminiscence has been demonstrated with children but not with adults, and evidence for both spontaneous recovery and reminiscence is confined largely to verbal tasks. One explanation for reminiscence is that children adopt a high criterion for recognizing a target during immediate recall but a more lenient criterion after delay. However, there are problems in applying these notions to the present findings. First, the proposed reminiscence mechanism ought to work for math as well as reading. Second, the time interval used in most classic spontaneous recovery and reminiscence studies is 1 or 2 days, not 3 or 4 months. And, finally, the learning literature does not suggest why reminiscence might occur for some children but not for others.

Gender and race. Neither student gender nor student race appeared to have a consistent moderating influence on the effect of summer vacation. While several authors examined gender differences in summer effects, none provided a theoretical rationale for why this individual difference might be important, so perhaps the lack of effect is not surprising.

When race has been examined, the argument has typically been made that race might be an important variable because of its association with income disparities. We found no researcher who predicted differential effects of summer break for different racial groups based on biological or cultural differences that would influence students' learning capacities or summer experiences. The economic interpretation appears to be borne out by the findings of the meta-analysis. In most of the included studies, when race was examined as a moderator of summer effects, income level was at least partially controlled by drawing student samples from similar economic backgrounds. Under these conditions, no race differences were found.

Grade. Finally, the meta-analysis revealed a linear influence of grade level on the effect of summer vacation. On average, first and second graders showed nonsignificant gains in achievement over summer relative to national norms, while students in fourth grade and beyond showed significant losses, some of which were quite dramatic. This finding is somewhat counterintuitive in that it seems reasonable to believe that factors that influence children's earliest experiences in school have the greatest impact on their learning.

One possible explanation for the meta-analytic finding focuses on an issue in measurement. Specifically, student scores in early grades may show less variance from national norms simply because there is a restriction in their possible lower range of values. This is called a *floor effect* in scaling. For example, a child in first grade can, by definition, score only one grade below the normed grade level. Thus, early grades have a constraint on the amount of negative change the child can manifest. Finally, it is important to point out that even though summer effects might be smallest in earlier grades, it is not unreasonable to argue that these differences are especially critical, as they may set in motion processes that influence all later learning.

Implications for Educational Policy

We began this article by casting the debate about the effects of summer vacation in the context of proposed changes in the school calendar. What, then, are the implications of our findings for the calendar debate?

First, educational policymakers could simply choose to live with the diminished learning opportunities and decay in skills that accompany the present dominant school calendar. They could accept the argument that, rather than increase the amount of time students spend in school, it makes more sense, both from economic and pedagogical viewpoints, to improve the way we use the time students already spend in school (Karweit, 1985). However, this review suggests that policymakers who take this position must also address whether or not schools and society are obligated to remedy the inequities in learning opportunities, and the consequent differences in achievement, that summer vacation creates for children from different economic backgrounds. They must also state why an optimal pedagogical strategy ought not include both an alternative calendar and more efficient use of time.

Second, this research synthesis might be used to help direct the efforts of summer enrichment and remedial programs. Specifically, the meta-analytic results suggest that when options for programs are limited, a primary focus on mathematics instruction in summer would seem to be needed most. Alternatively, if programs have the explicit purpose of mitigating inequities across income groups, then a focus on summer reading instruction for lower-income students would seem to be most beneficial.

Finally, the results of this synthesis can be used to argue for adopting changes in the school calendar. That is, these results legitimize the contention that calendar changes could make educational sense, in addition to better fitting the lifestyle of current American families. Such changes might also address issues related to economic inequities that affect children's life chances. However, proponents of calendar change cannot take the findings about summer vacation to mean that any alternative calendar is preferable to the present one. For instance, the present synthesis does not assess whether alternative schedule calendars, such as those that include the present number of school days but distribute shorter and more frequent vacations throughout the year, are actually more effective than the present calendar. This review also does not estimate how much, if any, of the summer loss could be recouped by adding days to the school year.

Directions for Future Research

There are several questions that the present body of evidence on summer vacation leaves unanswered. These questions can provide direction for future research. First, there is a need for more research to be conducted at both early grades and in high school. No study has examined the effect of summer break on students beyond the eighth grade. It would be especially informative to have investigations of early grades that examine student income level as a moderator of summer effects.

