Are Schools the Problem? The Effects of School on Learning and Obesity

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

by

Paul T. von Hippel
Graduate Program in Sociology

The Ohio State University
2010

Dissertation Committee:
Douglas B. Downey, Advisor
Claudia Buchmann
Liana C. Sayer
Abstract

This dissertation is part of a larger research agenda whose purpose, broadly speaking, is to identify the major sources of inequality and of secular change in children’s physical and intellectual growth. In a theoretical chapter I review the evidence for two guiding principles—namely that, compared to the rest of children’s lives, schools, at least in recent decades, have been (1) relatively stable and (2) relatively fair. In a methodological chapter I review the logic of seasonal research, which has contributed to the case for schools’ relative fairness by showing that disadvantaged children fall behind primarily during the summer vacation, when school is out.

In a chapter on year-round schools, I bolster the case that summer setback is a symptom of disadvantages in children’s non-school environments, which cannot be eliminated simply by rearranging the school calendar. In a chapter on obesity, I use seasonal analysis to show that, during the 20-year rise in child obesity, it was the summer vacation, more than the school year, that grew more fattening. These finding implicate children’s non-school environments as the primary source of both the child obesity epidemic and of inequality in academic achievement.

In my conclusion, I review my findings as part of the broader case that schools are relatively stable and relatively fair. I then discuss the implications for sociological research and for social policy.
Acknowledgments

I thank Doug Downey for encouraging me to think big, and Claudia Buchmann for reining in my excesses. Both contributions were essential to helping me develop an argument that makes challenging claims but still has the potential to convince a skeptic. I thank Liana Sayer for helping me to see the implications beyond the school system and for helping me with time use in every sense. I thank Cassie for being an easy baby, mostly, and Pam and Ann for handling her when she wasn’t. Without all of your contributions, this dissertation could not have gotten off the ground.
Vita

1990..........................................................B.A., Music, Yale University

1998..........................................................Ph.D., Music, Stanford University

2001 ...........................................................M.A.S., Statistics, Ohio State University

Publications


*James Coleman Award for best article in the past 2 years, Education section, American Sociological Association.

†The first two authors contributed equally.

*von Hippel, P.T. “Regression with missing Ys: An improved strategy for analyzing multiply-imputed data” Sociological Methodology 37, 83-117. 2007

*Clifford Clogg Award for best paper by a graduate student, Methodology Section, American Sociological Association.


*Willard Waller Award for best article in the past 3 years, Education Section, American Sociological Association.


1. “Critical value.”
2. “Difference of proportions.”
3. “Expected value.”
4. “Normalization.”


Fields of Study

Major Field: Sociology
# Table of Contents

Abstract ........................................................................................................................... ii  
Vita .................................................................................................................................. iv  
Publications .................................................................................................................... iv  
Fields of Study ............................................................................................................... vi  
Table of Contents .......................................................................................................... vii  
List of Tables .................................................................................................................. xi  
List of Figures ................................................................................................................. xii

Chapter 1 Schools, Inequality, and Change........................................................................ 1  
Schools and Inequality .................................................................................................... 2  
  Historic Reductions in School Inequality .................................................................... 3  
  Sources of Inequality, Inside and Outside Schools ..................................................... 7  
  Unmeasured Family Influences .................................................................................. 9  
  Seasonal Longitudinal Data ....................................................................................... 11  
  A Place for School Effects ......................................................................................... 15  
  Schools and Inequality: Resolving the Paradox ......................................................... 22  
  A New Framework for Thinking About Inequality ..................................................... 26  
Schools and Continuity ................................................................................................. 29  
Looking Forward ............................................................................................................ 31
<table>
<thead>
<tr>
<th>Chapter 2 Seasonal Research Design</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal Data and Counterfactuals</td>
<td>34</td>
</tr>
<tr>
<td>Seasonal Data and the Crossover Design</td>
<td>37</td>
</tr>
<tr>
<td>Seasonal Measurement Schedules</td>
<td>42</td>
</tr>
<tr>
<td>Hierarchical Linear Growth Models</td>
<td>44</td>
</tr>
<tr>
<td>Three-Level Model</td>
<td>45</td>
</tr>
<tr>
<td>Two-Level Model</td>
<td>50</td>
</tr>
<tr>
<td>Contrasts and “Impact”</td>
<td>53</td>
</tr>
<tr>
<td>Missing values</td>
<td>57</td>
</tr>
<tr>
<td>Chapter 3 What Happens to Seasonal Learning Patterns in Year-Round Schools?</td>
<td>60</td>
</tr>
<tr>
<td>Summer Learning</td>
<td>63</td>
</tr>
<tr>
<td>Year-Round Schools</td>
<td>65</td>
</tr>
<tr>
<td>Preview of Results</td>
<td>70</td>
</tr>
<tr>
<td>Past Research on Year-Round Schools</td>
<td>72</td>
</tr>
<tr>
<td>Our Contribution</td>
<td>74</td>
</tr>
<tr>
<td>What kinds of schools use year-round calendars?</td>
<td>75</td>
</tr>
<tr>
<td>Data</td>
<td>75</td>
</tr>
<tr>
<td>Comparison</td>
<td>77</td>
</tr>
<tr>
<td>Learning Rates in Nine-Month and Year-Round Schools</td>
<td>84</td>
</tr>
</tbody>
</table>
List of Tables

Table 3.1. School calendars. ................................................................. 77
Table 3.2. School characteristics. ......................................................... 78
Table 3.3. Child characteristics. ......................................................... 78
Table 3.4. Test dates and test scores. Means. ......................................... 87
Table 3.5. Seasonal learning rates in nine-month and year-round schools. 92
Table 3.6. Adjusted learning rates in nine-month and year-round schools. 96
Table 4.1. Participation in school breakfast and lunch programs. ............ 128
Table 4.2. Samples and design. ......................................................... 141
Table 4.3. Average BMI growth in two cohorts of schoolchildren. ........... 144
Table 4.4. Average growth rates obese and other children in the BSS, 1987-1991. 152
Table 4.5. School and child-level variances and correlations from the ECLS-K, 1998-2000. ................................................................. 153
Table 4.6. BMI and BMI growth by ethnicity, in two cohorts of schoolchildren. 157
Table 4.7. Fuller models of BMI growth in two cohorts of schoolchildren. ....... 161
List of Figures

Figure 2.1. Typical measurement schedule for seasonal research................................. 43

Figure 2.2. When learning rates are extrapolated beyond the test date, estimates of
summer learning are reduced............................................................................................ 44

Figure 3.1. Average learning in nine-month and year-round schools............................ 93

Figure 3.2. Average learning in year-round and nine-month schools, adjusted for other
child- and school-level differences................................................................................ 97

Figure 4.1. BMI for age. ................................................................................................. 116

Figure 4.2. Monthly growth rates at different percentiles of the BMI distribution. ....... 118

Figure 4.3. Average school-year and summer BMI growth in the ECLS-K and the BSS.
......................................................................................................................................... 145

Figure 4.4. Average school-year and summer BMI growth for obese and normal-weight
children. .......................................................................................................................... 154

Figure 4.5. The effect of ethnicity on BMI and BMI growth in two cohorts of
schoolchildren.................................................................................................................. 158
In this opening chapter, I make the case for two principles that shape much of my research agenda, both in this dissertation and elsewhere.

1. The first principle is that, compared to the rest of children’s lives, U.S. schools today are relatively fair.
2. The second principle is that, again compared to the rest of children’s lives, U.S. schools, at least in recent history, have been relatively slow to change.

I state these principles for two reasons. First, they are important, and yet they are not developed as explicitly in the literature as I would like. The case that schools are comparatively fair is rarely made, especially by sociologists (for exceptions, see Alexander 1997; Downey, von Hippel, and Broh 2004), and to my knowledge no one has drawn together all the streams of evidence that I draw together here. The case that schools are slow to change seems so open and shut that few scholars bother to make it, and for that reason the causes and effects of schools’ slowness are not well understood.

The second reason that I state these principles is that they have the qualities of what I take to be useful theory. Although the principles are not especially formal, they are simple and clear, and, most important, they can be used to make predictions.
The predictions that flow from these principles are straightforward. The fairness principle implies that, if a social outcome is distributed very unequally, schools probably make only a small contribution to the inequality. The slowness principle implies that, if a social change is very rapid, schools probably make only a small contribution to the change.

The subject of child obesity, which I analyze in Chapter 4, is in the crosshairs formed by these two principles. Child obesity is a problem that has emerged rapidly, and a problem that affects some types of children much more than others. In light of schools’ relative stasis and uniformity, we would predict that schools have played only a minor role in the emergence of child obesity. The results of our analysis are consistent with this prediction, suggesting that schools have caused at most 20% of the rise in child obesity, and have increased obesity more among the affluent than among the poor.

In the remainder of this chapter, I review the evidence for both principles, taking care to acknowledge apparently contrary evidence and put it in context. I begin with the case that schools are relatively fair, then move on to the case that schools are relatively slow to change. I close with the implications for sociology and for policy.

_Schools and Inequality_

_“The initial state in which schools find children, and the continuing environments outside the school that compete for the child’s time, are unequal... [T]he school’s task is—besides increasing opportunity for all, through what it imparts—to reduce the unequalizing impact on adult life of these differential environments.”_ – James S. Coleman (1975)

For most of the twentieth century, the U.S. school system was widely regarded as systematically unfair, a major source of social inequality and a major hindrance to social
mobility. Today, however, my reading of the evidence suggests that the U.S. school system is remarkably fair, especially when compared to other aspects of U.S. society.

Two things have changed. First, the U.S. school system has grown fairer. Second, an accumulation of research evidence has attributed more and more of the inequality in educational outcomes to causes that lie outside of schools’ control.

*Historic Reductions in School Inequality*

One hundred years ago, around 1910, there could be no doubt that the school system was an important source of inequality. Children in poor communities went to poor schools, because four-fifths of school funding came from local residents through property taxes (Augenblick, Myers, and Amy Berk Anderson 1997), and incomes were at least as unequal as they are today (Piketty and Saez 2003). Schools in the South were poorer than those in the North, and in rural areas, many children still attended one-room ungraded common schools (Goldin 1999).

The most glaring examples of educational inequity were the South’s segregated black schools. In most Southern states, black teachers were paid less than half the salary of white teachers, and the black school year was ten to eighty days shorter than the white school year (Margo 1990). In some states, the student-teacher ratio was twice as high in black schools as in white schools (Margo 1990; Card and Krueger 1992). Black elementary schools often met in improvised quarters, such as a church or a shack, and the whole South had only sixty-four black public high schools, few of them in the rural areas where most black families lived (Thomas Jesse Jones 1917). These inequities were not accidental. They had been deliberately cultivated by white supremacist state and local
governments, starting around 1890, with the openly avowed goal of preventing blacks from achieving social equality (Margo 1990; Woodward 1957).

But 1910 was a low point. Over the following decades, racial inequities shrank considerably (Margo 1990; Card and Krueger 1992; Wilson 1980; Woodward 1957). Blacks began to leave the South for factory work in the North, and this put pressure on Southern governments to improve school conditions for the blacks who remained (Margo 1990). The NAACP, after its founding in 1909, carried out public relations and lobbying campaigns to raise awareness and change attitudes toward segregation. The NAACP also mounted a series of court challenges against the inequities of segregated education, and finally, in the 1954 *Brown* case, overturned the legal basis for school segregation itself, although in practice Southern schools remained segregated for more than a decade longer (Tushnet 1987). By the time national school conditions were surveyed in the 1966 Coleman report on *Equality of Educational Opportunity* (Coleman et al. 1966), many observers, including Coleman himself (Grant 1973), were surprised to learn that by conventional measures the resources of black and white schools were not so far from equality. For example, in elementary schools the teachers of white and black students had identical average salaries, and the average class size was 32 for black students and 29 for white students. In the rural South the differences were a bit larger, but in the urban South, some of the resource discrepancies actually favored black students, just slightly (Coleman et al. 1966). In short, by conventional measures the resource discrepancies between black and white schools were much smaller than they had been 30 or 50 years before. The primary remaining difference was in schools’ student composition, which still justified the labeling of entire schools as black and white. Nationally, two-thirds of black students
and four-fifths of white students—and, despite the *Brown* decision, nearly all of black and white students in the South—went to schools where at least 90% of their classmates shared their own race (Coleman et al. 1966).

In the South, black and white students did not remain separate for much longer. Although segregation had changed little in the twelve years after the *Brown* decision, after the Coleman report, and in part because of it (Grant 1973), desegregation proceeded rapidly. By 1970 Southern schools were the most integrated in the country: 33% of Southern black students attended majority white schools, up from 14% in 1967 and 2% in 1964; by 1988 the percentage had risen to 43%. With black and white students often attending the same schools, discrepancies in their access to school resources shrank even further. Even in South Carolina, where white students’ classes were half as large as black students’ classes in 1910, and still 10% smaller at the time of the Coleman report in 1966, white and black students’ class sizes had converged to absolute equality by 1990.

Some of the ground gained during desegregation has been lost. Major court decisions from 1991 to 2007 permitted resegregation, so that by 2005 the percentage of Southern black students attending majority white schools had slipped to 27%—no lower than levels in the Midwest, but a level not seen in the South since 1969 (Orfield and Chumgmei Lee 2007). Desegregation has had little impact in regions outside the South, and next to no effect on Hispanics (Orfield and Chumgmei Lee 2004). Yet today it remains safe to say that black and white students are far more integrated, and perforce more equal in their school resources, than they were at the time of the Coleman report.

It is not just racial disparities in school resources that have shrunk. Economic disparities have shrunk as well. Today schools serving poor children have substantial
access to funds from outside of their immediate community. Until 1935, four-fifths of the funding for public schools came from local property taxes, and the one-fifth that came from state governments was often distributed in a way that favored wealthy districts (Augenblick et al. 1997). But the equitability of state funding began to improve in the 1920s, and the level of state funding jumped in the 1930s and again in the 1970s. Beginning in the 1970s, a series of court challenges forced various state governments to redistribute funding in a way that explicitly leveled funding disparities by compensating for inequalities in local resources (Augenblick et al. 1997; Hoxby 2001; Card and Krueger 1996; Corcoran et al. 2004).

Federal funds compensate for local disparities as well. Since 1965 Title I of the Elementary and Secondary Education Act (ESEA) has directed federal funds toward schools serving poor children, and the 2001 reauthorization of ESEA, known as No Child Left Behind, explicitly holds schools responsible for reducing test-score gaps associated with race and poverty.

Today local property taxes provide less than half of school funding. State governments provide 48%, and the federal government provides an additional 9% (NEA Research 2010). Federal and state funding has not wholly equalized school resources, partly because local resources have been growing more disparate along with the general increase in income inequality (Card and Abigail Payne 2002). In addition, there is some evidence that poor communities tend to reduce local funding when federal and state funding flow in (Card and Abigail Payne 2002). Nevertheless, state and federal funds do provide a kind of resource floor for the poorest communities, and if a poor district is able and willing to supplement outside funds with a manageable level of local funding, its
schools can have much better resources than was possible under older funding models. On the whole, financial inequality between school districts is 20-30% smaller today than it was in 1972 (Corcoran et al. 2004).

It is not just school resources that have become more equal. Students’ achievement has become more equal as well. On the National Assessment of Educational Progress, the black-white and Hispanic-white gaps in both reading and math skills have shrunk by a third since 1973, although practically all of the shrinkage occurred between 1973 and 1986. Total inequality, not just the inequality between ethnic groups, also appears to have shrunk. In mathematics, the “interdecile range” between the 10th and 90th percentiles is 15% smaller today than it was in 1978. In reading, since 1971 the interdecile range has held constant for 13- and 17-year-olds, but has shrunk by 15% for 9-year-olds.

On the whole, while inequities certainly remain, the U.S. school system is substantially more equal today than it was 50 years ago, and 50 years ago it was dramatically more equal than it was 100 years ago. The reduction in inequality since 1960 is all the more remarkable because it happened during a period when inequalities in household income and wealth were growing to levels approximating those of 1910. In relative terms, compared to the rest of U.S. society, the school system may now be as fair as it has ever been.

Sources of Inequality, Inside and Outside Schools

In the previous section, we documented the progressive equalization of school resources and opportunities. We also documented the gradual shrinkage of achievement gaps between advantaged and disadvantaged children. It is easy to imagine that the
equalization of resources caused some of the shrinkage in achievement gaps—and in fact it may have (Alexander 1997; Card and Abigail Payne 2002). But the relationship between school resources and student achievement is not as strong as many reformers imagine.

Modern research on the relationship between resources and achievement began with the Coleman report on *Equality of Educational Opportunity* (Coleman et al. 1966), an analysis of the test scores and background characteristics of over 40,000 teachers and 570,000 students at the beginning of 1st, 3rd, 6th, 9th, and 12th grade, as well as school characteristics reported by about 4,000 principals. As we mentioned earlier, one of the surprising findings of the Coleman report was that the resource gaps between black and white schools were small. Further analysis of the Coleman data found that resource gaps between schools serving poor and affluent children were small as well (Jencks et al. 1972).

An equally surprising finding of the Coleman report was that schools and their characteristics accounted for little of the variation in test scores. At the beginning of first grade, before schools had a chance to matter, there were already substantial gaps between racial and economic groups. These gaps were larger for 3rd-9th graders, who had been in school longer, but for children of all ages most of the variation in test scores lay within rather than between schools. Conventional measures of school quality—such as teachers’ salaries, experience, and education, class size, and physical facilities—had little or no association with student test scores. What really seemed to matter for test scores were the characteristics of the students and their families—such as race, parental education, and the family’s economic position. These characteristics mattered not just for individual
students but for entire student bodies. For example, the score of a black or poor student tended to be lower if most of his or her classmates were also black, poor, or low-scoring.

Commentary on the Coleman report pointed out several weaknesses in the research design (e.g., Armor 1972). Two weaknesses are probably the most important. First, some measures of school quality were omitted or crudely measured. Second, the data represented a cross-sectional snapshot of student achievement at a single point in time, rather than a longitudinal portrait that might show why achievement gaps grew from 1st grade to 3rd, 6th, 9th, and beyond. More recent research, using longitudinal data with better measures of school inputs, does in fact suggest that variation in some of those inputs can affect achievement growth. We will discuss some of this school research a little later.

Unmeasured Family Influences

What critics of the Coleman report rarely acknowledged was that the limitations of Coleman’s research design were not limited to the schools’ side of the inequality story. Some of the child and family measures were crudely measured as well, and the lack of longitudinal information obscured family as well as school effects on achievement growth. In fact, some of the most powerful research to emerge since the Coleman report shows that families, or more generally non-school variables, matter in ways that go far beyond what was acknowledged in the Coleman study.

The Coleman report measured family background crudely. The researchers could not interview the parents and had to ask children as young as first graders for information. Even if the children were accurate—and young children often are not (Kerckhoff, Mason, and Poss 1973)—the information requested of them was often
coarse. In an attempt to gauge economic position, for example, children were asked whether the family owned a car, but they were not asked whether the car was a new Cadillac or an old VW Bug. Later research has demonstrated that more of the variation in test scores—including much of the variation associated with race—can be explained by asking parents’ directly about family income or, even better, family wealth (e.g., Orr 2003).

Probably the most neglected family variable in education research is genetics. At the time of the Coleman report, the evidence for genetic influences on human abilities and behavior was limited, and the most visible work on the topic was an ugly debate about the possibility of a genetic relationship between race, class, and IQ (Jensen 1969)—a debate that repeated a generation later (Herrnstein and Charles Murray 1994, chapter 13). Starting around 1980, however, a remarkable body of genetic evidence began to accumulate. Perhaps the most startling research was the Minnesota Study of Twins Reared Apart, which showed in a large survey what had only been demonstrated in small studies before: that identical twins are almost as psychologically similar if they are separated soon after birth as if they are raised in the same home (Bouchard et al. 1990). Although we understand very little about the mechanisms by which genes shape psychology, evidence from the Minnesota study and other sources suggests that genes (some mediated by the environment) explain about half of the variation of many psychological characteristics—not just IQ, but occupational interests as well as personality traits such as achievement orientation and social potency (Bouchard et al. 1990). These traits are of course relevant to success in school and beyond, and later research has confirmed that genetics plays a very large role in shaping educational
achievement and attainment (Nielsen 2006; Lee Anne Thompson, Detterman, and Plomin 1991; Miller, Mulvey, and Nick Martin 2001). In a study of twins and other siblings during high school, Nielsen (2006) estimated that 60% of the variance in high school GPA, verbal IQ, and college plans is explained by genetic factors, while less than 10% is explained by aspects of the environment that are shared by siblings, which includes almost every part of the home environment that has ever been measured in a survey. It may be, in fact, that many of the home resources and parental behavior that appear to influence children’s academic achievement may in part be proxies for genetic characteristics that are shared between parents and offspring (Nielsen 2006).

The importance of genes does not mean that we are consigned at conception to some genetic caste. Genes influence parts of our psychology, but they do not fully determine our fate, and there is evidence that non-genetic influences matter more at the low end of the socioeconomic spectrum than at the high end (Turkheimer et al. 2003). What genetic evidence means about education is that family influences shape achievement in ways that go far beyond the enriching experiences that affluent children have at home, and far beyond any in-school advantages that affluent parents can secure for their children. Even if schools were absolutely uniform, even if children were assigned to new parents at birth, there would still be a tendency for economically successful parents to give birth to academically successful children. We should not assume that this tendency means that schools are unfair.

*Seasonal Longitudinal Data*

After the release of the Coleman report, it was suggested that longitudinal data might reveal ways in which school resources can impact achievement and achievement
gaps (Armor 1972). And to some extent this has happened. But to a greater extent longitudinal data has demonstrated the degree to which achievement gaps are shaped by families.

The longitudinal data that reveal the most about schools and families are data collected on a seasonal schedule. If children are measured twice each year—once when school begins in the fall, and once when school ends in the spring—we can compare how quickly inequality grows during different seasons: during the fall, winter, and spring, when school is in session, and during the summer, when school is out. School-year growth is influenced by a mix of school and non-school factors that are hard to disentangle, but summer growth is influenced by non-school factors alone (Heyns 1978).

The earliest seasonal analysis was carried out shortly after the Coleman report, using data from over 600 New York City elementary schools in 1965-1967. The researchers found that the gap in reading skill between affluent white schools and poor minority schools grew at a much faster rate during summer vacation than during the academic year (Hayes and Grether 1969, 1983). This basic pattern has since been replicated among elementary students in New Haven (Murnane 1975), Atlanta (Heyns 1978), and Baltimore (Entwisle and Alexander 1992), and in two national samples (Downey et al. 2004; Klibanoff and Haggart 1981). The pattern has also been replicated abroad, in Belgium (Verachtert et al. 2009) and Holland (Luyten, Schildkamp, and Verachtert 2009), although replication apparently failed in Sweden, perhaps because socioeconomic differences are smaller there (Lindahl 2001).

The finding that achievement gaps grow fastest in summer suggests that those gaps are due primarily to inequalities outside of school (Hayes and Grether 1983; Heyns
The amount that summer vacation contributes to the socioeconomic achievement gap varies from study to study, but it is always substantial. In New York City in the 1960s, summer learning accounted for 80% of the growth in the achievement gap between 2nd and 6th grade (Hayes and Grether 1983). In Baltimore in the 1980s, summer learning accounted for 100%—that is, all—of the growth in the achievement gap between 1st grade and 9th grade (Alexander, Entwisle, and Olson 2007). In Chapter 3 I find evidence that various ethnic and socioeconomic gaps actually shrink during kindergarten and then grow during summer, so that the summer accounts for more than 100% of the 12-month gap in achievement growth. Perhaps the smallest estimate of summer’s contribution is Benson and Borman’s (2010) report that early in children’s schooling, the summer accounts for about half of the growth in the socioeconomic achievement gap. But even this contribution is remarkably large. Since the summer vacation covers only a fifth of the calendar year, on a month-for-month basis the summer is clearly contributing more than its share of inequality.