Also with regard to income level, it would be of interest to know how students from families with different income levels spend their summer vacations. We suspect that such studies, both surveys and ethnographies, would provide important insights into the etiology of the differences in summer loss between middle- and lower-class students.

There is also a need for studies that estimate the "pure" effect of summer vacation by employing test dates that more accurately capture the vacation interval. As part of these studies, researchers should vary the date of the fall testing so as to estimate the time needed to recoup any summer loss.

The added precision in the testing interval should be accompanied by increased precision in the distinction between subject matters. Many recent studies have used total math and reading scores from standardized achievement tests or have focused on single subtests that correlate highly with other subtests. However, the existence of high correlations among students' scores on different subtests does not imply equal effects of summer across subject areas. This review's finding that summer has dramatically different effects on different subject areas dictates that researchers maintain the subject matter distinction in future investigations.

And finally, researchers can begin to examine the broader question of how much time can pass between lessons and tests before there is an overall negative influence on the instructional program. When the issue is viewed from this perspective, it becomes clear that certain other breaks from instruction—including weekends, spring break, and winter break—might be considered. We suspect there is much less research on these intervals. Also, it is not clear that they give rise to the debates concerning equity, special education, and lifestyle changes associated with the summer break.

Conclusions

Our review of 39 studies indicated that achievement test scores decline over summer vacation. The loss equals about one month on a grade-level equivalent scale or one tenth of a standard deviation relative to spring test scores. The effect of summer break is more detrimental for math than for reading, and most detrimental for math computation and spelling. There was also evidence that the summer break has roughly equal negative effects on the math skills of students from middle- and lower-income families, but greater negative effects on the reading skills of lower-income students. In fact, middle-class students appeared to gain on grade-equivalent reading recognition tests over summer, while lower-class students lost on them. There were no moderating effects for student gender or race, but the negative effect of summer did increase with increases in students' grade levels.

The finding that subject area influenced the amount of summer loss may be due to the differential availability of opportunities to practice different academic material over summer (with reading practice more available than math practice) and to differences in the material's susceptibility to memory decay (with factual and procedural knowledge more easily forgotten than conceptual knowledge). The income differences also may be related to differences in opportunities to practice and learn (with more books and reading opportunities available to middle-class students).

The results suggest that summer school programs might best focus on instruction in mathematics if they are targeted at all students, or on reading instruction if their purpose is to mitigate differential summer effects associated with economic inequities. Proponents of school calendar changes appear correct in arguing that summer vacation has a negative impact on learning and that this impact is not equal across all students. Hopefully, such knowledge will guide our policy priorities and choices as American education is restructured to help our children meet the challenges of the 21st century.

References

Allinder, R. M., Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1992). Effects of summer break on math and spelling performance as a function of grade level. *The Elementary School Journal, 92,* 451–460.

Arnold, R. D. (1968, April). *Retention in reading of disadvantaged Mexican-American children during the summer months.* Paper presented at the meeting of the International Reading Association, Boston, MA.

Association of California School Administrators. (1988). *A primer on year-round education.* Sacramento, CA: Author.

Ballinger, C. E., Kirschenbaum, N., & Poimbeauf, R. P. (1987). *The year-round school: Where learning never stops.* Bloomington, IN: Phi Delta Kappa.

Barrett, M. J. (1990). The case for more school days. *Atlantic Monthly, 266,* 78–106.

Beggs, D. L., & Hieronymus, A. N. (1968). Uniformity of growth in the basic skills throughout the school year and during the summer. *Journal of Educational Measurement, 5,* 91–97.

Bender, J. F. (1944). Retention of experience by school children. *Proceedings of the Oklahoma Academy of Science, 25,* 59–61.

Botwin, J. R. (1965). *Summer and school year correlates of academic achievement.* Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, OH.

Brueckner, L. J., & Distad, H. W. (1924). The effect of the summer vacation on the reading ability of first-grade children. *The Elementary School Journal, 24,* 698–707.

Bruene, E. (1928). Effect of the summer vacation on the achievement of pupils in the fourth, fifth, and sixth grades. *Journal of Educational Research, 18,* 309–314.