It may be tempting to say that the summer gap represents the contribution of non-school inequalities to the achievement gap, while the school-year gap represents the contribution of school-based inequalities. But this surely understates the non-school contribution. Many of the non-school inequalities that are evident during summer continue to matter during the school year (during weekends, at the very least) so that some share of the school-year learning gap is due to non-school influences as well. This means that if the summer accounts for half of the annual growth in the achievement gap (as in Benson and Borman 2010), well over half of the annual growth in the gap must be due to non-school inequalities. Or, to choose a less conservative estimate, if the summer
accounts for 80% of the annual growth in the achievement gap (Hayes and Grether 1983), it is conceivable that 90% or more of the growth in the gap is due to non-school inequalities.

It is even possible, as we suggested earlier, that non-school inequalities account for more than 100% of the achievement gap, so that school actually serves to reduce inequality (cf. Coleman 1975). To make this argument, you do not have to find that the achievement gap shrinks during the school year, as I find in Chapter 3. It is only necessary to find, as nearly all seasonal studies do, that inequality grows more slowly during the school year than during the summer vacation. From a counterfactual perspective, the reduction in learning gaps during the school year suggests that schools reduce the growth in inequality that would occur in their absence. For example, in Downey et al.’s (2004) analysis of learning in kindergarten, summer, and first grade, the authors point out that, on a certain reading test, the gap between a typical high- and low-SES child grew an annual rate of 1.54 points, from 6.90 points at the end of kindergarten to 8.44 points at the end of first grade. However, if these children had diverged all year at the rate observed during the summer, over twelve months the gap would have grown by 5.47 points, to 11.37 points by the end of first grade. With school in the mix, the gap grew by 1.54 points; without school, it would have grown by 5.47 points. It appears that school reduced the growth in the reading gap by a factor of 3.55 (5.47/1.54). In other words, non-school inequalities accounted for 355% of the 12-month growth in the achievement gap, and school reduced that growth by 255%.

Perhaps the strongest evidence that schools reduce inequality comes from the section of Downey et al.’s (2004) study that analyzed “total inequality” in seasonal
learning rates. Other seasonal studies have focused on the gap between socioeconomically advantaged and disadvantaged children, but the definition of socioeconomic disadvantage varies from one study to another, and it can always be argued (e.g., Cook 1996; Benson and Borman 2010) that the definition used in a particular study is inadequate. Downey et al. (2004) sidestepped these definitional issues by looking at total variation in achievement and achievement growth, the vast majority of which cannot be attributed to any conventional measure of socioeconomic status.

Analysis of the total variance in achievement and achievement growth only strengthened the conclusions of seasonal research. First, during the school year the variation in learning rates was smaller than it was during the summer. Second, children who were behind at the beginning of kindergarten tended to fall further behind during summer vacation, but tended to catch up, or at least fall behind more slowly, when school was in session. Both results corroborate the view that schools reduce inequality in learning rates.

A Place for School Effects

In short, improvements in research since the Coleman report have actually increased the role of families in explaining achievement inequality. Better measures of family variables, especially genetic variation, have explained more of the variation in achievement. And seasonal studies have shown that inequality grows fastest when school is out of session.

None of this means that schools do not matter. In fact, seasonal research suggests a new way to think about the importance of schools. Instead of focusing so much on schools’ relative effects—the difference between one school and another, or between the
experiences that two children have in the same school—seasonal results suggest that we should think more about the “absolute effect of schooling”—the difference between attending school and not attending at all (Luyten 2006). Seasonal results are just one source of evidence that the absolute effect of schooling is substantial. Other evidence comes from examining the effects of school strikes (William E. Caldwell and Jeffreys 1983; Belot and Webbink 2010), the effect of changes to school calendars (Pischke 2007; see Chapter 3), and the effect of delaying a child’s entry because his or her birthday falls after an arbitrary cutoff (Cahan and Nora Cohen 1989; Luyten 2006). (See chapter 2 for further discussion.)

Seasonal research also confirms that there do exist relative differences between one school and another. Some schools are substantially better than others at increasing the rate at which students learn when they finish summer vacation and start the school year (Downey, von Hippel, and Hughes 2008). Remarkably, the schools that increase learning most are almost as common in disadvantaged poor and minority neighborhoods as they are in neighborhoods populated by affluent whites (Downey et al. 2008). The surprisingly uniform distribution of high-impact schools constitutes further evidence that the modern school system is relatively fair.

Further evidence for relative school effects comes from the story, some of which we have already told, of segregated schools in the South. In 1910, black schools had dramatically fewer resources than white schools. These resource differences accounted in part for blacks’ much lower rates of literacy and possibly for some of their disadvantage in earnings (Margo 1990; Card and Krueger 1996). As the resource gap between black and white schools shrank between 1910 and 1950, so did the literacy and earnings gap
between black and white adults (Margo 1990; Card and Krueger 1996). By 1966, when
the Coleman study was carried out, black and white schools had reached near parity. But
had the Coleman study been carried out 30-50 years earlier, it would have found larger
resource differences between black and white schools, and larger consequences of those
differences. In fact, one of those consequences was visible in Coleman’s finding that
black parents had lower educational attainment and earnings than white parents. These
parental deficits, one of the family effects that depressed the achievement of black
children, were in part a consequence of the abysmal resources of the black Southern
schools that many black parents had attended a generation earlier.¹

Even in our era of comparatively equal resources, longitudinal research has
demonstrated that differences in school resources can affect learning, although many of
these effects are smaller than was once thought, and small compared to total inequality.
We will briefly review the evidence on four of the best-studied educational resources:
teacher expectations, curriculum tracking, school composition, and teacher skill.

Teachers expect some students to learn more than others, and there is a concern
that these expectations may become self-fulfilling. The classic Pygmalion classroom
experiment, carried out around the same time as the Coleman study, showed that teachers
could be fooled into expecting that certain students were poised for growth, when in fact
the designated students had been chosen at random. Although teachers’ induced
expectations had no basis in students’ actual potential, students who were labeled with
high expectations did in fact tend to outlearn their peers over the subsequent school year
(Rosenthal and Jacobson 1968). Survey research confirms that teacher expectations are

¹ The transformation of one generation’s school effects into the next generation’s family effects is known as
“intergenerational drag” (Margo 1990).
strongly associated with student achievement (e.g., Roscigno 1998), and at first this seems to suggest that teacher expectations may explain a substantial part of inequality in achievement.

The difficulty with this conclusion is that, unless teachers are participating in an experiment, their expectations are not random and do have a strong basis in reality. Teachers quite naturally expect high achievement from children who are already high-achieving and well-prepared. Teacher expectations for their students tend to be accurate, and the better teachers know their students, the more accurate their expectations are (Raudenbush 1984). Even when teachers rely on a stereotypical expectation that poor or minority children will underperform, the strength of that expectation is consistent with those groups’ average achievement levels (Jussim and Harber 2005). The main reason that teacher expectations are associated with achievement is not that expectations affect achievement, but that achievement affects expectations. Net of this reverse effect, there is some evidence that expectations affect achievement, but the causal effect is much smaller than the raw association: about a tenth of a standard deviation on average, and perhaps twice that for stigmatized types of students (Jussim and Harber 2005).

Tracking and ability grouping is another topic that has been extensively studied. In high schools and middle schools, it is common practice to assign students who appear more academically promising into higher course tracks where the materials are more challenging, the expectations are higher, and the teachers are often better qualified (Gamoran 2010; Oakes 1985). Although tracking has been widely criticized since the 1980s, and many schools have stopped using the label (Hallinan 2004), analysis of high school transcripts suggests that the practice remains widespread in all but name (Lucas
Ability grouping is a kind of precursor to tracking, in which children in kindergarten or first grade are assigned to work with a small group of peers who have similar skills and often similar social backgrounds as well (Rist 1970; Condron 2008).

Tracking and ability grouping are a form of institutionalized expectation, and the problems encountered in studying them are similar to the problems in studying expectations. While the association between achievement and group or track assignment is strong, much of the association comes from the effect of achievement on track or group assignment, rather than the effect of track or group on achievement. Students who are assigned to a high track tend to have high prior achievement and tend to expend more effort on their schoolwork, and in some studies that control for these differences, the effect of track on achievement disappears completely (Carbonaro 2005). A broader review of careful longitudinal research suggests that tracking and perhaps ability grouping do have some effect on achievement inequality (Gamoran 2010), but the effect is much smaller than the simple association between track and achievement would suggest.

Teacher skill is also a topic that has received considerable attention. Longitudinal research suggests that teachers differ substantially in effectiveness; within schools, the difference between one teacher’s class and another’s can account for something like 10% of the student-level variation in that year’s gains (Clotfelter, Ladd, and Vigdor 2006; Rockoff 2004). Although this is a decent-sized effect, it should be emphasized the effect of a teacher is limited to that year’s achievement gains. Total achievement is a sum of many years’ gains, including time spent out of school during summer vacation and the first five years of life (Alexander et al. 2007). So the effect of even an outstanding fifth
grade teacher is likely to be small compared to the inequality that has opened up by the end of fifth grade (e.g., Decker, Mayer, and Glazerman 2004). This is why longitudinal data is necessary to see that teachers really matter. It should also be pointed out that little of the variation in teacher effectiveness is reliably associated with readily observed teacher characteristics. Characteristics that do explain a small part of the variation include teacher experience, especially the experience gained during the first two or three years (Rockoff 2004), and the selectiveness of the college that the teacher attended (Clotfelter, Ladd, and Vigdor 2007).

In addition to the effects of teachers and curriculum, it is important to consider the effect of a school’s student body. To return to the Coleman study, the only school-level effect that seemed at all strong in the Coleman data was the composition of the school’s student body. Net of a student’s individual characteristics, the achievement of an individual student was associated with the characteristics of other students in his or her school. For example, children who were black or poor tended to do better in schools where other children were white, affluent, or high-scoring.

The student body effect, also known as the effect of student peers, student composition, or student mix, was one of the most widely publicized results of the Coleman report, and became an important argument for accelerating school integration. In the Coleman data, the effect of student body characteristics was immense, accounting in a recent reanalysis for 50% of the between-school variation in achievement (Borman and Maritza Dowling 2010). Later estimates of the student body effect have been smaller, often much smaller, and there is an important reason for this. What appears to be a student body effect may in large part be a proxy for unmeasured characteristics of the
individual student (Hanushek et al. 2001). The possibility is easy to illustrate using the Coleman data, because some of the Coleman measures of family characteristics were quite coarse. As we mentioned earlier, one of the Coleman measures of affluence was a question asking a student whether his or her family had a car. The answer explained some of the variation in student achievement, but surely much more would be explained if we knew whether the car was a new Cadillac or an old VW Bug. The Coleman data do not tell us what kind of car the child’s family had, but the characteristics of other children in the school provide a good clue. If only a few of the children in a school have a family car, it is unlikely that any of those cars is a Cadillac. On the other hand, if all of the other children have a family car, a Cadillac is not out of the question. In this way, information pooled across a student’s peers can proxy for unmeasured detail about the student’s own situation, and this can make peer effects look much larger than they really are. The problem of overestimated peer effects gets smaller when family variables are measured more carefully, but given the difficulty of capturing even a small fraction of family effects in a survey, it is hard to eliminate the problem of spurious student body effects. A recent fixed-effect analysis suggested that peer poverty does not affect individual test scores at all, although peer test scores do (Hanushek et al. 2001). In short, the true effect of student body composition is much smaller than early studies suggested.

Although some effects of school resources may be statistically significant, they are typically small compared to total inequality in achievement. An example comes from a randomized evaluation of Teach For America (TFA). The teachers recruited by TFA come from highly selective colleges and are screened to make sure that they share TFA’s belief that effective teaching can close the achievement gap. These two qualities—
education at a selective college and high expectations for disadvantaged students—are among the qualities that research suggests make for effective teaching. Yet for the students in the study, TFA teachers were no more effective than other teachers at increasing reading skill. TFA teachers were about 10% more effective at increasing math skill—a sizable effect, on the fact of it—but this meant that children’s math scores only improved from the 14th to the 17th percentile. Three percentile points is progress, but it would take eleven years of three-point gains for children to reach the 50th percentile, a rough proxy for closing the achievement gap. The effect of TFA, although significant, was tiny when compared to the problem of socioeconomic inequality.

In sum, modern longitudinal studies do suggest that changes in school resources can affect achievement, although the effects on achievement are typically small.

*Schools and Inequality: Resolving the Paradox*

How are we to reconcile the findings of specific school effects, many of which disadvantage poor and minority children, with the broader finding, particularly from seasonal research, that schools create little of the inequality of achievement, and may even help to reduce it? There are several answers to this paradox.

First, although some troubling but atypical case studies might lead readers to believe otherwise (Rist 1973; Kozol 1991), the inequalities that exist within and between schools today are rather small compared to a number of outside standards. As our history of inequality found, school-resource inequalities are smaller than they were fifty years ago, and much smaller than they were one hundred years ago.

Inequalities in school are also much smaller than inequalities in families. For almost any school inequality that has been identified, a home-based inequality can be
identified that makes the school-based inequality look trifling. Teacher education is an example. Compared to middle-class children, poor children tend to have less educated teachers. But poor children also tend to have less educated parents, and the differences among parents are much larger than the differences among teachers. In North Carolina, for example, all fifth-grade teachers are college graduates, and a quarter have a graduate degree (Clotfelter, Ladd, and Vigdor 2006). By contrast, only a quarter of fifth-graders’ parents are college graduates, and a tenth dropped out of high school. One might imagine that education is more relevant to teaching than to parenting, but actually the reverse is true. Parental education is one of the most robust predictors of student test scores (Davis-Kean 2005), while teacher education (beyond a bachelor’s degree) seems to have little benefit for student achievement (Rockoff 2004). In North Carolina the estimated effect of advanced teacher education is actually negative (Clotfelter, Ladd, and Vigdor 2006).

Class size offers another opportunity to compare home and school inequalities. Although the differences are not large, black and Hispanic students do tend to have larger average class sizes than white students (Milesi and Gamoran 2006; Coleman et al. 1966). Larger class sizes tend to reduce learning (Angrist and Lavy 1999; Mosteller 1995; Jepsen and Rivkin 2002; Molnar et al. 1999; but see Hoxby 2000; Milesi and Gamoran 2006) probably because in larger classes teachers’ finite attention is more diluted across children (Downey 2001). But the dilution of scarce adult attention across competing children is much more unequal in homes than in schools. At school, the bottom quintile for kindergarten class size is 17 students, and the top quintile is 26; that is, a large class has about half again as many children as a small one (Milesi and Gamoran 2006). This is a small difference in adult attention compared to the difference between living with one
parent and living with both, or between being an only child and being one of two, three, or four. Resource dilution hurts disadvantaged groups more at home than at school, because disadvantaged groups have only slightly larger class sizes (Milesi and Gamoran 2006; Coleman et al. 1966), but substantially higher rates of fertility, single motherhood, and divorce (Blake 1989; McLanahan and Sandefur 1997).

In sum, school-based inequalities are smaller than they were in the past, and substantially smaller than inequalities that exist between children’s homes.

Although we typically emphasize ways that schools advantage affluent and high-achieving children, there are a number of ways that schools give extra resources and attention to poor and low-achieving children. Compared to middle-class children, poor children are somewhat more likely to attend full-day rather than half-day kindergarten, an increase in school exposure that has benefits for achievement (Cooper et al. 2010). In a recent survey, 60% of teachers reported that helping low achievers is a top priority, while only 25% placed that kind of priority on helping high achievers (Duffett, Farkas, and Loveless 2008). Teachers are certainly encouraged to focus on low achievers by the incentives in the Obama administration’s Blueprint for Reform (Duncan and Carmen Martin 2010), which outdoes the No Child Left Behind Act in threatening punitive action against schools with low achievement scores or large socioeconomic gaps. In fact, nearly all of state and federal education legislation since the War on Poverty has been aimed—first through extra resources, and now through accountability measures—at helping poor and low-achieving students to close the achievement gap.

Although sociologists rarely comment on the extra resources and attention given to low-achieving students, scholars and advocates who focus on “gifted” students have
noticed. To advocates for the gifted, the fact that the top decile of test scores is rising slower than the bottom decile is not a sign of healthy equity, as we argued above; instead, it is troubling because it suggests that the highest-achieving students are not being adequately challenged (Finn and Petrilli 2008). The prospect of a curriculum without tracking and ability groups is also troubling to advocates for the gifted, because it suggests that the challenges for advanced students will be reduced even further (Joseph S. Renzulli and Reis 1991).

This perspective of advocates for the gifted draws our attention to another way that schools reduce inequality. Schools reduce inequality by choosing not to track children more. Most sociologists take an ungrouped classroom as a sort of reference point, and from this perspective it appears that tracking and grouping create inequality. But if we view tracking and grouping as the norm, we realize that an ungrouped classroom, by not customizing lessons for children at levels, tends to compress achievement toward the mean. In an ungrouped classroom, teachers pitch their lesson plans to the middle of the achievement distribution, which results in a curriculum that is challenging for children at the bottom but limiting to those at the top. This may be important because, although tracking is pervasive in high school and ability grouping is common in kindergarten and first grade, there is a long span, running roughly from second grade to eighth, when children of different skill levels spend the majority of their time in common ungrouped classrooms.

The plight of gifted children is rarely discussed by sociologists, and it feels a little presumptuous even to bring it up. Why worry about gifted children? They don’t need help from school; they’re going to be fine. This question may highlight another way in
which our education culture reduces inequality. Education researchers, myself included, spend the vast majority of our energy trying to understand disadvantaged children—and occasionally succeeding in helping them (Treisman 1992; Case, Griffin, and Kelly 1999). By contrast, the research community spends very little energy trying to understand and help children at the top.

**A New Framework for Thinking About Inequality**

Education researchers sometimes try to divide inequality into a portion that is attributable to schools and a portion that is due to families and other non-school influences. A popular interpretation of the Coleman report, for example, is that inequality comes 80% from families and 20% from schools (Boulton 2010). Not everyone agrees with this interpretation, but is the correct division 60% vs. 40%? Is it 90% vs. 10%? The very question may be poorly framed because, as we saw in our discussion of seasonal research, there are ways of looking at the problem which suggest that families account for substantially *more* than 100% of inequality, so that schools are in the business of reducing inequality rather than adding to it.

A useful way of thinking about inequality is suggested by the formula for adding variances. Let A be a variable that represents individual student achievement. We can break A into two components: S, the component of achievement that is due to school, and N, the component of achievement that is due to non-school influences:

\[ A = S + N \]

If we think of inequality as the variance in achievement, then it easy to imagine that the total inequality or variance can be obtained simply by adding the variance of the school component and the variance of the non-school component:
Var(A) = Var(S) + Var(N)

But this equality only holds if the school and non-school components are *uncorrelated*. If there is some correlation between the school and non-school components, then the covariance has to be added to the variances:

Var(A) = Var(S) + Var(N) + Cov(S,N)

The interpretation of this covariance is crucial. A positive covariance would mean that the same children who get a large achievement contribution from non-school influences also get a large achievement contribution from their schools. For example, children from advantaged family backgrounds may benefit in school from high teacher expectations and high track assignments. This is the kind of pattern that is typically emphasized by sociologists.

But, as we have already discussed, some parts of the school and non-school contribution are *negatively* correlated. For example, children from disadvantaged family backgrounds may benefit in school from attending full-day rather than half-day kindergarten, from getting extra teacher attention to help them catch up, and from a curriculum that, being pitched near the middle of the achievement distribution, presents a target that is challenging but not unattainable.

In addition to positive and negative correlations, there are parts of the school and non-school contributions that have zero correlation. Within a given school, for example, one teacher may be more effective than another. But if students are assigned to teachers in an arbitrary way—say, by shuffling a deck of index cards—then whether a given student gets the better or worse teacher depends on the luck of the draw. There will be no
correlation between students’ non-school advantages and the advantage that they do or don’t receive by assignment to a more effective teacher.

So the origins of inequality do not depend only on whether inequality is greater inside or outside school—that is, whether \( \text{Var}(S) > \text{Var}(N) \) or \( \text{Var}(S) < \text{Var}(N) \). Inequality also depends on whether the school and non-school contributions covary positively or negatively. If the covariance is positive—if \( \text{Cov}(N,S) > 0 \)—then school and non-school inequalities do not simply add; there is an extra boost to inequality that comes from the fact that school and non-school inequalities advantage the same students. On the other hand, if the covariance is negative—if \( \text{Cov}(N,S) < 0 \)—then the contribution of school inequality is reduced by the fact that it advantages students who are disadvantaged elsewhere. In fact, if school inequality is small enough, and has a strong negative correlation with non-school inequality—that is, if \( \text{Cov}(N,S) < -\text{Var}(N) \)—then total inequality could actually be less than nonschool inequality. As we discussed earlier, more than 100% of inequality might come from the non-school environment, so that schools actually reduce inequality rather than contributing to it.

Within this framework, evidence that schools reduce inequality can be gleaned from the seasonal research of Downey, von Hippel, and Hughes (2008). The authors defined the non-school contribution as the learning rate observed during the summer, and they defined the school contribution (which they called “impact”) as the difference between the summer learning rate and the school-year learning rate. Between one school and another, the authors found that the variance of the non-school component was 7 times the variance of the school component, and the correlation between the school and non-school components was very strongly negative, at -0.94. This pattern of results is wholly
consistent with the argument that on balance schools do not contribute to inequality. Instead, they reduce it.

*Schools and Continuity*

In the previous section, we presented evidence that the school environment varies much less than the non-school environment (cf. Downey et al. 2004 especially Figure 1). In making this argument we were thinking about variation in space: variation from one school to another, or variation from one child to another within the same school. But the same generalization can also be used to summarize variation over time, which at least in recent history is smaller in school than it is elsewhere.

Compared to the rest of society, schools change very slowly. Despite the apparent churn of crises and reforms, what happens in most schools today, particularly in the classroom, looks very similar to what happened a generation ago (Tyack and Tobin 1994; Charles M. Payne 2008). There are several perspectives on schools’ slow rate of change. From the “public choice” perspective, schools systems are monopolies that have been “captured by providers”—teachers’ unions and administrators—whose interests often favor uniformity and stasis (Lubienski 2003). In addition, schools have an exceptionally complex governance structure requiring buy-in from multiple constituencies—not just from administrators and unions, but from school boards and parents’ organizations—so that sincere attempts at change can easily lead to stalemate (Charles M. Payne 2008).

Arguments like these are fundamental to the charter school movement, which sought to simplify school governance and make schools more accountable to “consumers” (students and parents), whose interests, advocates expected, would favor diversity and innovation. Yet with important exceptions (Tough 2009; Mathews 2009),
charter schools have not been as innovative as proponents hoped (Lubienski 2003). There must be additional forces holding schools in place.

The forces for statis in schools are legion, because schools are not just a convenient medium for conveying educational services. Schools have a broad range of functions, and truly disruptive innovations are bound to undermine at least one of them. For example, traditional schools provide day care and socialization, and an educational innovation like internet courses leaves those needs unmet. Parents expect a “real school” to have certain features, and innovative schools that do away with features like courses, classrooms, and levels leave many parents disoriented and threatened (Tyack and Tobin 1994). Finally, schools are part of a larger web of institutions, many of which are invested in the status quo. Schools that abandon curriculum and transcripts have to explain themselves to college admissions officers (Tyack and Tobin 1994), and schools that shorten summer vacation encounter organized resistance from summer camps and amusement parks (Chapter 3).

This argument may seem a bit contradictory since in the previous section we discussed two examples of major changes in the U.S. school system: the increase in the quality of segregated black schools between 1910 and 1950 and the rapid desegregation of Southern schools, starting in earnest in 1967-1970 and continuing at a slower pace until 1988 or so. It should be remembered, though, that both these changes met decades of organized resistance, and that desegregation has to some extent been reversed. It is not easy to change schools, and since the battles over segregation the pace of change has been very slow compared to changes in the rest of children’s lives.
Perhaps the biggest change to schools over the past 35 years has been the increase in competition for public school funds, initially through voucher programs and more recently through charter schools. Yet even today only about 10% of students attend charter schools. By contrast, over the same 35-year period, the fraction of mothers with school-age children who work outside the home has increased by more than 20%. The rise in working mothers has changed what children eat and who prepares it, where children spend their summers, what rewards girls foresee getting out of their education, and many other factors relevant to health and achievement (Bianchi, John P. Robinson, and Milkie 2007). Next to the rise of working mothers, even a reform as major as the charter school movement looks rather minor.

Looking Forward

The principles elaborated in this introduction guide a large part of my research agenda, including the empirical chapters in this dissertation.

In my first empirical chapter (Chapter 3), I study the effect of a calendar reform—the year-round school schedule—which aims to increase achievement among disadvantaged groups. The year-round calendar proposes to do this by eliminating the long summer vacation, when the relative achievement of disadvantaged groups declines most rapidly, and instead scheduling a number of shorter vacations throughout the year. As it turns out, this change does little or nothing to increase achievement, because rearranging the calendar does nothing to reduce the contribution to inequality of children’s non-school environments. The results are consistent with the principle that the majority of inequality comes from outside schools, so that minor changes to school practices, such as conversion to a year-round calendar, can do little to reduce inequality.
In the second empirical chapter (Chapter 4), I look at the relationship between schools and childhood obesity. I find that schools haven’t changed enough to explain more than a fifth of the rise in child obesity, and don’t vary enough to explain the differences in obesity between black, Hispanic, and white children. Both of these results are consistent with the principles that, compared to other aspects of children’s lives, schools are relatively fair and relatively slow to change.
Chapter 2
Seasonal Research Design

Some of the most revealing findings in the school-effects literature come from analyses of seasonal data. Seasonal data allow researchers to compare children’s physical or intellectual growth when they are in school, during fall, winter, and spring, and when they are out of school, during summer vacation. In some seasonal data, as well as non-seasonal data, there are also baseline measurements taken at the beginning of first grade or kindergarten, which allow researchers to estimate the growth that occurs outside of school in the months and years before school begins.

Compared to other common research designs in sociology and education, seasonal designs offer major advantages to researchers who wish to make causal arguments about the effects of schooling. By comparing in-school and out-of-school growth, seasonal researchers can evaluate the effect of school attendance on growth in body mass index, growth in average achievement, or growth in the achievement gap between students from different backgrounds. By contrast, using non-seasonal data it can be very difficult to evaluate the effect of schooling on growth.

In this chapter, I review the advantages of the seasonal research design, situating the discussion within the framework of counterfactual causal arguments and crossover experimental designs. I conclude that, although seasonal research falls short of the ideals presented by counterfactual arguments and experiments, the seasonal design nevertheless
has considerable advantages over other research commonly available when studying observational data.

I then present two multilevel growth models for seasonal data—a detailed three-level model and a simpler but more flexible two-level model. I point out assumptions that are made in the modeling process and present ways to test those assumptions or test the model’s sensitivity to them.

Finally, I review a recent attempt to use seasonal estimates as a way of ranking schools on “impact” or effectiveness. I discuss some of these assumptions made by the “impact” measure, and discuss ways that those assumptions could be softened or tested to produce an improved measure.

Seasonal Data and Counterfactuals

It has become increasingly common to frame causal argument in terms of alternative potential outcomes or counterfactuals (Morgan and Winship 2007). According to the counterfactual perspective, the statement that X causes Y means that if the value of X were different, the value of Y would be different as well. More formally, every unit of observation—which can be a student, a school, or even a national school system—has an observed value of Y, which corresponds to the actual value of X, as well as a potential value of Y that would be observed in the counterfactual situation where X had a different value. For a counterfactual argument to be meaningful, we must be able to imagine manipulating X in a way that didn’t affect other influences on Y.

From the counterfactual perspective, to say, for example, that schools increase inequality (or increase obesity), means that children would be more equal (or less obese), in the counterfactual situation where they did not attend school. Most research on school
effects fails to address this counterfactual. The finding that ninth-grade test scores are unequal, for example, or the finding that test-score inequality grows between first grade and ninth grade, does not necessarily mean that schools cause test-score inequality. To establish causality, we would need to know how unequal test scores would be, and how quickly inequality would grow, if children did not attend school.

Schools are such an integral part of life in developed countries that it is hard to imagine the pure counterfactual situation where children do not attend school at all. And it is just about impossible to imagine the situation truly required by a counterfactual argument, in which children do not attend school but nothing else that affects children’s achievement (or overweight) has changed. If schools were suddenly and permanently shut down, parents would react. Some parents would leave the work force; some would buy all-day care; some would demand day care subsidies from the government. Stay-at-home parents would start acting more like home-schoolers, and day-care centers would start offering instructional services that used to be offered by schools. In short, everything would change to compensate for the fact that schools had disappeared. The counterfactual of a modern society without schools is too radical to imagine.

Despite the impossibility of imagining a modern counterfactual society with no schools at all, we do get occasional glimpses of what such a society might look like. One glimpse comes from observing the expansion of mass education. Since 1950 a number of countries have spawned mass primary educations systems from next to nothing (Meyer, Ramirez, and Soysal 1992). These countries provide an opportunity to observe the differences between closely spaced cohorts with low and high levels of primary schooling. Even in the United States, black children in some Southern states experienced
sudden and dramatic increases in educational opportunity during the twentieth century. These changes, too, provide an opportunity to observe the effect of education on later outcomes such as wages (Card and Krueger 1996)—although the effects of educational changes are hard to separate from the effects of changes in discriminatory employment practices.

Delays or interruptions to schooling provide another glimpse into a counterfactual world without schools. Teacher strikes can children out of school for a month or more, typically to the detriment of children’s achievement and later earnings (William E. Caldwell and Jeffreys 1983; Belot and Webbink 2010). During the Civil Rights movement, schools were at various times shut down by strikes, walkouts, or government closures (Podair 2004; Lassiter and Lewis 1998). Changes to the school calendar can also affect school exposure; in Germany after reunification, for example, the switch from a spring to a fall start date kept students out of school for six months—a loss that reduced children’s test scores (Pischke 2007). Finally, birth-date cutoffs can affect children’s school exposure. Many school systems recommend or even require that a child be born before a certain date in order to begin formal schooling. A child born just before the cutoff will therefore begin school almost a year earlier in life than a child born just after the cutoff. The “discontinuity” between the test scores of children born before and after the cutoff date sheds light on the difference between spending a year of early childhood in school and spending that same year at home or in day care (Luyten 2006).

All of these circumstances provide glimpses into a counterfactual world where children do not attend schools. But these glimpses are narrow and limited in their generality. For the vast majority of children—children who are not affected by a strike, a
walkout, a closure, a change in school calendars, or a cutoff date close to their birthday—
these circumstances provide little insight into how that child would fare in a
counterfactual world without schools. If we want to evaluate the whole population of
schools and children, specialized circumstances are not sufficient.

For the vast majority of children, the only sustained opportunity we have to
observe life outside of school comes during summer vacation and during the five or six
years before formal schooling begins. These are exactly the periods illuminated by
seasonal and baseline measurements. By letting us compare how quickly and equally
children grow during schooled and unschooled periods, seasonal and baseline data give
us our best insights into the effects of school on the whole population.

Seasonal data do not fully answer the counterfactual question, because the way
that children spend summer vacation is surely different from the way that adults would
have them spend their time if school ceased to exist. But seasonal data certainly show us
the relative contributions of school and non-school periods in a society—our society—
that has a little of both. In our society, seasonal data suggest that most of children’s
academic skills come from periods of schooling, while most of the inequality in academic
skills comes from periods outside schools (Klibanoff and Haggart 1981; Hayes and
Downey et al. 2004). Similarly, non-school periods account for most of children’s excess
weight gain (von Hippel et al. 2007)—at least today (see Chapter 4).

Seasonal Data and the Crossover Design

The seasonal research design is similar to the crossover design that is frequently
used in clinical trials (B. Jones 1998) and occasionally used in other fields including
education (Lawson, Nordland, and Devito 1974). In a crossover trial, each subject spends some time in a treatment condition and some time in a control condition (or possibly an alternative treatment). A seasonal design is akin to a crossover trial in which the treatment is school and the control is vacation.

Seasonal and crossover designs share a crucial strength. In both designs, the researcher need not worry about differences between subjects receiving the treatment and subjects receiving the control. There can be no differences since each subject is observed under both conditions. Each subject serves as his or her own control.

In one respect, seasonal research is actually better than a typical crossover experiment. Whereas crossover experiments have to enlist volunteers who may or may not be representative of the population, seasonal research could potentially encompass the whole school-age population. Any child whose skills are measured in the fall and spring of successive school years is a participant in seasonal research. This group already includes children in the ten percent of U.S. schools that use the seasonal testing services of the Northwest Evaluation Association.

In other respects, though, seasonal research falls short of an ideal crossover design, as nearly any observational study must fall short of a designed experiment. By comparing the seasonal design to the ideal of a crossover experiment, we can highlight some of the weaknesses and assumptions of the seasonal approach. Fortunately, some of these assumptions can be tested—and will be tested in this dissertation. Where the assumptions bear out, the tests reassure us about the validity of seasonal results. Where the assumptions fail, we should exercise caution in interpreting seasonal results.
Compared to the ideal of a crossover experiment, one shortcoming of the seasonal design is that all participants receive the treatment and control at approximately the same time. That is, nearly all children are on vacation during the summer, and in school during fall, winter, and spring. In a clinical trial, this would never be permitted because the effect of the treatment will be confounded with the effect of the season. To take an extreme example, suppose that we designed a crossover trial of a flu vaccine, and proposed that all subjects receive a placebo in the winter and cross over to the vaccine in the spring. Such a trial would never be approved. Flu risk is much lower in spring than in winter, so if subjects only receive the vaccine in spring it is foreordained that flu incidence will be lower during the vaccination period—even if the vaccine is completely ineffective. Regulators would insist that the trial be balanced so that half the subjects received the vaccine in the winter and half received it in the spring. Only in a balanced design could the effect of the vaccine be disentangled from the effect of the season.

Seasonal school research suffers from a similar imbalance. It may appear that children learn at slower and more unequal rates when school is out—but perhaps learning is bound to slow and vary during the summer, whether school is in session or not. Likewise, it appears that children, at least today, gain weight faster when school is out. But perhaps there is some other aspect of summer—for example, the appeal of ice cream in hot weather—that affects weight more than the absence of school.

The confounding between schooling and season can be tested by looking at the four percent of children who attend schools that run on a “year-round” calendar. Year-round schools do not have a long summer vacation; instead they have shorter periods of school and vacation distributed more evenly throughout the year. Reassuringly, in
Chapter 3 I find that the learning of year-round students is also evenly distributed throughout the year. Evidently, it is schooling, not season, that influences learning.

Compared to a seasonal research design, a crossover trial has an additional advantage: it can be conducted blind. In a pharmaceutical crossover trial, for example, subjects receive a pill (or a shot, or a nasal spray) during the treatment period, but they also receive a pill during the control period. The control pill—known as a dummy or placebo—looks, feels, and tastes the same as the treatment pill; the only difference is the presence or absence of the active ingredient. Subjects are blind to whether they are being treated or not. If the personnel who have contact with the subject are also blind to the treatment, then the trial is double-blind. The purpose of blinding is to ensure that subjects and study personnel do not systematically change their behavior when they know that they are being treated. For example, subjects who think they are receiving an active flu vaccine may be less careful about washing their hands—while study personnel may be more careful. Changes in hand-washing may affect flu risk, and it is important that these effects not be confounded with the effect of the vaccine. Blinding ensures this.

Seasonal research, by contrast, is not blinded—nor can it be. Students and teachers are most definitely aware of when school is in session, and they adjust their behavior accordingly. The very notion of vacation encourages children, parents, and teachers to be less concerned about what is happening to children’s academic skills, and to devote more energy to non-academic pursuits. We encountered this problem earlier when I observed that summer vacation is not a pure counterfactual. What happens during summer vacation goes beyond the mere absence of school.
Even an ideal crossover trial is not methodologically watertight. The most prominent weakness of the crossover trial is the possibility of *carry-forward* effects. In a pharmaceutical crossover trial, for example, when a patient moves from a treatment period to a control period, it takes a while for the active ingredient to fully leave the patient’s blood. Similarly, when the patient moves from a control period to a treatment period, it takes a while for the active ingredient to build up. So the beginning of a treatment period is not a full treatment, and the beginning of a control period is not a pure control. Clinical trials deal with this problem by inserting short “washout” periods between treatment and control, or by ensuring that the treatment and control periods are long relative to the time required for the medication to ramp up or down.

Carry-forward is a possibility in seasonal school research as well. It may take a few weeks of fall for children to ramp up into an academic mindset, and it may take a few weeks of summer for children to fully relax. Carry-backward is a possibility as well. Children may begin to slack off toward the end of the academic year, and they may start dusting off their books before a new academic year begins.

While I acknowledge the possibility of carry-forward and carry-backward, I should point out that they do not threaten the validity of seasonal findings. First, carry-forward and carry-backward periods would likely be short compared to the 40 weeks of the school year, and possibly short compared to the 12 weeks of summer vacation as well. Second, both carry-forward and carry-backward periods would have the effect of *reducing* the contrast between school and summer. Since the typical finding is that the contrast between school and summer is already very strong, the possibility of carry-
forward and -backward merely suggests that the school-summer contrast may be even larger than past reports suggest.

Seasonal Measurement Schedules

Until now, I have written as though we knew exactly how much children gain during the school year and during summer vacation. In reality, seasonal gains have to be estimated from imperfect measurements taken on an imperfect schedule. Estimating gains requires some assumptions and entails some error. Fortunately the potential for error is relatively small compared to the very large differences typically observed between school-year and summer gain rates.

Seasonal data typically include two measurements academic year—one measurement in the spring and one in the fall. If these measurements coincided with the beginning and end of the school year, we would know, except for measurement error, exactly how much each child gains during the school year and during the summer. As Figure 2.1 illustrates, however, most schools schedule fall measurements weeks or months after the first day of school, and most schools schedule spring measurements well before the last day of school. In addition, measurements dates vary considerably from one school to another. (Within the same school, however, students are usually measured within days of each other.)
The discrepancy between the school schedule and the measurement schedule means that we have to assume something about what happens between the measurement dates and the first and last day of school. Seasonal research has evolved in how it handles the issue of these unmeasured school days. Early seasonal research ignored the problem, defining summer gains as any gains that took place between the spring measurement and the fall measurement. This definition exaggerated summer gains by attributing to summer any school-year gains that took place after the spring measurement or before the fall measurement—a problem that early seasonal researchers acknowledged (Klibanoff and Haggart 1981). More recent seasonal research extrapolates school-year gains beyond the measurement dates to the first and last day of school. In some data dates for the beginning and end of school are missing, but it is reasonable to plug in plausible dates from, say, the last week in August and the first week in June. Supplemental analyses indicate that estimates of learning change very little when typical school dates are used in place of the actual dates reported by schools. This is not surprising since the academic calendar is highly standardized: start and end dates vary little from one school to another. The important variation is in the dates that children are measured.

Extrapolation of school-year learning beyond the measurement dates results in smaller gains or even losses being attributed to summer, as illustrated in Figure 2.2.

<table>
<thead>
<tr>
<th></th>
<th>School year 1</th>
<th>Summer</th>
<th>School year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>In school</td>
<td>Out</td>
<td>In school</td>
</tr>
<tr>
<td>School 2</td>
<td>In school</td>
<td>Out</td>
<td>In school</td>
</tr>
<tr>
<td>Measurement</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 2.1. Typical measurement schedule for seasonal research.
Typically a linear extrapolation is used, which assumes that gains in the first and last months of the school year are about as fast as gains in the middle. A curved extrapolation should also be used, allowing for the possibility that learning ramps up at the beginning of the school year or tapers off at the end. Supplementary analyses, however, suggest that learning is approximately linear during the school year. Even if learning is slightly slower in the days between measurement and vacation, the typical finding of large differences between school-year and summer gains would persist.

![Graph showing reading improvement over months from Kindergarten to First grade.](image)

Figure 2.2. When learning rates are extrapolated beyond the test date, estimates of summer learning are reduced.

[From Downey, von Hippel, and Broh (2004). For a similar figure see Klibanoff and Haggart (1981).]

**Hierarchical Linear Growth Models**

To turn seasonal measurements into estimates of growth, I use a three-level hierarchical linear growth model (Bryk and Raudenbush 1988; Raudenbush and Bryk 2001). Two versions of the model can be specified: a two-level model with measurements
nested within schools, and a three-level model with measurements nested within children and children nested within schools. The three-level model provides detailed information about variation among students, while the two-level model provides more flexibility when variation among schools is the focus.

Three-Level Model

Level 1

Level 1 of the three-level model is a child-specific linear growth model. On each measurement occasion $t$ we regress each measurement on the number of months that child $c$ in school $s$ been exposed to each school year and each summer—i.e.,

$$Y_{mcs} = \alpha_{0cs} + \alpha_{1cs} \text{Schoolyear1}_{mcs} + \alpha_{2cs} \text{Summer1}_{mcs} + \alpha_{3cs} \text{Schoolyear2}_{mcs} + \ldots + e_{mcs}$$

where $e_{mcs} \sim iid\ normal\ (0, \sigma^2_m)$.

This model can be written more compactly in vector notation,

$$Y_{mcs} = \text{Exposures}_{mcs} \alpha_{cs} + e_{mcs}$$

where $\text{Exposures}_{mcs} = (1, \text{Schoolyear1}_{mcs}, \text{Summer1}_{mcs}, \text{Schoolyear2}_{mcs}, \ldots)^T$

and $\alpha_{cs} = (\alpha_{0cs}, \alpha_{1cs}, \alpha_{2cs}, \alpha_{3cs}, \ldots)$.

Here $Y_{mcs}$ represents the measurement of the dependent variable—either achievement or body mass index—for child $c=1,2,\ldots$ in school $s=1,2,\ldots$ on measurement occasion $m=1,2,\ldots$. The $\text{Exposures}$ variables represent the number of months that the child has been exposed to $\text{Schoolyear1}$, $\text{Summer1}$, $\text{Schoolyear2}$, and possibly subsequent school years and summers. The intercept $\alpha_{0cs}$ represents the initial value of the dependent variable for that child, extrapolated back to the beginning of the first school year. The
slopes $\alpha_{1cs}$, $\alpha_{2cs}$ … represent the child’s rates of growth during subsequent school years and summers. The error term $e_{mcs}$ represents random measurement error, or the difference between the true value and the measured value of the dependent variable. This error is assumed to be independent of the *Exposures* variables and normally distributed with a variance of $\sigma^2_m$ that can be different for each measurement occasion $m$. Note that the error variance $\sigma^2_m$ cannot be estimated simultaneously with the slopes and intercept $\alpha_{cs}$. There is one $\alpha_{cs}$ parameter for each measurement $Y_{mcs}$, and this means that there are no degrees of freedom left to estimate error. An estimate of the error variance must be obtained from another source.

I use different sources to estimate the error variance $\sigma^2_m$. When the dependent variable $Y_{mcs}$ is a test score, I estimate the measurement error using psychometric properties of the test. For example, if the test has a total variance of 100 and is 95% reliable, then the measurement error must account for 5% of the variance—so the error variance must be 5. On the other hand, when the dependent variable $Y_{mcs}$ is a body mass index, I estimate the measurement error by using the difference between closely spaced measurements. In the Early Childhood Longitudinal Study, for example, each child’s BMI was measured twice on each measurement occasion. Within-occasion discrepancies between the measurements suggests that the error variance was approximately 0.01 in squared BMI units. These methods of estimating error are not perfect, and an argument can be made that they overestimate or more likely underestimate measurement error.²

² In analyzing the Early Childhood Longitudinal Study, Kindergarten cohort of 1998, I estimated the measurement error in test scores using reliabilities calculated by applying item-response theory to the “internal consistency” of answers to questions on the same test (Rock and Judith M. Pollack 2002). An alternative estimate would use “test-retest” reliability, calculated by having children take the same test a second time. A reviewer once suggested that test-retest reliability would be smaller, though it could also be larger; in any case, it is unavailable. I estimate the measurement error in BMI by comparing height and
When I re-fit models with doubled or tripled error variances, however, the results change very little. Mistakes in estimating the error variance affect only level 1 of the model, and even there the effects are small. Even when doubled or tripled, the error remains small compared to the total variance.

**Level 2**

At level 2 of the model, I let the child-level growth parameters $\alpha_{cs}$ vary around a average $\beta_{0s}$ for each school $s$. In the simplest form of the model, all the child-level variation takes the form of a child level random effect $a_c$, which I assume to have a multivariate normal distribution across children:

$$\alpha_{cs} = \beta_{0s} + a_c, \text{ where } a_c \sim iid \text{ multinormal } (0, \Sigma_c)$$

A more developed model would also includes a vector of child-level regressors $X_c = (X_{1c}, X_{2c}, \ldots)$ with regression weights $\beta_{1s} = (\beta_{1s}, \beta_{2s}, \ldots)^T$:

$$\alpha_{cs} = \beta_{0s} + \beta_{1s} X_c + a_c$$

The child-level regressors contain child- or family-level characteristics such as race and ethnicity, household income, and parents’ education. If these regressors explain some child-level variation, then the variation in the child-level random effect $a_c$ will typically be reduced—especially if the regressors are centered around their school-specific means. Also, if the regressors are school-mean centered, I can continue to interpret the intercept as a school-level mean.

Despite the advantages of school-mean centering, it is sometimes preferable to center the regressors a different way, so that the intercept represents the mean for a group weight measurements taken by the same personnel on the same child using the same equipment on the same day. The estimated error would surely be larger if I varied the equipment and personnel and let a few days elapse between measurements.
of particular interest. In my chapter on year-round schools, for example, I centered the child-level regressors so that the intercept represented the mean for the Hispanic children who are most likely to attend year-round schools.

Level 3

At level 3 of the model, I let the school-average growth parameters $\mathbf{\beta}_{0s}$ vary around a grand average for all schools. In the simplest form of the model, all the school-level variation takes the form of a school-level random effect $\mathbf{b}_s$, which I assume to have a multivariate normal distribution across schools:

$$\mathbf{\beta}_{0s} = \mathbf{\gamma} + \mathbf{b}_s, \text{ where } \mathbf{b}_s \sim iid \text{ multinormal } (0, \mathbf{\Sigma}_s)$$

It is also possible, at least theoretically, to let some of the school-level regression weights $\mathbf{\beta}_{1s}$ vary randomly across schools, but I have not found this option helpful, either in fitting my models or in addressing the questions that I view as most important.

A more developed form of the level 3 model would also includes a vector of school-level regressors $\mathbf{X}_s = (X_{1s}, X_{2s}, \ldots)$ with regression weights $\mathbf{\gamma}_1 = (\gamma_1, \gamma_2, \ldots)^T$:

$$\mathbf{\beta}_{0s} = \mathbf{\gamma} + \mathbf{\gamma}_1 \mathbf{X}_s + \mathbf{b}_s$$

The school-level regressors can represent, for example, whether the school is public or private, whether it is in the city, the suburbs, or the country; some researchers also include child-level variables aggregated to the school level, such as the percentage of a schools students who qualify for free or reduced-price lunches. As in level 2, if the regressors explain some school-level variation, then the variation in the child-level random effect $\mathbf{a}_c$ will typically be reduced—especially if the regressors are centered around their grand means. Also, if the regressors are grand-mean centered, I can continue to interpret the intercept as a grand mean.
Mixed model

Levels 2 and 3 of the model can be combined into a mixed two-level equation:

\[ \alpha_{cs} = \gamma_0 + \gamma_1 X_s + \beta_{1s} X_c + b_s + a_c \]

This is a straightforward regression equation, except that it has two random components, one for the school level and one for the child level. These components are assumed to be independent of one another and independent of the school- and child-level regressors \( X_s \) and \( X_c \).