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

Carter, L. F. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher, 13*(7), 4–13.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Cook, R. C. (1942). Vacation retention of fundamentals by primary-grade pupils. *The Elementary School Journal, 43,* 214–219.
Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Education Psychology, 79,* 347–362.
Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.
Cooper, H. (1989). *Integrating research: A guide for literature reviews.* Newbury Park, CA: Sage.
Davis, R. A., & Rood, E. J. (1947). Remembering and forgetting arithmetical abilities. *The Journal of Educational Psychology, 38,* 216–222.
Elam, S. M., Rose, L. C., & Gallop, A. M. (1996). The 28th annual Phi Delta Kappa/Gallop poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 78,* 41–59.
Elder, H. E. (1927). The effect of the summer vacation on silent-reading ability in the intermediate grades. *The Elementary School Journal, 27,* 541–546.
Entwistle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57,* 72–84.
Entwistle, D. R., & Alexander, K. L. (1994). Winter setback: The racial composition of schools and learning to read. *American Sociological Review, 59,* 446–460.
Fitzsimmons, M. E. (1969). *Analytical study of the retention of selected skills.* Unpublished doctoral dissertation, Fordham University, Bronx, NY.
Gandara, P., & Fish, J. (1994). Year-round schooling as an avenue to major structural reform. *Educational Evaluation and Policy Analysis, 16,* 67–85.
Garfinkel, M. A. (1919). The effect of the summer vacation on ability in the fundamentals of arithmetic. *The Journal of Educational Psychology, 10,* 44–48.
Gastright, J. F. (1979, September). *Summer loss in reading achievement: A comparison of fall to spring reading achievement gains with fall to fall reading achievement gains.* Paper presented at the annual meeting of the Evaluation Network, Cincinnati, OH.
Geary, D. C. (1995). Reflections of evolution and culture in children's cognition. *American Psychologist, 50,* 24–37.
Ginsburg, A., Baker, K., Sweet, D., & Rosenthal, A. (1981, April). *Summer learning and the effects of schooling: A replication of Heyns.* Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA.
Glass, A. L., & Holyoak, K. J. (1986). *Cognition.* New York: Random House.
Grenier, M. A. (1975). An investigation of summer mathematics achievement loss and the fall recovery time (Doctoral dissertation, University of Georgia, 1975). *Dissertation Abstracts International, 36,* 3069A.
Hawn, H. C., Ellett, C. D., & Des Jardines, L. (1981, March). *Differences in mathematics achievement between males and females in Grades 1–3.* Paper presented at the annual meeting of the Eastern Educational Research Association, Philadelphia, PA.
Hayes, D. P., & Grether, J. (1969, April). *The school year and vacations: When do students learn?* Paper presented at the meeting of the Eastern Sociological Association, New York.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.
Heyns, B. (1978). *Summer learning and the effects of schooling.* New York: Academic