All three levels of the model can be combined into a more complicated three-level mixed equation:

\[ Y_{mcs} = Exposures_{mcs} (\gamma_0 + \gamma_1 X_s + \beta_{1s} X_c + b_s + a_c) + e_{mcs} \]

which shows explicitly how variation in gains is modeled using interactions between the \( Exposures_{mcs} \) variables and the school and child-level regressors.

The mixed model also has a very complex error structure, with school-, child-, and test-level errors \( (b_s, a_c, e_{mcs}) \), which are assumed to be independent of one another, independent of the school- and child-level regressors \( X_s \) and \( X_c \), and independent of the \( Exposures_{mcs} \) variables. To make things even more complex, the \( Exposures_{mcs} \) variables interact with the school- and child-level errors. The distribution of the dependent and independent variables is complex as well, with \( X_s \) varying across schools, \( X_c \) varying across schools and children, and \( Y_{mcs} \) varying across schools, children, and measurement occasions. The \( Exposures_{mcs} \) variables also vary across schools, children, and measurement occasions, but the child level accounts for very little of the variation since children from the same school are typically measured within days of each other.
Two-Level Model

The three-level model provides very specific information about the growth of individual children, but in some analyses I focus on the school level and individual children are not of interest. In such circumstances, I may use a two-level model instead, which is simpler, makes fewer assumptions, and often produces faster results when fit with statistical software.³

Another reason to consider a two-level model is to model the possibility that gains are not linear within the school year. The three-level model typically assumes linear gains; it has to, because it has one child-level parameter for every measurement occasion. For example, if there are four measurement occasions, in the fall and spring of successive school years, we can estimate at most four child-level parameters: (1) an intercept or starting point, (2) a linear gain rate for the first school year, (3) a linear gain rate for summer vacation, and (4) a linear gain rate for the second school year. I cannot add further child-level parameters to allow for nonlinear gains. Just as two measurements define a straight line, four measurements define three straight lines (and an intercept).

If we want the flexibility to allow for nonlinear gains, we can get it by using a two-level model. The two-level model gives up specificity about the child-level, and gains flexibility at the school level.

³ Many standard statistics procedure, including the MIXED procedure in SAS, are slow to fit the three-level model to a large data set. HLM software is much faster, often taking a minute or two to fit a model that takes an hour in SAS.
Level 1

In a two-level model, on each measurement occasion \( m \), the simplest form of the level 1 equation regresses each measurement on the number of months that child \( c \) in school \( s \) been exposed to each school year and each summer:

\[
Y_{mcs} = \beta_0s + \beta_{1s} \text{Schoolyear}_m + \beta_{2s} \text{Summer}_m + \ldots + u_{mc}\,
\]

where \( u_{cs} = (u_{1cs}, u_{2cs}, \ldots) \sim iid \text{ multinormal } (0, \Sigma_u) \)

Or in vector form:

\[
Y_{mcs} = \beta_0s \text{Exposures}_{mcs} + u_{mc}\,
\]

This is similar to level 1 of the three-level model, but there are two important differences. First, instead of giving each child \( c \) their own growth parameters \( \alpha_{cs} \), I use the same growth parameters \( \beta_{0s} \) for every child in the same school \( s \). Naturally these school-level growth parameters don’t fit individual children very well, so the residual term \( u_{mc} \) has to be much larger and more complex than the corresponding error term \( e_{mc} \) in the two-level model. Specifically, if \( u_{cs} = (u_{1cs}, u_{2cs}, \ldots) \) is the vector containing child \( c \)'s residuals for each measurement occasion, then this residual vector can be assumed to have a multinormal distribution with mean 0 and covariance matrix \( \Sigma_u \). This covariance matrix allows for the fact that residuals from different occasions can have different variances (heteroskedasticity), and residuals from the same child will be correlated (autocorrelation). I can constrain the residual covariance matrix to fit an autoregressive model, but this constraint is not necessary.

A more developed form of the level 1 model could include child-level covariates \( X_c \), as well as nonlinear transformations \( T(\text{Exposures}_{mcs}) \), such as splines or squares, of the school-year and summer exposure variables. Nonlinear transformations allow for the
possibility that school-level gains may be faster or slower at different times of year. The two-level model can accommodate nonlinear gains because, unlike the three-level model, the two-level model does not lock us into a linear gains assumption at the child level.

Interactions between the $\text{Exposures}_{mcs}$ and the child-level covariates are also a possibility; these would account for the possibility that different types of children gain at different rates. An example of a well-developed level 1 model, including both child-level covariates and transformed exposure variables, might look like this:

$$Y_{mcs} = \beta_0 s T(\text{Exposures}_{mcs}) + \beta_1 s X_c + \beta_2 s T(\text{Exposures}_{mcs}) + u_{mcs}$$

**Level 2**

Level 2 of the two-level model is just like level 3 of the three-level model. I let the school-level growth parameters $\beta_{0s}$ vary around a grand average for all schools. Some of the school-level variation is explained by school-level covariates $X_s$, and some takes the form of a school-level random effect $b_s$:

$$\beta_{0s} = \gamma_0 + \gamma_1 X_s + b_s, \text{ where } b_s \sim iid \text{ multinormal } (0, \Sigma_s)$$

**Mixed model**

Both levels of the model can be combined into a single mixed equation:

$$Y_{mcs} = (\gamma_0 + \gamma_1 X_s + b_s) T(\text{Exposures}_{mcs}) + \beta_1 s X_c + \beta_2 s T(\text{Exposures}_{mcs}) + u_{mcs}$$

which shows explicitly how variation in gains is modeled using interactions between the $\text{Exposures}_{mcs}$ variables (possibly transformed) and the school and child-level regressors.

Again the mixed model has a very complex error structure, with level 1 and level 2 errors ($u_{cs}, b_s$) that are assumed to be independent of one another, independent of the school- and child-level regressors $X_s$ and $X_c$, and independent of the $\text{Exposures}_{mcs}$
variables. Again, additional complexity comes from the fact that the $\text{Exposures}_{mcs}$ variables interact with the school- and child-level regressors and errors.

**Contrasts and “Impact”**

The crucial questions in seasonal research require comparing quantities across season. Is the average gain rate greater in the summer or in the school year? What about the variance in gain rates? What about the gap in gain rates between races or socioeconomic groups?

I can answer these questions by using linear contrasts, which can estimate the difference between any two quantities: two means, two coefficients, two variances. The statistical theory behind contrasts is simple. For example, let $q = (v, s)$ be a vector maximum likelihood estimates for summer vacation and for the following school year. The estimated quantities can be anything: average gains, the average gap in gains between rich and poor students, the variance in gains. Then if $c = (-1, 1)$ is a contrast vector, $cq$ is an estimate for the difference between the school year and summer quantities. If $S$ is the covariance matrix for the estimates in $q$, then the point estimate $cq$ has a large-sample normal sampling distribution with a standard error of $(cSc^T)^{1/2}$ (Richard A. Johnson and Wichern 2001). Contrasts can also be used for more complicated comparisons such as the difference between a quantity during the summer and the average of the same quantities across the school year before and the school year afterward. Contrasts offer a valid way of testing whether gains are faster or less variable during the school year than during summer vacation (Downey et al. 2004; von Hippel et al. 2007).
But the interpretation of contrasts has been pushed further. Downey, von Hippel, and Hughes (2008) have proposed that the contrast or difference between school-year and summer learning rates—or more specifically, between learning rates during the school year and learning rates during the summer before—can be interpreted as the unique contribution or “impact” of schooling on children’s learning. The idea is that summer learning represents the unique contribution of non-school factors—e.g., the home environment, the neighborhood, genetics—while school-year learning represents the combination of all factors, school and non-school. It follows that simply subtracting the summer learning rate from the school-year learning rate is equivalent to removing non-school factors from school-year learning. What remains must be the unique contribution of schools.

The assumptions behind the impact measure can be questioned on both theoretical and empirical grounds. The first assumption is that the non-school environment contributes the same amount to learning during the school year as during the summer. This seems doubtful. During the school year children have less time to spend in their non-school environments, and they surely spend their non-school time differently than during summer vacation. This insight emerged earlier when I pointed out that summer vacation does not represent a counterfactual world without school, and when I pointed out that the seasons do not add and subtract school from children’s lives without their knowledge, like a medication in a blind crossover experiment. One of the key findings of seasonal research—that SES gaps grows more slowly during school than during summer vacation—confirms that the contribution of at least some non-school factors does indeed shrink when the school year begins (Klibanoff and Haggart 1981; Hayes and Grether...
A second assumption of the impact measure is that learning follows an additive model. That is, children learn a certain amount because of non-school factors, and school adds its unique contribution to the total. According to this additive way of thinking, the learning that schools adds is independent of the non-school contribution, although the school and non-school contributions may be positively correlated. Positive correlation would occur if the advantaged children who learn fastest during summer vacation probably also attend better schools during the academic year.

The empirical research on impact, however, suggest that something other than an additive process might be at work. In the early years of school, it turns out that summer learning actually has a negative school-level correlation with kindergarten learning (-0.30), with first-grade learning (-0.19) and with first-grade impact (-0.82) (Downey et al. 2008). This negative correlation means that the children who learn most slowly during summer vacation actually learn somewhat faster and get much more school impact than other children during the kindergarten and first grade school years. If impact is a measure of school quality, if what schools add to learning is a measure of their effectiveness, then these negative correlations suggest that the most disadvantaged children—the children whose summer environments do the least to build academic skills—attend the most effective schools. This is not impossible, but it would be very surprising.

The negative correlation between school year and summer learning suggest that the contributions of school and non-school factors may follow something other than an additive model. It may be that a given school doesn’t simply add a fixed amount to
children’s learning, it adds as much as it must to help children reach a required level of proficiency. The kinds of children who learn quickly during the summer don’t have to learn as much during the school year. Conversely, the kinds of children learn slowly, or even lose skills, during summer have to make up for lost time during school. This interpretation makes some sense for kindergarten and first grade, since those early school years are geared toward helping children reach a level of proficiency. Future research should examine whether school-year and summer learning continue to be negatively correlated in the later years of school.

Despite these issues, the idea that summer learning provides a useful proxy for non-school influences on learning—quite possibly a better proxy than conventional measures such as household income—is promising and deserves further investigation. It may be that the best measure of school impact would not simply subtract summer learning from school year learning; instead, the best impact measure would subtract only a fraction of the summer learning rate. Alternatively, it may be that the idea of subtraction is misguided and a better approach would be simply to match schools on their summer learning rates. That is, among schools with similar summer learning rates, the most effective school is the one with the fastest school-year learning. A matching approach would mean that I could not compare the effectiveness of schools whose students learn at very different rates during the summer. But it may be that those comparisons would not be meaningful anyway. The kind of school that helps slow summer learners to catch up may be qualitatively different from the kind of school that challenges fast summer learners to make further gains.
In evaluating the pros and cons of the impact measure, our discussion has relied primarily on conceptual arguments and on the internal consistency of statistical analyses. What I would really like to do is compare the impact measure to some sort of external “gold standard” for measuring school quality. Unfortunately, such gold is rare, since other ways of evaluating schools—e.g., on achievement, achievement gains, school resources, nearby property values—make at least as many questionable assumptions as the impact measure. Only when children are randomly assigned to attend different schools—as when seats in charter or magnet schools are awarded by lottery—is there really a gold standard for evaluating school effectiveness. The ideal setting for developing the impact measure would be an experiment that included both random assignment and a seasonal measurement schedule.

**Missing values**

Like other social-science data sets, seasonal data often have missing values. Missing values can be filled in using multiple imputation strategy, but imputation is challenging when the imputed data are to be used in a multilevel growth model. A principle of imputation is that the imputation model should be consistent with the analysis model, and a multilevel growth model is exceptionally complex. The research on multilevel imputation is limited, and popular imputation software—I used the MI procedure in SAS 9—assumes that the data to be imputed have only one level and each case is independent.

In order to impute missing values credibly, I had to investigate and develop several imputation techniques. My first problem was to distinguish the measurement level (level 1 in the three-level model) from the child level (level 2). That is, I needed to
account for the correlation among repeated measurements of the same child. I dealt with this correlation by following Allison’s (2002) suggestion and formatting the data so that all measurements on the same child appeared on a single row. The imputation software would not recognize repeated measurements, but it would recognize that values that appear in the same row can be correlated.

My next problem was to distinguish the child level from the school level. There are some variables that vary from child to child in the same school (for example, household income) and some variables that only vary from one school to another (for example, whether the school is public or private). I handled this distinction by rolling up a school-level data set containing all school-level variables $X_s$ as well as school-level averages of child-level variables $X_c$. I counted repeated measurements $Y_{mcs}$ as child-level variables since I had already rolled them up as separate columns in each child’s row. I counted the $\text{Exposures}_{mcs}$ variables as school-level since 90% of their variation lies between schools rather than within them. I then imputed missing values in the school-level file, appended the imputed school-level variables to the child-level file, and imputed missing child-level variables conditionally on the imputed school-level variables.\footnote{I carried out this procedure once. A more advanced approach might iterate the process, since imputing the child-level variables changes the averages that are rolled up to the school-level file. I doubt iteration would have improved the results much, since in practice there tend to be few missing values at the school level.}

Finally, I unrolled the child-level file so that each measurement of the same child got its own row. This is the format—one row per measurement—that is typically required to fit a multilevel growth model.

A further problem was how to account for the interactions evident in the mixed-model equations. Note that there will be interactions in both the school-level and the child-level file. I simulated a variety of methods for imputing interactions and concluded
that the best general-purpose method is one that I call *interact, then impute* (von Hippel 2009). Under this method, I calculate interactions in the incomplete data, and then impute missing interactions like any other variable. In practice, I found that there was high collinearity among the interactions, and this sometimes prevented the imputation procedure from converging on the parameters of an imputation model. Following a suggestion by Schafer (1997), I helped the imputation to converge by using a ridge prior, which artificially reduces collinearity.

The final problem was that, despite the pains that I took to impute variables carefully, my imputation model was surely imperfect. The complex error structure of the mixed model is nearly impossible to reproduce in imputed data, and any attempt to replicate it is a compromise. I reduced the consequences of this compromise by developing a procedure that I call *multiple imputation, then deletion* (MID) (von Hippel 2007). Under MID, I impute the data as described above, but then delete any measurement occasions with imputed $Y$. Deleting these cases helps to reduce dependency on imperfectly imputed data, especially since occasions with imputed $Y$ often have imputed $X$s as well. There is no cost to deleting these occasions, since even if the imputation model is perfect, imputed $Y$s add nothing but noise to statistical estimates (von Hippel 2007).

In the course of developing this imputation strategy, a number of reasonable alternatives were tried. None of these alternatives materially affected the results. The important effects in seasonal research are typically so clear that minor variations in the imputation procedure cannot shake them.
Chapter 3

What Happens to Seasonal Learning Patterns in Year-Round Schools?

The core finding of seasonal research is that learning proceeds very differently when school is session than when school lets out for summer vacation. When school is in session, nearly all children learn quickly, and children from disadvantaged backgrounds learn nearly as quickly as children from the middle class. When school lets out for summer vacation, by contrast, learning slows down for nearly all children, and especially for disadvantaged children. One consequence is that the achievement gap between advantaged and disadvantaged children grows primarily during the summer rather than during the academic year.

In interpreting these patterns, researchers typically assume that schools cause the acceleration and uniformity of learning rates that is observed during the school year. The deliberate instruction provided by schools accelerates learning, and the uniform nature of the school environment ensures that children from different background learn at similar rates, despite the inequalities in children’s non-school backgrounds. When school lets out, the instruction and uniformity stop, so that children learn more slowly and at more unequal rates. The fact that this happens during the summer is arbitrary; the same thing would happen if schools let out at a different time of year.

Yet it could be that something special about summer contributes to the differences between school-year and summer learning. While it seems likely that schools play the predominant role, it is not implausible that something else about summer—for example,
the attractions of outdoor recreation—play a role in slowing the acquisition of academic skills. The disappointing results of some high-profile summer interventions (Heyns 1987; Carter 1984)—though by no means all of them (Cooper et al. 2000; Borman and N. Maritza Dowling 2006)—suggest that the summer environment presents special challenges to learning. If this is true, then perhaps schools get too much credit for the acceleration in learning that occurs when school begins in the fall.

To state the problem in methodological terms, the seasonal research design is a lot like the crossover design that is often used in clinical trials (B. Jones 1998). In a crossover design, each participant is exposed to at least one period of treatment (here school) and one period of control (vacation). Viewed as a crossover design, a weakness of seasonal research is that nearly all children are in the control condition simultaneously, during the summer, and also under the treatment condition simultaneously, during fall, winter, and spring. In a clinical trial, such a design would never be approved because the effect of the treatment is confounded with the effect of the season. To take an extreme example, suppose that researchers proposed a crossover trial of a flu vaccine, in which all subjects receive a placebo in the winter and cross over to the vaccine in the spring. The problem with this design is that flu risk is much lower in spring than in winter, so if subjects only receive the vaccine in spring it is foreordained that flu incidence will be lower during the vaccination period—even if the vaccine is completely ineffective. Regulators would insist that the trial be balanced so that half the subjects received the vaccine in the winter and half received it in the spring. Only in a balanced design could the effect of the vaccine be disentangled from the effect of the season.
In education research, we can address the question of seasonal confounding by looking at the three percent of children who attend schools that run on a “year-round” calendar. In year-round schools, season is not confounded with schooling because year-round schools do not have a long summer vacation. Instead, year-round schools have shorter periods of school and vacation distributed more evenly throughout the year. Year-round schools have about the same number of school days as nine-month schools—typically between 175 and 180 days of instruction—but they distribute those school days much more evenly across the year.

If it is mainly schooling and not season that affects learning, then year-round students, especially those from disadvantaged backgrounds, should display faster summer learning than nine-month students, because year-round schools provide instruction during summer while nine-month schools do not. By the same reasoning, nine-month students, especially those from disadvantaged backgrounds, should display faster learning than year-round students in fall, winter, and spring, because nine-month schools schedule all their instruction in fall, winter, and spring, while year-round schools have only three-quarters of their instruction in those seasons.

The distribution of learning in year-round schools has more than methodological implications. It has policy implications as well. Advocates for year-round schools often claim that they can increase average achievement and reduce achievement gaps by eliminating the problems of summer learning (Ballinger 2000). The analysis above suggests that advocates may be right about summer learning, but may not have fully considered the effects of year-round schools on other parts of the year.
In this paper, we review the background and justification for year-round schools, then test the effect of year-round calendars using longitudinal data that lets us estimate learning rates both during the summer and during the nine months of the traditional academic year. Consistent with our predictions, we find that year-round calendars increase learning during summer but reduce learning during fall, winter, and spring. These results validate the assumptions of past seasonal research, suggesting that it is schooling, not season, that affects learning. On the other hand, the results are somewhat discouraging with respect to the promise of year-round schools. On balance, over a twelve-month period, it appears that year-round calendars do not increase learning by more than a token amount.

Summer Learning

Summer vacation is bad for children’s academic achievement. In the early grades, children learn word and number skills much more slowly during summer vacation than they do during the school year (Cooper et al. 1996). In the later grades, when schoolwork grows more academic and less similar to out-of-school activities, most children return to school after summer vacation having forgotten over a month of the math, science, and social studies that they learned the previous year (Cooper et al. 1996). This summer forgetting represents more than a lost opportunity; it also reduces the amount that can be taught during the school year. Instead of introducing new material, many teachers must begin the fall by reviewing a substantial fraction of the material that students learned the previous spring but forgot over the summer.

Summer learning is particularly slow for poor children with less-educated parents (Heyns 1978; Entwisle and Alexander 1992; Downey et al. 2004; Hayes and Grether
1983). In fact, it is mainly during the summers between academic years that poor children lose ground to their middle-class peers. Although poor children are already behind when they enter school in kindergarten or first grade, during the school year they nearly hold their own by learning almost as fast, on average, as their more affluent peers (Heyns 1978; Entwisle and Alexander 1992; Downey et al. 2004; Hayes and Grether 1983). It is the summer vacation that sets poor children further and further back.

Compounded year after year, summer setback steadily erodes the competitive position of poor children, at first in academic achievement and later in the competition for jobs. By ninth grade, the achievement gap between poor and middle-class children is more than three times larger than it was at the start of first grade—and most of the expansion in the gap has occurred during summer vacations, not during the school years (Alexander et al. 2007; Hayes and Grether 1983). Poor children’s low ninth-grade achievement, in turn, routes them away from college-preparatory courses and reduces their chances of graduating from high school, or of attending college if they do graduate (Alexander et al. 2007). And if poor young adults lack a college degree, or at least a high-school diploma, it is much harder for them to escape poverty today than it was a generation ago (Barton 2005).

Sociologists typically attribute summer learning loss to deficiencies in children’s home lives and neighborhood resources (Heyns 1978; Alexander et al. 2001; Downey et al. 2004). Compared to schools, most children’s homes do relatively little to develop academic skills, and the level of stimulation is especially low in the homes of poor families (Linver, Brooks-Gunn, and Kohen 2002, 1999; Risley and Hart 1995; Gary W Evans 2004). In school, by contrast, children share a common environment that helps the
poor learn almost as quickly as the middle class. Although some schools are better than others, the differences between rich and poor schools are much smaller than the differences between rich and poor homes and neighborhoods (Downey et al. 2004).

**Year-Round Schools**

At least three educational interventions attempt to increase summer learning. The first and most obvious intervention is summer school. Most evaluations of summer school programs find that they have a positive effect on achievement (Cooper et al. 2000). The positive effect may be larger for middle-class children than for low-achieving poor children, but since poor children are more likely to attend (Cooper et al. 2000), on balance summer programs probably tend to reduce inequality across socioeconomic groups. Summer school programs are difficult to evaluate, because the children who enroll in them are typically much lower-achieving than the children who do not enroll. The best summer-school evaluations match participants to nonparticipants with similar achievement levels, or, even better, assign children at random to summer school or non-summer-school conditions. On average, the best-designed studies find that summer school can boost achievement scores by 10% to 15% of a standard deviation (Cooper et al. 2000), which is impressive given that summer school programs last less than three months. Nevertheless, the effect is small in an absolute sense, and monthly learning rates during the summer rarely approach the monthly learning rates that are observed during the academic year (Cooper et al. 2000).

A second intervention is the extended-year school calendar, which lengthens the academic year from the usual 175-180 school days to more than 200 school days including three or more weeks in the summer. At least four east Asian countries—
Singapore, Korea, Japan, and Hong Kong—have academic years with more than 200 school days, and these are also the highest scoring countries on international tests of mathematics skill (Trends in International Mathematics and Science Study 2003). Yet east Asian countries differ from the U.S. in other ways as well, and it is not clear to what extent their high test scores are due to their longer school years. In the U.S., less than one elementary school in a thousand uses an extended calendar, so that U.S. evaluations of the extended-year calendar are rare and can focus on as few as three schools (Gandara and Fish 1994). In addition, the effect of an extended-year calendar can be hard to evaluate because schools that extend the year often adopt other reforms at the same time. Charter schools in the Knowledge Is Power Program (KIPP), for example, have high scores that might be attributed to their extended-year calendar, yet KIPP schools also require extensive parental involvement, and they tend to attract students who are high-scoring, relative to other children in their impoverished neighborhoods, even before they enroll (Carnoy et al. 2005).

A third attempt to increase summer learning is the year-round school calendar—the focus of this paper. Although still rare, the year-round calendar is much more popular than the extended-year calendar, and is currently in use at about 3% of US public elementary schools. Unlike an extended-year calendar, with which it is often confused, a year-round calendar does not actually increase the days of instruction beyond the usual 175 to 180. Instead, year-round schools take those 175-180 days and redistribute them, replacing the usual schedule—nine months on, three months off—with a more

---

5 Among the 935 schools surveyed by the Early Childhood Longitudinal Study, Kindergarten cohort, none scheduled more than 183 school days.
6 In 2007, 2,127 elementary schools followed a year-round calendar (National Association for Year Round Education 2007). This figure is about 3% of the 67,291 public schools with elementary grades in 2005-06 (Nation Center for Education Statistics 2009).
“balanced” schedule of short instruction periods alternating with shorter vacations across all four seasons of the year. There are several year-round calendars in use; the most popular alternate nine- or twelve-week instructional periods with three- or four-week vacations or “intersessions.” In some intersessions, schools can include supplemental academic instruction, though many offer nonacademic activities such as sports (National Association of Year-Round Education 2005).

Year-round calendars are not always meant to increase summer learning. In 40 percent of year-round schools, the year-round calendar is part of a multi-tracking plan whose purpose is to alleviate crowding (Cooper et al. 2003). In a multi-tracked school, the student body breaks into three or four groups who attend school on a staggered schedule. When one group is on vacation, the other groups are in session, so that only two-thirds to three-quarters of the students are in the building at a given time. Class sizes are reduced, and higher enrollments can be accommodated without building new schools. Some fast-growing school districts have become dependent on the year-round calendar as a way to handle increasing enrollments. In greater Las Vegas, for example, nearly half of all elementary schools use a year-round calendar, and it has been estimated that switching to a traditional calendar would require half a billion dollars in new school construction (Year Round Calendar Study Group 2007).

In 60 percent of year-round schools, however, the calendar is not part of a multi-tracking plan and is evidently chosen for reasons other than crowding reduction. Both advocates and adopters of year-round calendars often claim explicitly that the calendar can reduce summer learning loss and therefore boost achievement (Stenvall 1999). The former president of the National Association for Year-Round Education, for example, has
asked, “Why does it take so long for some secondary school communities to understand that one way to reduce summer learning loss is to reduce the summer vacation?” (Ballinger 2000). The chief education officer of the Chicago Public Schools has said that most of Chicago’s 29 year-round schools adopted the year-round calendar in hopes of reducing summer learning loss (Sadovi and Little 2007); in fact, Chicago’s Mayor Daley believes that year-round calendars, as well as summer learning programs, are essential to international competitiveness—although his ultimate goal is extending the school year (Spielman and Grossman 2005).

Notice that the goals of crowd reduction and summer learning increases are not mutually exclusive. Even in multi-track schools, there may be some expectation that year-round calendars will increase test scores.

But can year-round calendars really fix the problem of summer learning? Remember that year-round calendars do not eliminate or even reduce summer vacation; instead, they take the days of summer vacation and redistribute most of them to other times of the year. Unless there is something special about the months of June, July, and August, it seems likely that redistributing vacation time to other months will simply move the problem of summer learning around. Instead of a single season where academic learning slows down or reverses, on a year-round calendar we might expect a larger number of short slowdowns and reversals, distributed across all seasons of the year. These three- and four-week slippages will be less noticeable and smaller in their individual impact—but over the course of a calendar year, the cumulative result may be similar to the major setback usually observed during summer vacation.
Although sociologists occasionally voice support for year-round education (e.g., Alexander et al. 2007), there is little in the sociology of education to suggest that year-round calendars can remedy the disadvantages apparent during summer vacation. If summer setback is a symptom of deficiencies in poor children’s non-school environments (Downey et al. 2004), then year-round calendars do not really address the problem. Year-round calendars do not increase the time that children spend in school, nor do they increase the academic stimulation that children experience outside of school.

While the sociological perspective is discouraging, cognitive psychology offers some hope that year-round calendars can increase achievement. Psychological experiments have repeatedly shown that “spaced practice” is better than “massed practice”. That is, people learn and retain more when they practice not in a single marathon session, but in a series of short sessions—especially if the intervals between those sessions are also short (Dempster 1988; Donovan and Radosevich 1999). This “spacing effect,” or “distribution-of-practice effect,” has been demonstrated in school-like activities such as memorizing vocabulary words and retaining ideas from readings and lectures (Dempster 1988). At first glance, then, we might hope that learning would increase if the three-month summer vacation were replaced with a series of three- or four-week breaks.

On the other hand, the kind of spacing effect observed in the laboratory may have little relevance to the school calendar. Most spacing research has focused on short time periods, with intervals of seconds, minutes, or hours between practice sessions. There has been less research that looked at intervals of days, and none that looked at intervals of weeks or months (Donovan and Radosevich 1999; J. Donovan, personal communication,
Sept. 21, 2006). The research on longer time periods has been fairly discouraging; when the interval between practice sessions is longer than a day, spaced practice is only a little more effective than massed practice, and spaced practice may even be less effective than massed practice if the mental tasks being learned are especially demanding (Donovan and Radosevich 1999, especially Table 3).

In short, although psychological findings suggest that the traditional three-month summer break is not optimal for learning, research provides no very strong reason to expect that year-round breaks of three or four weeks are any better.

**Preview of Results**

In this paper, we test the effect of year-round calendars using longitudinal data that lets us estimate learning rates both during the summer and during the nine months of the traditional academic year. Although a number of authors have speculated that year-round schools can increase achievement by increasing summer learning, ours is the first piece of research to look directly at what happens to summer and non-summer learning under a year-round calendar.

The results suggest that year-round calendars do not, on average, increase total learning; instead, they merely redistribute periods of learning and forgetting across the calendar year. During the summer, children in year-round schools do learn more quickly than children who are on vacation, but during the rest of the year, children in year-round schools learn more slowly than children in traditional schools. This pattern of results is consistent with the view that learning varies with school exposure. During the summer, year-round schools are in session while nine-month schools are on vacation—so during the summer year-round students outlearn nine-month students. But during the rest of the
year, year-round schools have regular breaks of three to four week, while nine-month schools are in session more continually—so nine-month students outlearn year-round students during fall, winter, and spring. On balance, over a twelve-month period including both summer and the other three seasons, the amount learned in year-round schools is almost exactly the same as the amount learned in schools using a traditional nine-month calendar.

Although past research has suggested that year-round calendars may at least be helpful to disadvantaged children, our own results for such children are mixed. As it turns out, most children in year-round schools are at least moderately disadvantaged, so our basic finding that children learned no more in year-round schools than did comparable children elsewhere may be viewed as evidence against the claim that year-round schools help the poor. If we look further down the socioeconomic ladder and focus on children who are deeply disadvantaged—that is, children who are poor even compared to the general year-round population—we find weak evidence that such children learn more on a year-round calendar than on a traditional calendar, but the evidence is too inconsistent and uncertain to convince a skeptic.

On balance, the results suggest that year-round calendars do little to fix the general problem of summer slowdown, or to fix the problem of poor children’s summer losses relative to more affluent children. Instead, year-round schools mainly hide the disadvantages evident in the summer by sweeping those disadvantages under the rug of fall, winter, and spring.
Past Research on Year-Round Schools

In a meta-analysis, Cooper et al. (2003) reviewed 39 past studies of year-round school calendars. Depending on how the results are interpreted, the results of the meta-analysis may be viewed as neutral on the effect of year-round calendars, as mildly encouraging, or as uninformative.

On one hand, most of the studies in the review found that children in year-round schools had higher average scores than students following a traditional school calendar. On the other hand, the percentage of studies showing an advantage for year-round schools was hardly overwhelming, at 62%, and the average effect size was trivial, at about 5% of a standard deviation. The effect did seem to be a bit larger, about 10% of a standard deviation, for students from disadvantaged communities.

As Cooper et al. (2003) acknowledged, the results of their meta-analysis should be viewed with skepticism, because most research on year-round calendars has been poor in quality. 59% of past studies did not control for confounding differences between students in year-round and traditional schools (Cooper et al. 2003)—which is troubling since, as we will show, year-round students tend to have a variety of non-school disadvantages that reduce their achievement. In addition, Cooper et al. (2003) were concerned that no prior study “explicitly controlled for the number of days that school had been in session before achievement outcomes were measured.” For example, if a fall test is given in early September, year-round students have a head start since they have already been in session more than a month while their nine-month competitors are just getting started. One the other hand, if a spring test is given in early June, nine-month students have an unfair advantage: students in a nine-month school would have almost a
full year of instruction behind them, while students in a year-round school would still
have a month of instruction to go.

McMillen (2001) pointed out other weaknesses of past studies, including
“collapsing achievement outcomes into categories such as ‘at or below grade
level’…failure to report any tests of statistical significance or measures of effect size…,
failure to differentiate between year-round and extended-year schools,” avoidance of peer
review, and failure to “account statistically for the nesting of students within schools.”
McMillen’s (2001) own study—perhaps the best research on year-round schools to
date—avoided most of these weaknesses. Focusing on public schools in North Carolina,
McMillen (2001) compared 106 year-round schools to 1,364 nine-month schools in a
design that accounted for student demographics, prior achievement, and the nesting of
students within schools. The results showed no significant effects of school calendars on
achievement, except for initially low-achieving students, whose final scores may have
been boosted by a small amount, again less than 5% of a standard deviation.

In addition to the shortcomings pointed out earlier, no prior research has
examined the effect of year-round calendars on summer learning. Although it is
commonly assumed that the achievement benefits of a year-round calendar would come
mainly from an increase in summer learning\(^7\), no study has looked at this question
directly. In saying this, we do not mean to criticize past researchers; it is impossible to
look at summer learning unless students are given tests near the end of the spring and the

\(^7\) For example, Cooper et al. (2003) suggest that the average summer learning loss—about 10% of a
standard deviation in achievement—represents an upper bound on the potential impact of year-round
calendars. On the other hand, McMillen (2001) suggests that year-round calendars may boost achievement
through “intersession” instruction—that is, extra instruction targeting at-risk students during the three- or
four-week breaks between regular instruction periods. Of course, intersession instruction effectively
increases the days of instruction well beyond 180, turning a year-round school into an extended-year
school.
beginning of the next fall. In our study, we were fortunate to have data from such spring and fall tests—but such a testing schedule is very unusual.

Our Contribution

Our study, then, is the first to look explicitly at summer learning in year-round schools. Using longitudinal data, we track students’ learning through the summer and through the nine months of the conventional academic year. Ours is also the first study to address Cooper et al.’s (2003) concern and account for the number of days that school has been in session at the time of each test. And we address McMillen’s (2001) concern by accounting for the nesting of children in schools using a multilevel model.

We try to equalize year-round and nine-month students on confounding variables, and a special feature of the data lets us show that our equalization is more or less successful. Specifically, the data let us estimate children’s achievement levels in the summer before kindergarten—before school calendars have had a chance to affect learning. Before kindergarten there is a significant gap between the test scores of children bound for year-round and nine-month schools, and this gap suggests that year-round students have disadvantages in their out-of-school environments. But we can more or less eliminate this pre-kindergarten gap using a small set of school- and child-level control variables. Our ability to explain the pre-kindergarten gap provides some assurance that we have controlled adequately for preexisting differences between children in year-round and nine-month schools, so that are comparisons are not biased by out-of-school differences over which the schools have no control.
What kinds of schools use year-round calendars?

We begin by comparing the characteristics of nine-month and year-round calendars. This comparison sheds some light on what kinds of schools use year-round calendars, and why those calendars were adopted. The comparison also provides us with a list of differences to control for when we estimate the effect of year-round calendars on learning.

Data

To make this comparison, we use the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)—a national survey carried out under the direction of the National Center for Education Statistics, US Department of Education (http://nces.ed.gov/ecls/Kindergarten.asp). During kindergarten and first grade, the ECLS-K gave reading and math tests to 17,030 children enrolled in 992 public and private elementary schools.\(^8\)

There are a couple of ways to check whether a school surveyed by the ECLS-K follows a year-round calendar. First, for nine-tenths of schools, the ECLS-K recorded a variable (called F4YrRnd) that explicitly indicated whether the school is year-round or not. In addition, for two-thirds of schools, the survey provides the first and last day of kindergarten, which can be used to calculate the length of the school’s kindergarten year.\(^9\)

---

\(^8\) The number 992 refers to the schools where children began kindergarten. Some students later transferred out of their original schools, and the survey followed some of these students to their new schools. Our analyses exclude the schools that these children transferred to, and also exclude any tests that a child took after a transfer. In addition, since our focus is on the effect of summer vacation, we exclude any tests taken after a child attended summer school. After post-transfer and post-summer-school test scores were eliminated, there remained 17,030 children who contributed at least one test score—this is the number reported in the text above. Reassuringly, the results are not materially different if post-transfer and post-summer-school test scores are included and attributed to the school where the child began kindergarten.

\(^9\) Dates for the first and last day of first grade are available from the public release of the ECLS-K, but these dates have many missing values. Dates for kindergarten are more complete, but only available to researchers who hold a restricted-data license.
For the most part, these two sources of information agree, but there are a few discrepancies that highlight coding errors and ambiguities. The length of the kindergarten school year has a cleanly bimodal distribution, and we would expect that schools near the primary mode of 9.3 months\textsuperscript{10} would be classified by the survey as traditional, while schools near the secondary mode at 11.6 months would be classified as year-round. Only three schools violate this expectation\textsuperscript{11}. In addition, there is a gray area between 10.2 and 10.6 months, where three schools are classified as year-round and three are not. It seems desirable to eliminate this overlap, since there is no point in making distinctions between nominally “year-round” schools and nominally “traditional” schools whose school calendars are roughly the same.

To reconcile these discrepancies, we classified a school as year-round if and only if its kindergarten year was at least eleven months. Eleven months is an arbitrary cutoff, but changing the cutoff by a few weeks would not change the classification of schools, since the distribution of school-year length has a clean break with no schools at all between 10.65 and 11.05 months. Among schools that did not report the length of their school year, we classified a school as year-round only if it was classified as such by the survey.

Under this definition, 27 of the 748 public schools in the ECLS-K were classified as year-round, which is about the number we would expect since, as we calculated earlier, 3% of all US public elementary schools run year-round. Under a less conservative

\textsuperscript{10} Throughout this paper, we measure time using a “standard month” that is exactly 365/12=30.4 days long (or 366/12 days long in a leap year). This seems preferable to using a calendar month, which can range in length from 28 to 31 days.

\textsuperscript{11} Two schools with kindergarten years of 9.2 and 9.7 months are classified as year-round, while one school with an 11.1-month kindergarten year is classified as traditional.
method for identifying year-round schools, up to 35 of the surveyed schools could be classified as year-round, but our results would not materially change.

Comparison

Table 3.1, Table 3.2, and Table 3.3 provide a detailed comparison of year-round and nine-month schools. Missing values of the tabled variables were filled in using multiple imputation (Rubin 1987); for details of the imputation model, see the Methodology chapter.

Table 3.1. School calendars.

<table>
<thead>
<tr>
<th></th>
<th>Nine-month (N=965)</th>
<th>Year-round (N=27)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten school days</td>
<td>178</td>
<td>176</td>
<td>-2.5* days (-4.4,-0.6)</td>
</tr>
<tr>
<td>First grade school days</td>
<td>179</td>
<td>177</td>
<td>-1.6* days (-3.0,-0.2)</td>
</tr>
<tr>
<td>Kindergarten begins</td>
<td>August 26, 1998</td>
<td>July 18, 1998</td>
<td>-1.30*** months (-1.49,-1.10)</td>
</tr>
<tr>
<td>Kindergarten ends</td>
<td>June 6, 1999</td>
<td>June 30, 1999</td>
<td>0.81*** months (0.64,0.98)</td>
</tr>
<tr>
<td>First grade begins</td>
<td>August 25, 1999</td>
<td>July 13, 1999</td>
<td>-1.43*** months (-1.64,-1.22)</td>
</tr>
<tr>
<td>First grade ends</td>
<td>June 5, 2000</td>
<td>June 23, 2000</td>
<td>0.59*** months (0.40,0.79)</td>
</tr>
</tbody>
</table>
Table 3.2. School characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Nine-month (N=965)</th>
<th>Year-round (N=27)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td>25%</td>
<td>3%</td>
<td>-23%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-29%,-17%)</td>
</tr>
<tr>
<td>Urban</td>
<td>39%</td>
<td>50%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-10%,32%)</td>
</tr>
<tr>
<td>Suburban</td>
<td>36%</td>
<td>48%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-09%,33%)</td>
</tr>
<tr>
<td>Western census region</td>
<td>23%</td>
<td>84%</td>
<td>62%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(46%,77%)</td>
</tr>
<tr>
<td>Public</td>
<td>69%</td>
<td>100%</td>
<td>31%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(27%,34%)</td>
</tr>
<tr>
<td>Percent free lunch</td>
<td>7%</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-3%,3%)</td>
</tr>
<tr>
<td>Percent reduced-price lunch</td>
<td>28%</td>
<td>43%</td>
<td>15%*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2%,27%)</td>
</tr>
<tr>
<td>Percent minority</td>
<td>36%</td>
<td>67%</td>
<td>31%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19%,44%)</td>
</tr>
<tr>
<td>Half-day kindergarten</td>
<td>44%</td>
<td>86%</td>
<td>42%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(29%,55%)</td>
</tr>
<tr>
<td>Crowded (1–5)</td>
<td>2.5</td>
<td>3.4</td>
<td>0.9**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.4,1.5)</td>
</tr>
</tbody>
</table>

Table 3.3. Child characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Nine-month</th>
<th>Year-round</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>16%</td>
<td>53%</td>
<td>36%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(24%,49%)</td>
</tr>
<tr>
<td>Asian</td>
<td>6%</td>
<td>11%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-1%,12%)</td>
</tr>
<tr>
<td>Black</td>
<td>15%</td>
<td>5%</td>
<td>-9%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-13%,-5%)</td>
</tr>
<tr>
<td>White</td>
<td>58%</td>
<td>27%</td>
<td>-31%***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-43%,-2%)</td>
</tr>
<tr>
<td>Mixed race</td>
<td>2%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2%,2%)</td>
</tr>
<tr>
<td>Native American/Pacific Islander</td>
<td>3%</td>
<td>2%</td>
<td>-1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-3%,0%)</td>
</tr>
<tr>
<td>SES (standardized)</td>
<td>0.01</td>
<td>-0.44</td>
<td>-0.44**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-0.72,-0.16)</td>
</tr>
<tr>
<td>Age on July 18, 1998</td>
<td>65.29 months</td>
<td>63.81 months</td>
<td>-1.47*** months</td>
</tr>
<tr>
<td>(average start of year-round kindergarten)</td>
<td></td>
<td></td>
<td>(-2.09,-0.86)</td>
</tr>
<tr>
<td>Age on August 26, 1998</td>
<td>66.57 months</td>
<td>65.09 months</td>
<td>-1.47*** months</td>
</tr>
<tr>
<td>(average start of nine-month kindergarten)</td>
<td></td>
<td></td>
<td>(-2.09,-0.86)</td>
</tr>
</tbody>
</table>

†p<.10, *p<.05, **p<.01, ***p<.001. Parentheses enclose 95% confidence intervals. Comparisons account for the clustering of children within schools.
Year-round schools do not schedule any more school days than nine-month schools—both calendar types typically have between 175 and 180 days of instruction—but year-round schools spread those days out over a larger part of the year. While nine-month schools typically begin in late August and end in early June, year-round schools typically start in mid-July and end in late June (Table 3.1).

Among the 27 year-round schools in the survey, only one is in a rural area (Table 3.2). This is not surprising, since the traditional academic calendar was designed in part to let rural children work on farms during the summer (Worsnop 1996). Although year-round schools are distinctly non-rural, they are not concentrated in urban areas. Instead, year-round calendars are about common in the suburbs as they are in cities.

Four-fifths of year-round schools are in the Western census region (Table 3.2), and half are in the state of California alone (National Association of Year-Round Education 2005). Year-round calendars may be popular in California because they got an early start there. California is an early adopter of policy innovations (Rogers 2003), and school reforms are no exception. California was the second state to pass a charter-school law (Linda A. Renzulli and Roscigno 2005); it has the country’s most ambitious class-size reduction program (Stecher et al. 2001); and it was a pioneer in year-round education as well. In 1971 two districts in San Diego County were among the first districts to adopt a year-round calendar; by 1974 thirteen other California districts had followed suit; and today there are 164 California districts with year-round schools (National Association of Year-Round Education 2005). It may be that, like the charter-school movement, the calendar-reform movement has spread by geographic contagion (Linda A. Renzulli and Roscigno 2005). That is, once one district or state tries an innovation such as a year-
round calendar, that innovation starts to look more promising to its neighbors, even if it is not yet evident whether the innovation is effective. The idea that a neighbor’s adoption of a practice can make that practice more appealing is known as “observability” in the diffusion-of-innovations literature (Rogers 2003), and as “social proof” in the psychology of persuasion (Cialdini 1985). Indirect evidence for geographic diffusion can be seen in the states around California: not only is California the number one state in its number of year-round schools, but its eastern neighbors, Arizona and Nevada, are number two and number five (National Association of Year-Round Education 2005). On the other hand, California’s northern neighbor, Oregon, is not a leader in year-round education; the entire state of Oregon has only three year-round schools (National Association of Year-Round Education 2005).

Although year-round calendars are popular in the West, in other parts of the country resistance to year-round schools is institutionalized. In seven Southern states, groups have sprung up with names such as Save Alabama Summers and Texans for a Traditional School Year (http://schoolyear.info/stcoalition.html). Some of these groups get funding and support from the American [Summer] Camp Association and the International Association of Amusement Parks and Attractions (Cumming 1993; MacFarquhar 1995; Chaker 2005), which are understandably reluctant to lose summer customers. In 2006, Texans for a Traditional School Year cheered the passage of a state law forbidding schools from starting before the fourth Monday in August (79th Texas Legislature, 3rd called session, House Bill 1). With this law Texas joined six other

---

12 This pattern is not an artifact of differences in state populations. Although California is the most populous state, its share of the U.S. population (12%) is much smaller than its share of year-round schools (50%). Similarly, although Arizona and Nevada rank second and fifth in their share of year-round schools, they rank only 19th and 35th in population.
Southern and Midwestern states with laws requiring that all public schools start on the same day. Four of these states—North Carolina and Texas in the South, Wisconsin and Minnesota in the Midwest—require schools to start no earlier than the last week of August, a start date that is most compatible with a traditional calendar (Janofsky 2005).

Children attending year-round schools tend to be poor, but their poverty is not extreme. Compared to nine-month schools, year-round schools have a 15% higher rate of poor students who eat reduced-price lunches—but the percentage of very poor students who eat free lunches is no higher at year-round schools than it is elsewhere (Table 3.3). Similarly, on a measure of socioeconomic status (SES) constructed by the ECLS-K—a standardized composite of household income, parents’ education, and parents’ occupational status—children at year-round schools typically have lower-than-average SES, but the SES gap between children at year-round schools and nine-month schools is only about half of a standard deviation (Table 3.3). Nearly all year-round schools are public rather than private; none of the ECLS-K’s 244 private schools were classified as year-round, which is not surprising since only 1 in 700 private schools uses a year-round calendar (Naylor 1995).

Year-round schools are disproportionately likely to suffer from crowding (Table 3.2). On a 1-to-5 measure of crowding, year-round principals give their schools an average score of 3.4, compared to an average score of 2.5 in nine-month schools. This difference is not at all surprising, since the multi-track variant of the year-round calendars is a strategy to handle crowding by rotating different parts of the student body in and out of the school at different times of the year. Crowding is one reason why year-round

---

13 The ECLS-K measure of SES is a sum of standardized variables, but the sum itself is not standardized. To make interpretation easier, we have standardized the sum.
schools are popular in California (Cooper et al. 2003; Ready, Valerie E. Lee, and Welner 2004). Crowding may also be the reason why year-round schools are twice as likely as other schools to offer half-day rather than full-day kindergarten. Half-day kindergarten can also be an answer to crowding if half a school’s students attend in the morning and half attend in the afternoon.¹⁴

It is not surprising that year-round calendars are more common in disadvantaged communities than they are elsewhere. The problems that the year-round calendar is supposed to address—both overcrowding and summer setback—are most severe in poor schools. In addition, disadvantaged schools may be more likely to use nearly any educational reform. Schools serving poor children are widely, though often unfairly, viewed as failing (Downey et al. 2008), and this label makes high-poverty schools targets for educational reform. Schools that are perceived as failing may be more receptive to change, and if they are not receptive, they have few resources and little credibility to resist reforms imposed by outsiders.