Press.
Heyns, B. (1987). Schooling and cognitive development: Is there as season for learning. *Child Development, 58,* 1151–1160.
Houston, J. P. (1991). *Fundamentals of learning and memory.* San Diego, CA: Harcourt Brace Jovanovich.
Irmina, S. M. (1928). The effects of summer vacation upon the retention of the elementary school subjects. *Educational Research Bulletins, 3,* 99.
Jamar, I. (1994). *Fall testing: Are some students differentially disadvantaged?* Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.
Johns, J. L., & Vacca, R. T. (1984). An inquiry into summer loss in reading: Grades 1–7. *Reading Horizons, 24,* 208–215.
Karweit, N. (1985). Should we lengthen the school year? *Educational Researcher, 14*(6), 9–15.
Katsiyannis, A. (1991). Extended school year policies: An established necessity. *Remedial and Special Education, 12,* 24–28.
Keys, N., & Lawson, J. V. (1937). Summer versus winter gains in school achievement. *School and Society, 46,* 541–544.
Klibanoff, L. S., & Haggart, S. A. (1981). *Summer growth and the effectiveness of summer school* (Rep. No. 8). Mountain View, CA: RMC Research Corporation.
Kolberg, O. W. (1934). A study of summer-time forgetting. *The Elementary School Journal, 35,* 281–287.
Kramer, G. A. (1927). Do children forget during the vacation? *Baltimore Bulletin of Education, 6,* 56–60.
Lahey, M. F. (1941). Permanence of retention of first-year algebra. *The Journal of Educational Psychology, 32,* 401–413.
Mazzarella, J. A. (1984). Longer day, longer year: Will they make a difference? *Principal, 63,* 14–20.
McCormick, C., & Mason, J. M. (1981). What happens to kindergarten children's knowledge about reading after a summer vacation? *The Reading Teacher, 35,* 164–172.
Merino, B. J. (1983). The impact of year-round schooling: A review. *Urban Education, 18,* 298–316.
Morgan, L. D. (1929). How effective is specific training in preventing loss due to the summer vacation? *The Journal of Educational Psychology, 20,* 466–471.
Mousley, W. (1973). Testing the 'summer learning loss' argument. *Phi Delta Kappan, 54,* 705.
Murname, R. (1975). *The impact of school resources on inner-city school children.* Cambridge, MA: Ballenger.
National Education Commission on Time and Learning. (1994). *Prisoners of time.* Washington, DC: U.S. Government Printing Office.
National Institute of Education. (1978). *Compensatory education study.* Washington, DC: U.S. Department of Health, Education, and Welfare.
Nelson, M. J. (1928). How much time is required in the fall for pupils of the elementary school to reach again the spring level of achievement? *Journal of Educational Research, 18,* 305–308.
Noonan, M. E. (1926). Influence of the summer vacation on the abilities of fifth and sixth grade children. *Contributions to Education, 204,* 103.
Parsley, K. M., & Powell, M. (1962). Achievement gains or losses during the academic year and over the summer vacation period: A study of trends in achievement by sex and grade level among students of average intelligence. *Genetic Psychology Monographs, 66,* 285–342.
Patterson, M. V., & Rensselaer, N. Y. (1925). The effect of the summer vacation on

children's mental ability and on their retention of arithmetic and reading. *Education, 46,* 222–228.

Pelavin, S. H., & David, J. L. (1977). *Evaluating long-term achievement: An analysis of longitudinal data from compensatory education programs* (SRI Project No. 4537). Washington, DC: Educational Policy Research Center.

Sardo-Brown, D., & Rooney, M. (1992). The vote on all-year schools. *American School Board Journal, 179,* 25–27.

Sargent, L. R., & Fidler, D. A. (1987). Extended school year programs: In support of the concept. *Education and Training in Mental Retardation, 22,* 3–9.

SAS Institute. (1985). *SAS user's guide: Statistics* (Version 5). Cary, NC: Author.

Schrepel, M., & Laslett, H. R. (1936). On the loss of knowledge by junior high-school pupils over the summer vacation. *The Journal of Educational Psychology, 27,* 299–303.

Scott, L. F. (1967). Summer loss in modern and traditional elementary school mathematics programs. *California Journal of Educational Research, 18,* 145–151.

Shaw, V. T. (1982). *Retention of selected reading and arithmetic skills by learning disabled pupils and non-disabled pupils over summer vacation.* Unpublished master's thesis, California State College, Stanislaus.

Soar, R. S., & Soar, R. M. (1969). Pupil subject matter growth during summer vacation. *Educational Leadership, 2,* 577–587.

White, W. F. (1906). Reviews before and after vacation. *American Education, 10,* 185–188.

Wintre, M. G. (1986). Challenging the assumption of generalized academic losses over summer. *Journal of Educational Research, 79,* 308–312.

Authors

HARRIS COOPER is Professor, Department of Psychology, McAlester Hall, University of Missouri, Columbia MO 65211; psychhc@showme.missouri.edu. He specializes in social psychology and research methodology.

BARBARA NYE is Director, Center of Excellence for Research and Policy on Basic Skills, Tennessee State University, 330 10th Ave. North, P.O. Box 141, Nashville, TN 37203-3401. She specializes in program evaluation and comprehensive family services.

KELLY CHARLTON is a graduate student, Department of Psychology, McAlester Hall, University of Missouri, Columbia, MO 65211. She specializes in social psychology.

JAMES LINDSAY is a graduate student, Department of Psychology, McAlester Hall, University of Missouri, Columbia, MO 65211. He specializes in social psychology.

SCOTT GREATHOUSE is a graduate student, Department of Psychology, McAlester Hall, University of Missouri, Columbia, MO 65211. He specializes in social psychology.

Received September 21, 1995
Revision received April 25, 1996
Accepted June 20, 1996