When year-round calendars are proposed in middle-class school districts, they run into organized resistance. Newspaper reports suggest that the parents leading the opposition to year-round schools are from the upper middle class. A leader of the Florida resistance, for example, is described as “a non-practicing lawyer and mother of two,” while a co-founder of Save Georgia Summers evidently has an affluent parent’s flexible schedule and access to sympathetic doctors; she reports spending a weekday “in the allergist’s office to get a note” excusing her child from hot early-August bus rides—rides necessitated by the early start of the year-round calendar (Janofsky 2005).

¹⁴ Alternatively, we might suspect that half-day kindergarten is a sign of disadvantage; that is, we might think that schools with low resource levels can stay open only half the day. But this explanation contradicts the evidence; in fact, schools with half-day kindergarten are more likely to be affluent than disadvantaged.
Poor parents, by contrast, tend to be less politically mobilized (Oberschall 1972; Jenkins and Wallace 1995; Verba, Schlozman, and Brady 1995). And given their low purchasing power, poor families may be less likely to attract support from business interests such as the theme-park and summer-camp industries. So when a year-round calendar is proposed in a poor community, it has a better chance of taking root.

With year-round schools concentrated in poor Western cities and suburbs, it is not surprising that about half the students in year-round schools are Hispanic. In fact, Hispanic children are three times as common in year-round schools as elsewhere (Table 3.3). We might also expect year-round schools to have a large number of African-American students, but actually black children are three times less common in year-round schools than they are in other schools. The lack of black students in year-round schools probably has something to do with geography. Although Hispanic and African-American children both tend to be poor, nearly half of Hispanics live in the Western states, where year-round schools are popular, while most African-Americans live in the South (U.S. Bureau of the Census 2002), where the resistance to year-round schools is strongest.

A final sign of disadvantage is that the children who attend year-round schools are younger than the children who attend nine-month schools. Disadvantaged parents are less likely to hold their children back from entering kindergarten, perhaps because poor parents have fewer resources to care for children outside of school (Downey and Hickman 2003). On any given day, children in a year-round school are six weeks younger, at the median, than children in the same grade at a nine-month school. In addition, the year-round school year starts, on average, five weeks before the nine-month school year (Table 3.3). This means that year-round students are eleven weeks younger
when their kindergarten starts than nine-month students are when their kindergarten starts. Younger children tend to have lower test scores than older children, so, in our models of achievement and learning, we will account for differences in age as well as test dates and school calendars.

Learning Rates in Nine-Month and Year-Round Schools

To sum up, then, year-round calendars are most common in the crowded public schools of Western cities and suburbs, and are mainly attended by moderately poor Hispanic children. We will account for these characteristics of year-round schools as we evaluate the impact of year-round calendars on children’s learning rates.

Data: Test Scores and Test Dates

A major asset of the ECLS-K is that its testing schedule allows us to estimate how quickly children learn during the summer as well as during the nine-months of the traditional academic year. The ECLS-K provides this information by measuring reading and math skills four times in the first two years—in the fall and spring of kindergarten 1998-99 and in the fall and spring of first grade 1999-2000. By comparing the spring and fall test scores, we can estimate the amount learned during kindergarten, summer, and first grade. The ECLS-K continued to follow children through fifth grade, but later test occasions were spaced two years apart and so do not permit separate estimates of summer and non-summer learning.

Excellent tests of reading and mathematics skill were designed especially for the ECLS-K (Rock and Pollack 2002). The tests tried to measure the full range of young children’s abilities—from rudimentary skills, such as recognizing isolated letters and numbers, to advanced skills, such as reading words in context and solving multiplication...
and division problems. Tests were administered in two stages: first students took a routing test to determine their general skill level (low, medium or high); then they took a more specialized test to determine more precisely where they fell among students of generally comparable skill. This two-stage approach helped to explore the full range of children’s abilities, avoiding ceiling and floor effects, while keeping the number of questions students had to answer reasonably small. After testing, children’s level of reading and math skill was estimated using item response theory (IRT), which is influenced not just by the number of questions answered correctly and incorrectly, but also by the difficulty, discrimination, and guessability of those questions.

In the ECLS-K data, test results are recorded using different measures. The public release of the data includes a scale score that estimates how many questions each child would have answered correctly had they been asked a list of 92 questions in reading and 64 questions in math. The restricted data includes a theta score which is approximately linear with the log odds of answering a question correctly. Our analyses use the theta score because it is more nearly an interval measure of children’s skills (Reardon 2008). The analyses reported here assumed that the theta score is 90% reliable—that is, that measurement error accounts for 10% of the total variance in theta scores. This estimate is consistent with psychometric estimates (Rock and Judith M. Pollack 2002), but reliability could be lower if estimated by other means (Reardon 2008). To explore this possibility, we replicated our analyses under the assumption that the tests were only 70 or 80 reliable; our results were not materially affected. We also replicated the analyses using the scale score instead of the theta score, and again we obtained very similar results. The principal
difference between the scale score and the theta score is that, since the scale score is not as close to an interval measure, children’s growth on the scale score is not as smooth.

Although the results for reading and math were very similar, we have more confidence in the math results because the reading tests were not given to students who lacked proficiency in English. This is a particular problem for a study of year-round schools because more than half of the students in year-round schools are Hispanic.

Not every child was tested on every occasion, and the number of test scores was especially low in the fall of first grade, when only a random 30% subsample of schools was visited for testing (Table 3.4). The undersampling of schools in the fall of first grade is unfortunate, since the fall of first grade test is crucial for estimating the summer and first grade learning rates. Out of the 27 year-round schools in the ECLS-K, only 7 were tested in the fall of first grade. Even though our power to estimate summer learning is limited, our summer-learning estimates are still valuable since no prior study of year-round schools has been able to look at summer learning at all. In an understudied area, even a limited sample can provide considerable insight.

Average scores for each of the four test occasions are given in (Table 3.4). On every test occasion, the average score was lower—sometimes significantly lower—in year-round schools than in nine-month schools. But comparing scores between year-round and nine-month schools requires more than a comparison of means. Not only do nine-month and year-round schools differ in their institutional and demographic characteristics, but there are also some tricky issues related to the timing of school years and tests.
Table 3.4. Test dates and test scores. Means.

<table>
<thead>
<tr>
<th>Kindergarten</th>
<th>First grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Fall</td>
</tr>
<tr>
<td>Year-round</td>
<td>October 11, 1998</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.65*** months</td>
</tr>
<tr>
<td></td>
<td>(-0.90,-0.41)</td>
</tr>
<tr>
<td>Schools visited</td>
<td>Nine-month</td>
</tr>
<tr>
<td></td>
<td>Year-round</td>
</tr>
<tr>
<td>Math Scores (theta)</td>
<td>Nine-month</td>
</tr>
<tr>
<td></td>
<td>Year-round</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Scores (theta)</td>
<td>Nine-month</td>
</tr>
<tr>
<td></td>
<td>Year-round</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Children tested</td>
<td>Nine-month</td>
</tr>
<tr>
<td></td>
<td>Year-round</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Children tested</td>
<td>Year-round</td>
</tr>
</tbody>
</table>

*p<.10, **p<.05, ***p<.01, ****p<.001. Parentheses enclose 95% confidence intervals. Comparisons account for the clustering of children within schools. Missing values (except for test scores) filled in by multiple imputation.

First, each school took tests on a different date. In the fall of kindergarten, for example, 90% of children took the tests in October or November, but 10% of children took the tests as early as September or as late as December. The difference in school calendars was confounded with the difference in test dates; in the fall of kindergarten, children in year-round schools took their tests an average of three weeks before children in nine-month schools (Table 3.4).

Even if all schools had been tested on the same dates, their scores would not be comparable because year-round schools start earlier and end later than nine month schools. On average, year-round schools start about five weeks before nine-month

---

15 Within schools, different children took the tests on different dates, but the within-school differences were trivial. Typically, children from the same school differed by no more than a day or two in their test dates.
schools (Table 3.1), which gives year-round schools a head start in learning material for their fall tests. Nine-month schools, by contrast, have an advantage on the spring tests, since nine-month students finish their 175–180 days of instruction two or three weeks earlier than year-round students (Table 3.1).

Even if each school were tested on the first and last day of its own school year, year-round and nine-month scores would still not be comparable, because the students in year-round and nine-month schools are not the same age. As we noticed earlier, on the first day of year-round kindergarten, children in year-round schools are eleven weeks younger, on average, than children in nine-month schools are on the first day of nine-month kindergarten (Table 3.3). As we will estimate later, a one-month difference in initial age is associated with about a 0.025 point difference in initial achievement, so net of confounding differences an eleven-week difference in age corresponds to about a 0.07 point difference in achievement on the theta scale.

To compensate for these differences, our models of achievement and learning will include controls for students’ age and for the days elapsed between each test date and the first and last day of each school year.

**Basic model**

To compare learning rates across nine-month and year-round schools, we fit a multilevel growth model (Raudenbush and Bryk 2002, chapter 6) to the test dates and test scores summarized in Table 3.4. Our model answered Cooper et al.’s (2003) concern and “explicitly controlled for the number of days that school had been in session before achievement outcomes were measured.” That is, the model made adjustments for the difference between the test dates in Table 3.4 and the first and last dates of the school
years in Table 3.1. In a sense, the model extrapolated back from the test dates to estimate the scores that would have been received had the tests been given on the first and last day of the school year. For a detailed specification of our multilevel growth model, see the methodology chapter.

The model estimated monthly learning rates during the periods defined as nine-month kindergarten, summer, and nine-month first grade. In a nine-month school, we define the boundaries of these periods using the start and end of the kindergarten and first grade school year, which vary a little from one school to another. In year-round schools, we use the start and end dates that are average for the nine month school year (Table 3.1). In addition to estimating learning rates during each period, we combine these estimates to derive the following quantities:

1) the school-year learning rate, defined as the average learning rate across nine-month kindergarten and nine-month first grade;

2) summer slowdown, defined as the difference between the summer learning rate and the school-year learning rate;

3) twelve-month gains, defined as the total learned over one of two twelve-month periods.
   a. One period consists of nine-month kindergarten and the following summer.
   b. The other period consists of nine-month first grade and the preceding summer.

The total learned in each period is estimated by multiplying the school year’s monthly learning rate by 9.4 months, then adding the summer’s
monthly learning rate multiplied by 2.6 months. These figures—9.4 months and 2.6 months—are the average lengths of the school year and the summer vacation in a nine-month school.

Taken together, all these quantities let us ask the questions most relevant to our research interests. Do year-round students learn faster during the summer? Do they learn more slowly during the nine-month school year? Is year-round students’ learning smoother across the year, with less summer slowdown? Over a full twelve months, do year round students learn more or less than students on a traditional calendar?

Figure 3.1 plots reading and mathematics growth on an average nine-month calendar, with each school year starting in late August and ending in early June. It is evident from the plots that on a year-round calendar the distribution of learning, like the distribution of school days, is much more even across the year. In nine-month schools, learning slows or stops during the summer, whereas in a year-round school the summer barely changes the rate of learning at all. This is mainly because summer learning is faster in a year-round school, but also because school-year learning is slower.

Table 3.5 gives estimates obtained by fitting the basic model. The estimates show that, during summer, year-round students learn both reading and mathematics significantly faster than nine-month students. But during the nine-month school year, year-round students learn more slowly than nine-month students, at least in reading (the difference in mathematics verges on significance, at p=.08). Across seasons, year-round students learn at a more even rate, with significantly less summer slowdown. It is ambiguous whether year-round students learn more than other students over a full twelve month period. In reading, there is no significant difference between the twelve-month
gains of year-round and nine-month students. In mathematics, year-round students have
greater twelve-month gains than nine-month students, but the difference is only
significant (p=.046) during the twelve-month periods that includes kindergarten; the
difference is not significant during the twelve-month period that includes first grade.

Figure 3.1 plots reading and mathematics growth on an average nine-month
calendar, with each school year starting in late August and ending in early June. It is
evident from the plots that on a year-round calendar the distribution of learning, like the
distribution of school days, is much more even across the year. In nine-month schools,
learning slows or stops during the summer, whereas in a year-round school the summer
barely changes the rate of learning at all. This is mainly because summer learning is
faster in a year-round school, but also because school-year learning is slower.
Table 3.5. Seasonal learning rates in nine-month and year-round schools.

### Mathematics

<table>
<thead>
<tr>
<th></th>
<th>Achievement</th>
<th>Monthly learning</th>
<th>Summertime slowdown</th>
<th>12-month gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start of 9-month kindergarten (8/26/98)</td>
<td>9-month kindergarten Summer</td>
<td>9-month first grade</td>
<td>Avg of 2 school years</td>
</tr>
<tr>
<td>Nine-month</td>
<td>-1.289***</td>
<td>0.086***</td>
<td>0.026***</td>
<td>0.077***</td>
</tr>
<tr>
<td>Year-round</td>
<td>-1.428***</td>
<td>0.083***</td>
<td>0.066***</td>
<td>0.071***</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.139*</td>
<td>-0.003</td>
<td>0.041**</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

### Reading

<table>
<thead>
<tr>
<th></th>
<th>Achievement</th>
<th>Monthly learning</th>
<th>Summertime slowdown</th>
<th>12-month gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start of 9-month kindergarten (8/26/98)</td>
<td>9-month kindergarten Summer</td>
<td>9-month first grade</td>
<td>Avg of 2 school years</td>
</tr>
<tr>
<td>Nine-month</td>
<td>-1.386***</td>
<td>0.095***</td>
<td>0.003</td>
<td>0.091***</td>
</tr>
<tr>
<td>Year-round</td>
<td>-1.465***</td>
<td>0.088***</td>
<td>0.054**</td>
<td>0.072***</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.079</td>
<td>-0.006</td>
<td>0.051**</td>
<td>-0.019*</td>
</tr>
</tbody>
</table>

\[^p<.10, \*p<.05, \**p<.01, \***p<.001.\]
Figure 3.1. Average learning in nine-month and year-round schools.
The lower panel of the mathematics and reading results in Figure 3.1 shows how the achievement gap between year-round and nine-month students waxes and wanes across the year. During the summer, when nine-month students are out of session, year-round students close the gap rapidly, but during the other nine months, when year-round students have more frequent breaks than nine-month students, the gap slowly widens again. This pattern is wholly consistent with the view that learning is fastest when school is in session.

*Fuller model*

Although the results of our basic analyses are suggestive, they are also tentative because they do not adjust for confounding differences between the students in year-round and nine-month schools. The simplest way to see that these differences are important is to look at Figure 3.1’s estimate of achievement on the first day of nine-month kindergarten. On that day, children who are about to start nine-month kindergarten have significantly higher mathematics achievement than children who have already been attending year-round kindergarten for over a month. This initial difference tells us nothing about the merits of nine-month kindergarten, which has yet to begin. What it tells us is that the children who attend nine-month kindergarten have out-of-school advantages that are evident before the first day of school; in addition, nine-month students are a little older than their year-round peers.

To make a fairer comparison, Table 3.6 fits an augmented model that includes controls for the important demographic and institutional differences between nine-month and year-round schools and students. At the school level, this augmented model controls
for crowding, for school sector (public vs. private), and for geography (Western vs. non-Western; rural vs. suburban vs. urban). At the child level, the model controls for ethnicity (white, black, Hispanic, Asian, other) and for SES, which the ECLS-K defines as a composite of parental education, parents’ occupational status, and household income.

In addition to these demographic and institutional controls, the augmented model includes two variables that affect children’s exposure to their school and non-school environments. At the school level, the model controls for the difference between half- and full-day kindergarten, while at the child level, the model controls for age at the start of nine-month kindergarten. In an alternate version of the model, we also controlled for summer school attendance, but in the version displayed in Table 3.6 we simply omitted any tests that were taken after a child attended summer school. The results of these different approaches were about the same.

Variables were coded so that the reference group contained the types of children and schools that are most likely to follow a year-round calendar—that is, the reference group consisted of younger-than-average Hispanics with lower-than-average SES, who attended crowded public schools with half-day kindergarten programs in Western cities. We highlighted this reference group by centering continuous variables (crowding, SES, age) around the mean values for year-round schools, and by coding dummy variables so that the reference categories were Hispanic children, public schools, urban schools, Western schools, and schools offering half-day kindergarten.
Table 3.6. Adjusted learning rates in nine-month and year-round schools.

Math

<table>
<thead>
<tr>
<th></th>
<th>Initial achievement</th>
<th>Monthly learning</th>
<th>Summer slowdown</th>
<th>12-month gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start of year-round</td>
<td>Start of 9-month</td>
<td>9-month</td>
<td>Avg of two</td>
</tr>
<tr>
<td></td>
<td>kindergarten</td>
<td>kindergarten</td>
<td>kindergarten</td>
<td>school</td>
</tr>
<tr>
<td></td>
<td>(7/18/98)</td>
<td>(8/26/98)</td>
<td>Summer</td>
<td>9-month grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1st school</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>years</td>
</tr>
<tr>
<td>Nine-month</td>
<td>-1.597***</td>
<td>-1.563***</td>
<td>0.069***</td>
<td>0.037***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.085***</td>
<td>0.063***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year-round</td>
<td>-1.640***</td>
<td>-1.532***</td>
<td>0.085***</td>
<td>0.063***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>-0.044</td>
<td>0.031</td>
<td>-0.005</td>
<td>0.027*</td>
</tr>
</tbody>
</table>

Controlled differences

School level

Crowded (1-5)   -0.006   0.000   0.000   0.000
Full-day kindergarten  0.000  0.007*** -0.008 -0.004^*
Private        0.120*** -0.004* -0.003 -0.005**
Non-western    -0.007   -0.001 -0.010   0.006
Rural          -0.048*   0.004*  0.004 -0.003
Suburban       0.034**   0.000   0.004 -0.001

Child level

Ethnicity (Hispanic reference)

Asian          0.215*** -0.006*** 0.012 -0.010***
Black          0.044**  -0.010*** 0.001   -0.003
White          0.191*** -0.005** -0.002 -0.003
Other non-Hispanic 0.062*** -0.001 -0.011 -0.002
SES (standardized) 0.167*** -0.002*** 0.004 -0.003**
Age (months), 8/25/1998  0.026*** -0.001*** 0.000 -0.001*

Interactions

SES x year-round  0.012 -0.004*  -0.014   0.005
Age x year-round  0.009*    0.000   -0.001  0.000
SES x age         0.004***  0.000*  -0.001  0.000

Reading

<table>
<thead>
<tr>
<th></th>
<th>Initial achievement</th>
<th>Monthly learning</th>
<th>Summer slowdown</th>
<th>12-month gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start of year-round</td>
<td>Start of 9-month</td>
<td>9-month</td>
<td>Avg of two</td>
</tr>
<tr>
<td></td>
<td>kindergarten</td>
<td>kindergarten</td>
<td>kindergarten</td>
<td>school</td>
</tr>
<tr>
<td></td>
<td>(7/18/98)</td>
<td>(8/26/98)</td>
<td>Summer</td>
<td>9-month grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1st school</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>years</td>
</tr>
<tr>
<td>Nine-month</td>
<td>-1.681***</td>
<td>-1.656***</td>
<td>0.101***</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.093***</td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year-round</td>
<td>-1.707***</td>
<td>-1.589***</td>
<td>0.093***</td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>-0.026</td>
<td>0.066*</td>
<td>-0.008*</td>
<td>0.047**</td>
</tr>
</tbody>
</table>

Controlled differences

School level

Crowded (1-5)   -0.004   -0.001   0.000   0.000
Full-day kindergarten  0.023*    0.009***  -0.014* -0.008***
Private        0.119***  -0.005*    0.000   -0.001
Non-western    -0.009   -0.004*   0.001   0.004*
Rural          -0.059***  0.003    0.010*   -0.002
Suburban       0.048**    -0.001   0.000   0.001

Child level

Ethnicity (Hispanic reference)

Asian          0.180***  -0.002   0.004   -0.004
Black          0.075***  -0.010*** 0.003   -0.001
White          0.154***  -0.006*** -0.003   -0.002
Other non-Hispanic 0.058*    -0.001   -0.011   0.002
SES (standardized) 0.177***  -0.003*** 0.005* -0.003***
Age (months), 8/25/1998  0.020*** -0.001*** 0.000  -0.001***

Interactions

SES x year-round  0.000 -0.002  -0.026*  0.011*
Age x year-round  0.006   0.000   0.002   -0.001
SES x age        0.004***  0.000*   0.000   0.000*

*p<.10, *p<.05, **p<.01, ***p<.001. Parentheses enclose 95% confidence intervals.
Figure 3.2. Average learning in year-round and nine-month schools, adjusted for other child- and school-level differences.
Table 3.6 and Figure 3.2 display the results of fitting this augmented model. The effects of the control variables are generally consistent with past research. We will begin by describing the effects of the control variables on children’s initial achievement levels. Compared to children in the reference group, older children, children of higher SES, and children in private or suburban schools all had high initial achievement scores, while children in rural schools scored relatively low. Every non-Hispanic ethnic group—including blacks, whites, Asian-Americans, and others—had higher initial test scores than Hispanic children.

From a seasonal perspective, what is especially interesting is that nearly all the variables that predict high initial test scores also predict low (not high) learning rates during the kindergarten school year, and sometimes during the first grade school year as well. This pattern is consistent with the view that schools compensate for initial inequalities. Compared to the reference group, for example, children of high socioeconomic status have higher initial achievement and higher summer learning rates (at least in reading), but they also have lower learning rates during both kindergarten and first grade. This is a very strong seasonal pattern suggesting that high socioeconomic status is only an advantage during out-of-school periods, and is actually a disadvantage when school is in session. Similar though weaker patterns can be found for the effects of being Asian, white, or black (instead of Hispanic); for the effect of being older than the median age at the start of kindergarten; and for the effect of attending a private rather than a public school. All of these advantages are associated with having significantly higher initial achievement, but all of them are also associated with having significantly lower learning rates, in either reading or math, during either the kindergarten or the first
grade school year. This makes sense if the early years of school are compensatory in function—that is, if a major purpose of early schooling is to help disadvantaged children to catch up (Alexander and Entwisle 1996; Downey et al. 2004).

The effects of the remaining institutional variables are straightforward. Crowding has no effect on initial achievement or on learning during either the school year or the summer. Attending full-day rather than half-day kindergarten increases learning rates during the kindergarten school year, but it appears that some of the gains from full-day kindergarten may be given back during the summer and the first-grade school year (cf. Cooper et al. 2010).

We could fit this model with a different set of control variables—and we have—but the key question is whether any set of variables can adequately compensate for preexisting differences between the children who attend nine-month and year-round schools. We can address this question by examining estimates of initial achievement on the first day of nine-month kindergarten. In the basic model, before we added the control variables (Figure 3.1 plots reading and mathematics growth on an average nine-month calendar, with each school year starting in late August and ending in early June. It is evident from the plots that on a year-round calendar the distribution of learning, like the distribution of school days, is much more even across the year. In nine-month schools, learning slows or stops during the summer, whereas in a year-round school the summer barely changes the rate of learning at all. This is mainly because summer learning is faster in a year-round school, but also because school-year learning is slower.

Table 3.5), we found that children in nine-month schools had a significant lead on the first day of nine-month kindergarten, at least in mathematics; in reading, the lead of
nine-month children was about half as large and not statistically significant. But when we incorporate the control variables (Table 3.6), the initial mathematics advantage of nine-month students shrinks by 75% and becomes non-significant, while the reading gap actually flips direction, so that year-round students have a small and almost significant lead (p=.09) on the first day of nine-month kindergarten. These results suggest that the control variables have erased the pre-kindergarten advantage of year-round students.

Figure 3.1 plots reading and mathematics growth on an average nine-month calendar, with each school year starting in late August and ending in early June. It is evident from the plots that on a year-round calendar the distribution of learning, like the distribution of school days, is much more even across the year. In nine-month schools, learning slows or stops during the summer, whereas in a year-round school the summer barely changes the rate of learning at all. This is mainly because summer learning is faster in a year-round school, but also because school-year learning is slower.

Given the out-of-school disadvantages of year-round students, it may at first seem surprising that they could have a lead at the start of kindergarten. But we should remember that we are talking about the start of kindergarten in nine-month schools. By the time nine-month schools started kindergarten, on an average date of August 26, 1998, year-round kindergarten had already been open for an average of six weeks (Table 3.1). If other differences are controlled, it makes sense that these six weeks would give year-round students a small lead.

To find a time when neither nine-month nor year-round students have been exposed to schools, we have to turn the calendar back a bit further, to the first day of year-round kindergarten, on an average date of July 18, 1998. We can estimate
achievement on this date by extrapolating backward under an assumption about the learning rate between the start of year-round kindergarten on July 18 and the start of nine-month kindergarten on August 26. Since year-round students were in kindergarten for this period, we assume that their learning rate was simply the adjusted average for kindergarten learning among year-round students. Since nine-month students were not in school during the same period, we need an estimate of the rate at which nine-month students were learning in the month or so just before kindergarten began. And our model provides such an estimate: the pre-kindergarten learning rate is the effect of age in months on initial achievement—.026 points per month in mathematics, or .020 points per month in reading. (This is the pre-Kindergarten learning rate for children in nine-month schools. In year-round schools, the effect of age on achievement may be different, and is accounted for by including an interaction between student age and year-round school.)

By extrapolating these learning rates back to the first day of year-round kindergarten, we get estimates of initial achievement on July 18, 1998, before either year-round or nine-month students have had a day of school. And at that time, our estimates of initial, unschooled achievement do not differ significantly between year-round and nine-month students. This provides some assurance that the preexisting differences between year-round and nine-month students have been adequately controlled. It may be that the differences could be controlled more completely, but any improvement would be small and hard to demonstrate given the limited precision of the statistical estimates. In addition, with only 27 year-round schools in the sample, we would be uncomfortable including more than the half-dozen school-level control variables that are already present in the model.
In this fuller model, with its plausible controls for confounding differences (Table 3.6), the estimated effects of year-round school were very similar to what they were in the basic model (). Again, the results suggested that, compared to nine-month students, year-round students had faster average learning rates during summer, but slower average learning rates during the rest of the year. Figure 3.2 plots average learning rates in year-round and nine-month schools. Again the patterns suggest that during the summer, year-round schools pull ahead, but during the rest of the year, nine-month schools make up the lost ground. Aside from having a different start date—July 18 instead of August 26, 1998—Figure 3.2 looks very similar to Figure 3.1. The main difference is that Figure 3.2’s use of control variables has pulled the year-round and nine-month trajectories much closer together—so close that the two types of schools trade leads twice a year, with year-round students pulling ahead in the summer, and nine-month students pulling ahead by the end of the nine-month academic year.

Again, the crucial question is how much children learned over a twelve-month period. And now that we have extended our estimates back to the beginning of year-round kindergarten, we have three twelve-month periods to choose from:

1) the twelve months running from the first day of kindergarten to the first day of first grade in year-round schools;

2) the twelve months running from the first day of kindergarten to the first day of first grade in nine-month schools; or

3) the twelve months running from the last day of kindergarten to the last day of first grade in nine-month schools
Again, the results are ambiguous. In the first twelve-month period, the differences favor year-round schools, and border on significance in both mathematics (p=.06) and reading (p=.08). In the second twelve-month periods, the differences also favor year-round schools, but are smaller and not close to statistical significance. In the third twelve-month period, however, the differences favor *nine-month* schools, and are statistically significant in reading (p=.04), though not in mathematics. In sum, among our six twelve-month comparison comparisons—three in mathematics and three in reading—four comparisons favored year-round schools, and two favored nine-month schools. Three of the differences were nonsignificant; two bordered on significance and favored year-round schools; and one was outright significant and favored nine-month schools. Averaging across the six comparisons, we could estimate that annual gains were about 2.5% larger in year-round than in nine-month schools—a very small difference, 0.025 points on the theta scale, or about the amount learned in a week and a half of school. A glance at Figure 3.2 confirms the impression that there is not much to choose between the two calendars. Year-round and nine-month schools trade leads twice during each calendar year, and the leads that they build up are neither large nor sustained. Compared to the large difference that some year-round reformers propose to remedy—the difference between school-year and summer learning rates—the difference between learning in year-round and nine-month schools, if there is one, is barely noticeable.

*Do Year-Round Schools At Least Help the Disadvantaged?*

Past research has suggested that, even if year-round schools do not help most students, they can help students from disadvantaged families (McMillen 2001, Cooper et al. 2003). In a sense, we have already tested this prediction because, as it turns out, most
children in year-round schools are somewhat disadvantaged. The fact that we found little or no benefit for the typical student in a year-round school means that we found little or no benefit for children who are somewhat disadvantaged.

We can pursue the suggestion further, however, and test whether year-round schools are beneficial for very disadvantaged children—that is, for children who are disadvantaged even when compared to the typical year-round student. The possibility that year-round schools help the very disadvantaged seems plausible at first, because year-round calendars increase summer learning, and summer is when poor children do most of their falling behind (Alexander, Entwisle, and Olson 2007). We can see poor children’s summer setback in Table 3.6, where the gap between high- and low-SES children widens during the summer (at least in reading) and actually narrows during the kindergarten and first grade school years.

Since year-round schools replace most of summer vacation with regular school instruction, it makes sense that year-round schools would reduce summer learning gaps between children of high and low socioeconomic status (SES). During summer, year-round schools increase children’s exposure to school and reduce their exposure to their home environments—and it is at home that poor children experience their greatest disadvantages (Lareau 2000; Linver, Brooks-Dunn, and Kohen 1999, 2002).

But again, the summer benefits of year-round schools may come with a cost. As we pointed out earlier, year-round schools do not just add days of instruction to the summer; they also subtract days of instruction from the rest of the year. So although year-round schools reduce poor children’s time at home during summer, they increase home exposure during fall, winter, and spring.
On balance we might predict that, though year-round schools reduce the growth of achievement gaps during the summer, they accelerate the growth of achievement gaps during the school year. To test this prediction, we follow past research by including in our model an interaction between year-round schooling and child-level SES (cf. McMillen 2001, Cooper et al. 2003). If year-round schools reduce the summer growth in SES achievement gaps, we would expect the summer to reveal a *negative* interaction between SES and year-round schools—indicating that a *decrease* in SES is associated with an *increase* in the benefits of year-round schools. Conversely, during the regular kindergarten and first grade school years, we would expect a *positive* interaction between year-round schooling and SES.

The results, in the bottom lines of Table 3.6, are marginally consistent with these predictions. In reading, the SES X year-round interaction is, as predicted, positive and significant during the first grade nine-month school year, and negative and borderline significant (p<.10) during summer—but nonsignificant during nine-month kindergarten. In math, the interaction is nonsignificant during summer and nine-month first grade; during kindergarten the interaction borders on significance (p<.10), but the sign of the interaction is negative, contrary to expectations.

When an effect is positive in one season and negative in another, the fairest way to evaluate its net impact is over a full twelve months. And again, the twelve-month results are ambiguous. During the first two twelve-month periods, the SES X year-round interaction is negative in both reading and mathematics; both of the reading interactions are borderline significant (p<.10), and so is one of the mathematics interactions. During the third twelve-month period, however, the SES X year-round interaction is positive and
nonsignificant in both reading and mathematics. On balance, the results lean just slightly toward the view that year-round schools provide especial benefits to low-SES children, but the pattern is neither clear nor consistent enough to convince a skeptic.

Discussion

Our results fit the view that year-round schools do little to increase or decrease average achievement gains—instead, they just distribute gains more evenly across the year. The results suggest that neither the year-round nor the nine-month calendar has a clear advantage over the other. Any advantage of year-round schools, if it exists, is small, elusive, and may be confined to children whose socioeconomic status is low even for the year-round population. Compared to the problem that some year-round reformers hope to address—the slowdown in summer learning, especially for disadvantaged children—the year-round calendar is clearly an inadequate solution.

The data that we used to address this question are not perfect, though they are the best data available. Just 27 year-round schools participated in the ECLS-K survey, and just 7 took the first grade autumn test needed to estimate summer and first-grade learning rates. Because the sample of year-round schools was small, our estimates of first-grade and summer learning in year-round schools are imprecise, and some of this imprecision seeps into our estimates of twelve-month gains.

A further concern is that the ECLS-K did not give tests at the very beginning and end of the school year. Fall tests were typically given in mid-October, and we have to extrapolate backward to estimate the scores that would have been obtained at the start of the school year in late August or, for year-round schools, in mid-July. Likewise, spring tests were typically given in early May, and we have to extrapolate forward to estimate
the scores that would have been obtained at the end of the school year in June. These extrapolations rely on the assumption that learning is approximately linear during the school year—that is, that students learn about as fast in September and June as they do between October and May. Linearity may be a serviceable approximation, but it is only an approximation, and the true first- and last-day scores probably differ a bit from our estimates.

A final shortcoming of the ECLS-K is that it does not distinguish between the different forms of year-round calendar. Given the small sample of year-round schools in the ECLS-K, it is doubtful that we could learn much by making fine distinctions among calendar types, but it would be useful to know how many of the year-round schools offer supplementary instruction in the intersessions between regular periods of school. We suspect that few of the schools offer intersession instruction since the ECLS-K’s year-round schools are mostly in crowded Western districts where the year-round calendar typically follows a multi-track system. Multi-track schools, which account for 40% of year-round schools nationwide, have different groups of children on different calendars and have few slack resources for intersession instruction. Nevertheless, it seems likely that at least a couple of the year-round schools in our sample offer intersession instruction. If so, then that extra instruction may slightly bias our estimates in favor of the year-round calendar. A fairer comparison group for year-round schools with intersession instruction would consist of nine-month schools where a substantial fraction of the student body attends summer school.

Our findings are limited to elementary school, and more specifically to the kindergarten and first grade school years. But this limitation is less severe than it may
seem, since about three quarters of year-round students are in elementary rather than secondary schools (Cooper et al. 2004). All of the year-round schools in Las Vegas, and all but one of the year-round schools in Chicago, are at the elementary level (Clark County School District Year Round Calendar Study Group 2007; Ihejirika 2007). Middle schools and high schools often find it difficult to adopt the year-round calendar, at least in its multitrack form, because it is hard to ensure that all required courses and a reasonable sampling of electives are available to students on three or four different tracks (Dale Erquiaga16, personal communication, October 2007). An elementary school, by contrast, where children spend most of the day in a single classroom, is relatively easy to put on a year-round calendar.

It remains the case that the patterns in kindergarten and first grade could be different from those later on. However, if our explanation for our results is correct—if the amount learned depends on the amount of time spent in school, regardless of how that time is arranged—then we can hypothesize that similar results would be obtained for older students. Further research is needed to test this hypothesis.

Against the faults of the ECLS-K, we have to weigh the fact that no other data set can speak to the question of summer learning in year-round schools. The data from the ECLS-K are not ideal, but until someone collects data that sheds a brighter light on summer learning in year-round schools, the evidence from this study is valuable because it is unique.

16 Dale Erquiaga was the facilitator for the Clark County Public Schools Year Round Calendar Study Group (2007).
Conclusion

Do year-round schools accelerate summer learning? Yes, they do—but they also reduce learning during the rest of the year. Year-round calendars don’t schedule any more school days than nine-month calendars, so they can only add school days to summer by subtracting school days from fall, winter, and spring. Because learning increases and decreases with the number of scheduled school days, year-round schools do not really solve the problem of summer setback—they simply spread it out across the year.

Just as year-round schools take the long summer vacation and break it up into several shorter breaks, it may be that year-round schools replace the usual large summer setback with three or four smaller setbacks that happen during the year-round calendar’s three- and four-week vacations. Although we cannot observe these micro-setbacks directly, they would explain why the catching up that year-round children do during the summer is largely canceled by their falling behind during the rest of the year.

These results are consistent with the sociological view that summer setback comes from cognitive disadvantages in students’ homes and neighborhoods (Heyns 1978; Entwisle and Alexander 1992; Downey et al. 2004). Non-school disadvantages cannot be erased by a year-round calendar, because a year-round calendar does not reduce the time that children spend outside of school, nor does it increase the academic stimulation of children’s out-of-school activities.

Although our findings contradict a major argument for year-round calendars, the results need not be taken as a repudiation of year-round schools. Instead, the results are more or less neutral with regard to the academic effects of the year-round calendar. Over twelve-month periods, year-round students learn about the same amount as children on a
traditional nine-month calendar. On purely academic grounds, we cannot advocate a year-round calendar, but we cannot recommend against it, either. If a school has some non-academic reason to favor a year-round calendar—for example, to cope with overcrowding—it appears that the year-round calendar can be adopted at little or no academic cost. And if a school has already switched to a year-round calendar, we see no reason—at least, no academic reason—for the school to switch back. On the other hand, if a school is considering a year-round calendar solely in hope of boosting academic achievement, it seems unlikely that those hopes will be realized. Changing to a year-round calendar may not be worth the disruption that comes with the change.

Although the findings are somewhat discouraging with respect to the promise of year-round education, they are encouraging with respect to the validity of seasonal research and its implications.

First, the results indicate that summer learning can be increased to levels close to those observed during the academic year. Past research on summer learning programs has been discouraging on this point, suggesting that, although summer programs usually increase learning, the rate of summer learning does not approach what is observed during the school year (Cooper et al. 2000). But most summer learning programs have different content than courses during the academic year; in addition, participation in summer schools is often voluntary (Cooper et al. 2000), and attrition can be a serious problem (Borman and N. Maritza Dowling 2006). In year-round schools, by contrast, the curriculum is the same during the summer as during the rest of the year, and attendance is mandatory.
The finding that schooling increases learning during the summer also validates the results of past seasonal research, which has assumed that summer learning slows down because school is out, not because of something else that happens during the summer. As we mentioned in the introduction, if season affects learning independently of school, then many of the conclusions of seasonal research come into question. Our results suggest that this is not a concern.

Finally, the results provide further evidence that academic learning can be increased simply by increasing the number of days that children spend in school. During fall, winter, and spring, nine-month schools have more days of instruction than year-round schools, and children in nine-month schools learn proportionately more. During the summer, by contrast, year-round schools offer dramatically more instruction than nine-month schools, and children in year-round schools learn dramatically more. It is no great stretch to imagine that a school which increased total days of instruction—instead of merely redistributing days from one part of the year to another—would also increase total learning. This is the premise of extended-year charter schools, such as the KIPP schools, the Harlem Promise Academy, and the Andre Agassi College Preparatory Academy. Extra days of instruction may also be part of the advantage of east Asian schools that offer 200 or more days of instruction per year. Perhaps future attempts to solve the problem of summer learning should focus on extending or supplementing the school year rather than just rearranging it.
Chapter 4
Have Schools Made U.S. Children Obese?

U.S. schools play a role not only in developing children’s knowledge and skills, but also in monitoring and maintaining children’s health. Schools provide health-related services such as screening tests for vision, hearing, and lice; courses in health and physical education; opportunities to exercise during recess; and meals that meet federal nutrition standards. To the degree that American children are reasonably healthy, it seems fair to give schools some of the credit. And when children’s health deteriorates, it seems reasonable to ask whether schools deserve some of the blame.

Over the past generation, the most worrisome change in children’s health risks has been the rapid rise in obesity. In the 20 years between 1980 and 2000 the prevalence of obesity among school-age children approximately tripled, from about 5% to 15% of children aged 6-19 (Ogden, Flegal, et al. 2002). The rise in obesity was much faster for black and Hispanic children than for white children, so that more than 20% of black and Hispanic children are now obese (Ogden, Flegal, et al. 2002; Ogden et al. 2008). Although the increase in child obesity has slowed or stalled since 2000 (Ogden et al. 2008), for the past ten years we have been living in an era when obesity rates are three times higher than they were just a generation ago.

\footnote{Some authors give more precise figures, but confidence intervals for obesity prevalence are typically three more percentage points wide, and vary from survey to survey and subgroup to subgroup (e.g., Ogden, Carroll, and Flegal 2008). Given the imprecision and variability of prevalence estimates, we have chosen to give accurate but slightly vague estimates such as “about 15%.”}
Given the general principles laid out in chapter one, it seems unlikely that schools account for more than a small part of the rise in child obesity. Compared to the rest of children’s lives, schools are relatively stable and uniform, and these qualities make schools unlikely culprits for a problem that has emerged quickly and that affects some groups—blacks and Hispanics—much more than others.

Nevertheless, there are good reasons to examine the role of schools in the child obesity epidemic. The idea that schools are stable and uniform comes primarily from research on children’s learning and achievement, and may not hold for outcomes related to health. Indeed, during the obesity epidemic the prevalence of obesity has risen faster for children of school age (ages 6-19) than it has for younger children (ages 2-5 and ages 2-23 months) (Ogden, Flegal, et al. 2002). Second, if school practices have played an important role in causing the child obesity epidemic, it would be valuable to understand what the relevant school practices are, since changing school practices is much more straightforward than changing other influences on children’s health.

In this chapter, we review the rise of the child obesity epidemic and consider potential causes, focusing special attention on the possible role of schools. We review the research on school practices that might be relevant to obesity, including federally regulated school breakfasts and lunches, unregulated “competitive foods” that are also sold in schools, and time scheduled for physical activity in recess and physical education. Although some of these practices have changed for the worse during the child obesity epidemic, we find that the combined effect of all school changes can explain only about 20% of the increase in child obesity. Unless we have overlooked something important
about the school environment, the main sources of the obesity epidemic must lie elsewhere.

Having reviewed the effects of specific changes in school practices, we estimate the total effects of changes to school and non-school environments on children’s BMI. We do this using seasonal data that let us compare rates of BMI growth when children are in school, during the academic year, and when children are out of school, during summer vacation. We make this school-vs.-vacation comparison for two cohorts of children—a cohort measured in 1987-91, about halfway through the rise in obesity, and a cohort measured in 1998-2000, when obesity was near its current peak.

We find that the earlier cohort of children gained body mass index (BMI) at least as quickly during the school year as during summer vacation. The later cohort, by contrast, gained BMI much faster during summer vacation than during the school year. Although data on the two cohorts are not perfectly comparable, the differences corroborate the idea that the rise in child obesity has been due primarily to changes in the non-school environment. Relative to the school environment, it is the non-school environment that has grown dramatically more fattening.

We conclude by reviewing the sociological and policy implications, discussing the role that schools play in society, and suggesting what kinds of policy interventions can and cannot be expected to make a dent in the problem of child obesity.

*Overweight Among School-Aged Children*

Since the 1960s, U.S. children’s weights and heights have been monitored by the National Health Examination Survey (NHES), which in 1971 expanded to become the National Health and Nutrition Examination Survey (NHANES). From the NHES in the
1960s through the NHANES III in 1976-80, the weights of school-aged children were steady. But sometime between the NHANES III in 1976-80 and the next NHANES in 1988-94, school-age children’s weights began a rapid increase (Ogden et al. 2007). By the NHANES of 2002, children of elementary-school age (6-11 years) were, on average, nine pounds heavier than were children of their parents’ generation, 20-40 years before. Older children were no better off; among children of middle and high school age (12-19 years), girls in 2002 were 12 pounds heavier and boys were 15 pounds heavier, on average, than were children of their parents’ generation (Ogden et al. 2004). These gains in weight were disproportionate to children’s gains in height, which increased only half an inch, on average, over the same period (Ogden et al. 2004).

In social science and epidemiology, obesity is almost always defined using body mass index (BMI), which is weight (in kilograms) divided by the square of height (in meters). BMI is moderately correlated with measures of fatness (Barlow and Dietz 1998) and is a moderate predictor of obesity-related health and social outcomes (Yusuf et al. 2005; Burkhauser and Cawley 2008; Barlow and Dietz 1998). In population research, BMI is widely used out of habit and convenience, yet it has also been criticized as a measure of obesity (Burkhauser and Cawley 2008; Ross et al. 1988). Like any measure based on weight, BMI does not distinguish fat from lean body mass, and so BMI does not measure fatness or predict health risks as well as alternative measures such as the waist-to-hip ratio (Yusuf et al. 2005; Kragelund and Omland 2005). Nevertheless, BMI remains the most common and convenient measure of obesity, and in many social surveys, including the two analyzed in this chapter, BMI is the only obesity measure available.
Figure 4.1 shows smoothed BMI percentiles for boys age 2 to 20, calculated by the Centers for Disease Control and Prevention (CDC) using data primarily taken from the NHES and NHANES surveys conducted between 1963 and 1980—i.e., before the rise in child obesity (Kuczmarski et al. 2000; Ogden, Kuczmarski, et al. 2002). For each age, the curves show the 50th, 85th, and 95th percentiles for children’s BMI. Boys and girls are plotted separately, as is customary, although they are very similar until age 17 or so.

![Figure 4.1. BMI for age.](image)

The monthly BMI percentiles underlying this chart can be found at [http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm#males](http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm#males) and [http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm#females](http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm#females).

By age 20, the 85th percentile for children is near a conventional threshold for adult overweight (a BMI of 26) and the 95th percentile is near a threshold for adult obesity (a BMI of 30). For this reason, we use the 85th percentile as a threshold for child overweight and the 95th as a threshold for child obesity. A common alternative is to call the 95th percentile the threshold for overweight and the 85th a threshold for being “at risk for overweight” (see Flegal, Tabak, and Ogden 2006 for discussion).

---

18 For children under 6, the calculations also used NHANES data from 1988-1994. In addition, some supplemental data sources were used to estimate the percentiles BMI at birth (Kuczmarski et al. 2000).
The curves in Figure 4.1 are based on cross-sectional data and do not necessarily represent individual trajectories for growth. For example, children whose BMI is at the 95th percentile at age 2 will typically revert toward a lower percentile by age 5 (Li, Park, and Roche 1998). Nevertheless, the curves do give some idea what normal and abnormal BMI growth look like. A child near the median has a J-shaped BMI trajectory, losing BMI from age 2 to age 6 and then gaining until age 20 (and beyond). The inflection point at the bottom of children’s J curve is called the “adiposity rebound” (Rolland-Cachera et al. 1984). The adiposity rebound occurs around age 6 for a child at the median, but earlier for a child who is overweight or obese. An early adiposity rebound increases sixfold the risk for adult obesity (Whitaker et al. 1998).

It is important to remember that the lines in Figure 4.1 represent what the percentiles were before the rise in child obesity. Today the percentiles have shifted dramatically. From 1963 until 1976-80, the 95th percentile was close to the curve drawn at the top of Figure 4.1, but by 1988-94 that same curve approximated the 90th percentile, and by 1999-2000 it was near the 85th percentile. In other words, before 1980 only 5% of the school-age population exceeded the age-and-gender-specific threshold for obesity, but by 1999-2000 about 15% of the school-age population—and more than 20% of the black and Hispanic population—exceeded that same threshold.

The curves in Figure 4.1 have been widely reproduced, but they are hard to read if you want to know the exact timing of the adiposity rebound, or what typical BMI growth is at a particular age—quantities that will be important in our later analyses of BMI.

---

19 This is a straightforward example of reversion toward the mean. Simply put, for a child in the 95th percentile there are many more ways to move down than to move up in the distribution. (See the "red T-shirt argument" in Samuels 1991.)
growth. To make these quantities more accessible, we differenced the monthly time series underlying Figure 4.1 to produce a plot of monthly growth rates, shown in Figure 4.2.

Figure 4.2. Monthly growth rates at different percentiles of the BMI distribution.

Obtained by differencing the monthly BMIs in the previous Figure.

The curves in Figure 4.2 clarify patterns of BMI growth in childhood. The adiposity rebound—the transition from negative to positive BMI growth, where the curve crosses the horizontal axis—takes place at age 6 for a boy of median BMI, at age 5 for an overweight boy, and closer age 4 for an obese boy. The rebounds for girls are slightly earlier. After the adiposity rebound, the growth rates for median, overweight, and obese children diverge until age 9 or so, then slowly converge, especially for boys, until by age 15 boys of different BMIs are gaining BMI at about the same rate. This suggests that most of the BMI gaps between median, overweight, and obese children grow in elementary and middle school rather than high school.

It is important to recognize that the curves in Figure 4.2, which were obtained by differencing age-specific BMI percentiles, do not represent percentiles for growth. Instead, they represent the short-term growth rates characteristic of children at different
percentiles of BMI. At age 16, for example, children at the 50\textsuperscript{th}, 85\textsuperscript{th}, and 95\textsuperscript{th} BMI percentiles have similar rates of BMI growth—but of course they have very different BMIs.\textsuperscript{20} Percentiles for BMI growth would be very different (Berkey and Colditz 2007); unfortunately, reference growth rates based on longitudinal data are currently unavailable.

**Risks Associated with Obesity**

Although it is possible that some of children’s weight gain results from healthy increases in muscle mass or bone density, much of the gain is clearly unhealthy. The rise in children’s weight has been accompanied by a rise in the incidence of obesity-related illness among older children, such as nonalcoholic fatty liver disease and type 2 diabetes (Ogden et al. 2007)—formerly known as “adult-onset” diabetes but now a disease of childhood as well. Child obesity is also associated with hypertension, a common precursor to cardiovascular disease (Sorof and Daniels 2002). There is also some evidence linking obesity to asthma, though it is not conclusive (Ford 2005).

It should be acknowledged that the risk of obesity-related disease in childhood is relatively low (Ogden et al. 2007). The more important risks come from carrying obesity and associated risk factors, such as hypertension, into adulthood (Must and Strauss 1999). Compared to children of median BMI, children who are obese past age 10 are six times as likely to become obese adults (Guo et al. 1994), and obese adults face increased risk

\textsuperscript{20} It is also worth emphasizing that the growth rates in Figure 4.2 represent short-term, month-to-month growth. As we remarked earlier, the BMI-for-age percentiles in Figure 4.1 are not long-term growth curve because a child who is at, say, the 95\textsuperscript{th} percentile at age 2 will commonly be at a lower percentile by age 5. Over the short term, though, month to month it seems safe to assume that children’s position in the BMI distribution is fairly stable, especially near the median. For that reason, the curves in Figure 4.2 may be interpreted as approximate monthly growth rates.
for potentially fatal illnesses, most notably diabetes and cardiovascular disease (Mokdad et al. 2003; Ogden et al. 2007; Yusuf et al. 2005).

The adverse consequences of obesity go beyond the physical. Obese children have an increased risk of depression and social isolation (Erickson et al. 2000; Strauss and Harold A. Pollack 2003). Research on adults suggests that the relationship between depression and obesity is a vicious circle, with obesity increasing depression and depression increasing obesity (Luppino et al. 2010).

The adverse effects of obesity are also not limited to the obese. The national cost of treating obesity-related illness is $147 billion and is growing substantially faster than other medical expenses (Finkelstein et al. 2009). These costs are shared by non-obese persons through private and social insurance programs. In addition to depleting America’s financial capital, the social isolation of obese individuals (Strauss and Harold A. Pollack 2003) could reduce America’s social capital, with adverse consequences for economic and civic life.

**Causes of Obesity**

The causes of obesity can be divided into proximate causes, such as eating and exercise, and distal causes such as marketing, changes in consumer demand, and government policies.

**Proximate causes**

Compared to other outcomes studied by social science, obesity has remarkably simple proximate causes. As described in any basic physiology textbook (e.g., Sherwood 2005), obesity comes from an *energy imbalance*, an excess of energy intake over energy expenditure, or of calories in over calories out (Garrow 1978). *Calories in* depend
primarily on calories in the diet, although the mixture of nutrients in the diet may also exert some influence, and there are important differences in how much energy different individuals harvest from the same food (Turnbaugh et al. 2006; Ley et al. 2006). Calories out have three major components: physical activity, basal metabolism, and the thermic effect of food. On average, physical activity accounts for only 30% of energy expenditure—less in sedentary people, more in active people—and comes primarily from everyday activities, such as fidgeting or walking down the hall, rather than intense and deliberate exercise. The thermic effect of food is the energy consumed by digestion and fat storage, and accounts for another 10% of energy expenditure—more for diets high in hard-to-digest foods such as fiber and protein, less for diets high in easily digestible foods such as simple carbohydrates. The lion’s share of energy expenditure—60% on average—comes from basal metabolism, the energy consumed by bodily functions other than digestion, such as pumping the heart and lungs and maintaining a normal body temperature. Basal metabolism is higher for young people, higher in people with high lean body mass, and higher in hot or cold environments where body temperature is harder to maintain.

Every component of energy balance can be influenced by environment or behavior, and given the major changes in Americans’ environment and behavior over the past generation, any energy component could have changed in a direction consistent (or inconsistent) with the increase in obesity. Unfortunately, changes in the energy balance are difficult to evaluate because data on the various energy components are either poor or nonexistent. For example, there are no regular national surveys of basal metabolism, so it is impossible to evaluate the claim that improvements in indoor climate control,
especially air conditioning, have reduced metabolic rate by making it easier to maintain body temperature (Keith et al. 2006; Masako Kobayashi and Maiko Kobayashi 2006).

There are regular national surveys of food consumption and of physical activity, but these surveys rely on self-report and are subject to serious reporting bias. Individuals, especially the obese, substantially under-report their own consumption and over-report their own activity (Lichtman et al. 1992). Surveys of consumption have improved in an attempt to reduce reporting bias (Enns, Mickle, and Goldman 2003), but this in itself makes it difficult to compare current and past consumption. An increase in self-reported consumption could mean that consumption has really increased, or it could just mean that reporting has grown more complete.

Attempts to infer consumption and activity trends from available data are confusing. Some authors report increases in consumption (Cutler, Glaeser, and Shapiro 2003), some report decreases (Prentice and Jebb 1995), and there is similar confusion around changes in activity levels. Some authors claim that consumption and activity have been overemphasized, and review other influences—from reductions in sleep to increases in age at childbirth—in an attempt not to miss anything important (e.g., Keith et al. 2006).

Recent research has clarified the picture by shifting the focus from self-reported consumption, which is a private matter, to food production, which is a matter of public record. The data on food production are not at all confusing, and strongly implicate consumption as the predominant contributor to the secular increase in obesity. The U.S. Department of Agriculture estimates food consumption by adjusting food production for
losses at the producer, retail, and consumer level\textsuperscript{21} (US Department of Agriculture, Economic Research Service 2010). During the rise in obesity, this estimate of food consumption, which had been stable or declining for decades, increased by about 450 calories per person per day (Swinburn et al. 2009; Cutler et al. 2003). The rise began in the mid 1980s, at almost exactly the same time as the initial rise in obesity (Cutler et al. 2003). And the amount of the increase, split in a plausible way between adults and children, is almost exactly the right size to explain the increases in both child and adult obesity (Swinburn et al. 2009).

On general principles, it is not hard to believe that the obesity epidemic is due primarily to an increase in food consumption. The increase in Americans’ weight is consistent with a 20% increase in energy intake or a 20% decrease in energy expenditure. With physical activity accounting for just 30% of total energy expenditure, it is hard to see how decreases in physical activity could reduce energy expenditure by anywhere close to 20%. A 20% increase in energy intake, on the other hand, is all too easy to imagine. A 20% increase in daily energy intake is equal to “about one can of soda and a small order of French fries” (Swinburn et al. 2009).

\textit{Distal causes}

If consumption has driven the increase in obesity, the question remains: what has driven the increase in consumption? Some authors emphasize demand-side factors, especially the movement of women from the home into the workforce (Patricia M.

\textsuperscript{21}Food losses may be underestimated, but changes in food consumption are unbiased unless the error in estimating food losses has grown over time (Swinburn, Sacks, and Ravussin 2009). While there are plausible arguments for why the error in estimating food losses might have grown or shrunk over time (Swinburn et al. 2009), it seems implausible that the error has grown by anything close to 450 calories per day.
Anderson, Butcher, and Levine 2003b). Although the increase in maternal employment has not reduced parental supervision as much as many commentators imagine (Bianchi 2000), it has certainly increased the opportunity cost of home-cooked meals and reduced the relative cost of high-calorie substitutes such as restaurant meals, takeout and delivery meals, and convenience foods that need minimal preparation.

Other authors emphasize supply-side factors such as food marketers’ attempts to grow sales beyond consumers’ “finite stomach” capacity by increasing portion sizes, encouraging snacks between meals, and shifting consumers from simple and recognizable agricultural products, such as potatoes or corn, toward more processed and higher-profit food items such as French fries and corn chips (Nestle 2007; Pollan 2006). Some authors emphasize the unintended consequences of government subsidies for corn and soybeans, which artificially reduce the cost of high-calorie foods such as corn-fed beef, high-fructose corn syrup, and partially hydrogenated soybean oil (Pollan 2006). Other authors emphasize technological breakthroughs in centralized food production, coupled with consumer technology, such as the home microwave, that make it easier to serve and eat centrally prepared food (Cutler et al. 2003). A generation ago, the easiest way to prepare a potato was to bake or boil it. Today it is easier to reheat a package of French fries (Cutler et al. 2003).

**School Causes**

How much have schools contributed to the child obesity epidemic? In addressing this question, we should look for evidence that something about schools food and exercise practices has changed. It is not enough to show that a school practice is unhealthy; for example, it is insufficient to show that children are eating a lot or
exercising very little in school. To explain the child obesity epidemic, we must show that schools have become more unhealthy—e.g., that children are eating more or exercising less in school—than was the case before the epidemic began a generation ago.

In this section, we add up the obesity effects of changes in schools’ diet and exercise practices over the past generation, and conclude that the combined effects of all changes could account for no more than about 20% of the observed increase in child obesity. We carry out this calculation by using the research literature to estimate the effect on obesity of changes in each school practice, and multiplying those effects by the amount that each school practice has changed during the era of rising obesity. Where estimates are lacking or uncertain, we make assumptions that are typically ungenerous to schools, so that our final tally is very likely an overestimate of how much schools have increased obesity. The results suggest strongly that the major sources of the child obesity epidemic are not to be found at school.

Because different authors measure effects in different ways, it will be helpful to have several ways to think about the size of the increase in child obesity. We have already seen the relevant figures, but it is helpful to review them here. During the rise in obesity, the percentage of children who are obese increased by about 10%, from 5% to 15% of the school-age population (Ogden, Flegal, et al. 2002). During the same period, the weight of elementary school students’ increased by 9 pounds, and the weight of middle and high school students increased by 12 pounds for girls and 15 pounds for boys (Ogden et al. 2004). All these changes were caused by a 20% increase in daily energy consumption—or a 350-calorie increase for a child of average weight (Swinburn et al.
2009). All of these figures will serve as useful comparisons for the effects of school practices on children’s weight, obesity level, and energy balance.

Given the evidence linking the obesity epidemic with changes in the food supply, the first school practices to examine relate to food. And the first thing to notice is that little of children’s food comes from school. Children do not attend school until they are five and a half, and afterward, on school days children get only 19% of their daily calories from the school cafeteria (Gleason and Suitor 2001). Since school is in session only half the days of the year, this means that the school cafeteria supplies only 8.5% of the calories that school-age children consume in a year. Starting from such a low base, it is hard to see how any plausible increase in the school food supply could account for much of the 20% increase in consumption that is responsible for the epidemic of child obesity.

But the children who are most prone to obesity get much more of their food from school. On school days, children enrolled in the National School Lunch Program or School Breakfast Program get 50% of their daily calories from their school cafeteria (Gleason and Suitor 2001). And children who participate in these federal meal programs are disproportionately likely to be black or Hispanic—the very ethnic groups at highest risk for obesity (Gleason and Suitor 2001). Could school meals be partly responsible for obesity among black and Hispanic students?

Schools provide two types of food—federally regulated school meals (lunch and breakfast) and comparatively unregulated “competitive foods.” We will examine the effect of school meals first.
Research on the relationship between school meals and obesity has produced conflicting results. Cross-sectional regressions with controls for poverty, ethnicity and other characteristics suggest that school lunches do not increase BMI, and school breakfasts may actually reduce BMI (Gleason and Dodd 2009). More rigorous research uses longitudinal data that follows young children as they move in and out of school meal programs. One longitudinal study suggests that eating a school-provided lunch, whether it is subsidized or not, increases the risk of obesity by 2% (Schanzenbach 2009). Another study suggests that school- and day-care meal subsidies reduce average BMI by 4 percentile points (Kimbro and Rigby 2010).

Even if school meals do increase obesity risk, that increase would do little to explain the rise in obesity since 1980. As we remarked earlier, to explain the rise in child obesity, it is not enough to show that a school practice is unhealthy. We must show that the practice has become substantially more unhealthy during the obesity epidemic. And school meals have not. Since the National School Lunch Act in 1946 and the inception of the School Breakfast Program in 1966, school lunches and breakfasts have been required to provide just one-third and one-quarter, respectively, of the recommended daily energy allowance for total calories. During the obesity epidemic, the only change to school breakfasts and lunches was a 1995 regulation reducing the proportion of total calories that could come from fat. In short, school-provided meals have not grown more fattening during the rise in child obesity.

Although school meals have changed little in content, they might have contributed to the obesity epidemic if participation in school meal programs had substantially

---

22 The REA was last revised in 1989. As of this writing, I have been unable to determine whether it went up or down.
increased. But this has not happened, either. As shown in Table 4.1, from 1980 to 2000, when most of the increase in obesity took place, participation in school lunch programs (both subsidized and full-price) fell from 58% to 51% of children who were enrolled in school. Participation in school breakfast programs increased by a compensating amount, from 8% to 14% of school-enrolled children. On balance, considering breakfast and lunch together, participation in school meal programs was roughly unchanged.

Table 4.1. Participation in school breakfast and lunch programs.

<table>
<thead>
<tr>
<th>Year</th>
<th>School enrollment, millions</th>
<th>School meals served, millions (with % of enrollment)</th>
<th>Breakfasts</th>
<th>Lunches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>Free</td>
</tr>
<tr>
<td>1980</td>
<td>46.2</td>
<td></td>
<td>3.6</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(8%)</td>
<td>(6%)</td>
</tr>
<tr>
<td>2000</td>
<td>53.4</td>
<td></td>
<td>7.5</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(14%)</td>
<td>(11%)</td>
</tr>
</tbody>
</table>


To appreciate how small school-meal effects are in comparison to the size of the obesity problem, suppose that we give school-meal research the most pessimistic possible reading. We will ignore the 7% decrease in school lunch participation and focus on the 6% increase in school breakfast participation. In addition, we will ignore the findings suggesting that school meals reduce obesity risk, and focus on the study suggesting that school lunches increase the risk of obesity of 2%. We will assume that this finding
applies to school breakfasts as well. Under these unrealistically grim assumptions, between 1980 and 2000, 6% of U.S. school children would have experienced a 2% increase in the risk of obesity, so that school meals would have induced just a 0.12% increase in child obesity (6% \times 2\%). The actual increase in obesity prevalence over this period was about 10%, so even under the most pessimistic assumptions school meals could account for just 1% of the increase.

Having ruled out regulated school meals as an important cause of child obesity, we turn to the “competitive foods” that are also sold in schools. Federal regulations divide competitive foods into two categories. The first category of competitive foods are nutritious enough that federal guidelines place no restrictions on their sale. These foods have to exceed a standard for nutrition, although the standard is not very high; calories are not limited, and foods can meet the nutrition standard by providing just 5% of the recommended daily allowance for any one of eight nutrients (protein, vitamin A, vitamin C, niacin, riboflavin, thiamine, calcium, or iron). Foods meeting this minimum standard can include ice cream, chips, crackers, and many fast food items.

The second category of competitive foods are “foods of minimal nutritional value” (FMNVs), such as candy and soda, that do not meet even the low nutritional standards described above. According to federal regulations, schools that participate in federal meal programs are not permitted to sell FMNVs in the cafeteria during mealtime—but are permitted to sell them just outside the cafeteria at mealtime, or inside the cafeteria at other times. Some local and state governments restrict the sale of competitive foods more tightly than required by federal law, but 90% of schools sell at least some competitive foods (United States Government Accountability Office 2005).
Schools can earn significant revenue from competitive foods. Merely selling competitive foods is just the entry level to this revenue stream. Schools can earn further revenue by advertising competitive foods, or by offering a particular soda vendor exclusive “pouring rights” that shut out competing brands. These practices are widespread. In the year 2000, nearly all schools sold soda; 75% sold sweet and salty snacks; 67% had a pouring rights contract; and 44% advertised foods on school property (Patricia M. Anderson, Butcher, and Levine 2003a). Among the 30% of high schools that sold the most competitive foods in 2003-2004, annual sales revenues averaged $125,000 per school (United States Government Accountability Office 2005), and a pouring rights contract can bring over a million dollars to a school district (Patricia M. Anderson and Butcher 2006). Although it might be expected that cash-strapped districts would be most likely to market competitive foods, competitive food marketing is actually more common in affluent districts. Apparently schools and their partners, like other revenue-seeking enterprises, try hardest to sell goods to the consumers who can best afford them (Patricia M. Anderson and Butcher 2006).

A recent cross-sectional analysis estimated that a 10% increase in competitive food practices is associated with a 1% increase in children’s BMI (Patricia M. Anderson and Butcher 2006). We lack full information on how quickly competitive food practices increased during the twenty-year rise in children’s obesity, but during the period 1994-2000 high schools were increasing competitive food practices at a rate of about 1% per year, an increase that would cause a 0.1% annual increase in children’s BMI, or about a 2% increase over twenty years. During the twenty-year rise in obesity, the actual rise in

---

23 Although federal law continues to permit all of these practices, since 2000 some have been restricted by states, local governments, and school districts.
children’s BMI was about 10%, so aggressive competitive food practices could potentially explain about 20% of the increase, at least in high school and middle school. In elementary schools, where competitive food marketing is much less pervasive (Patricia M. Anderson et al. 2003a), competitive food practices may explain less of the increase in child obesity.

The idea that school-based food marketing explains about 20% of the increase in obesity among older children fits well with the distribution of food marketing dollars. In 2006, the food industry spent $1.6 billion marketing to children and adolescents, and about 16% of that, or $250 million, was spent in schools (Federal Trade Commission 2008). More money was spent marketing to teenagers in middle and high schools than marketing to younger children in elementary schools (Federal Trade Commission 2008). If marketing budgets are efficiently distributed, it makes sense that the 16% of marketing dollars that are spent in school would account for something like 20% of the increase in consumption.

Having covered schools’ dietary practices, we now move on to their exercise practices. Some critics claim that accountability systems that emphasize reading and math scores—most prominently the federal No Child Left Behind law—have pressured schools to increase the time that they devote to reading and math instruction and reduce the time spent on lower priorities including recess and physical education (PE) (e.g., Rothstein, Jacobsen, and Wilder 2008). According to these critics, the schools most likely to allocate time away from physical activity would be those under the most pressure to increase their math and reading scores—namely, schools serving poor black and Hispanic children, who have also displayed the greatest increase in obesity.
The data on school-based physical activity are patchy. Federal surveillance of high school physical education did not begin until 1991, when the rise in obesity was already halfway to its peak. Since then, about half of students have consistently reported that they attended PE at least weekly, and of that half, the fraction who attended daily has fallen from four-fifths to three-fifths, with all of the decline occurring by 1995 (Centers for Disease Control and Prevention 2008). We do not know what happened before 1991, but if we extrapolate the trends backward, it seems plausible that at the beginning of obesity’s rise, in the early 1980s, it was still the case that only half of high-school students attended PE, but of that half, nearly all attended daily. If that is the case, then, at the beginning of the rise in obesity, half of all high school students attended daily PE and the other half attended less than weekly, if at all. By contrast, at the peak of the obesity epidemic only 30% of high school students attended PE daily, 20% attended PE at least weekly, and the remaining 50% attended less often.

Given the breadth of these attendance categories—daily, at least weekly, less than weekly—it is hard to know exactly how much average attendance has fallen, but it is not hard to estimate a worst case. Under worst-case thinking, children who attended PE at least weekly but not daily attended just once a week, and children who attended less than weekly did not attend at all. With PE attendance rounded down into worst-case bins—daily, weekly, and none—the average level of high-school PE attendance would be estimated to have fallen from 2.5 days a week to 1.7, a decline of 32%.24

---

24 To spell out the calculation: At the beginning of the rise in obesity, with half of students attending PE daily (5 times a week) and half not attending at all, the average attendance per week was 2.5 days (0.5x5+0.5x0). At the peak of obesity, with 30% attending 5 times a week, 20% attending once a week, and 50% attending at all, the average attendance per week was 1.7 days (0.3x5+0.2x1+0.5x0).
Physical activity in elementary school is not tracked by the federal government, but it has been tracked by the Center on Education Policy (CEP), an advocacy group, since the passage of the No Child Left Behind (NCLB) Act in 2001. Through 2007, the CEP had found no evidence of declines in elementary school PE time (Mcmurrer 2007)\textsuperscript{25}, despite the pressures introduced by NCLB. The same may be said of PE in high school, which as we remarked earlier has not decreased since 1995 (Centers for Disease Control and Prevention 2008). On the other hand, the CEP has found that since 2001, 20% of districts had reduced time for recess, apparently to make room for instruction in English and math, the subjects tested under NCLB (Mcmurrer 2007).\textsuperscript{26}

There is some evidence, then, that school-based physical activity has declined. Yet this decline can explain little of the increase in obesity. The reason is that physical activity levels in school were never very high, so that little was lost by reducing them. A recent study found that third-graders spend, on average, just two half hour sessions—an hour in total—in PE, and spend just 37\% of that hour time in moderate to vigorous physical activity (Nader 2003). For an average-weight third grader, the mix of activities in PE class burns only 133 calories per hour (Nader 2003), just 48 calories more than the child would burn sitting at his desk.\textsuperscript{27} Using our high-school estimate of a 32\% decline in PE time, we would estimate that elementary-school students have lost a half an hour of

\textsuperscript{25} In its publications, the CEP reports that 9\% of sampled districts have reduced PE time (Mcmurrer 2007, 2008), but does not report how many districts have increased PE or left it unchanged. Through correspondence with Jennifer Mcmurrer at the CEP (May 10, 2010), I learned that 7\% of sampled districts had increased PE, and 84\% had made no change. On balance, any net change in PE time has been insignificant.

\textsuperscript{26} Again, in its publications, the CEP reports that 20\% of sampled districts have reduced recess time (Mcmurrer 2007, 2008), but does not report how many districts have increased recess or left it unchanged. Through correspondence with Jennifer Mcmurrer at the CEP (May 10, 2010), I learned that only 1\% of districts had increased recess. On balance, the net reduction in recess time has been significant.

\textsuperscript{27} Calories burned in PE and by sitting are tabulated in Nader (2003). For ease of interpretation, we have converted some of Nader’s estimates from per-minute to per-hour energy expenditure.
PE time per week since the beginning of the rise in obesity—a change in energy balance of just 24 calories per week, or 3 calories per day, and only during the school year. Since the obesity epidemic was caused by a change in energy balance of 350 calories per day (Swinburn et al. 2009), only 1% of this change can be attributed to reductions in PE.

Our finding that changes in elementary school PE have little effect on energy balance is consistent with a recent evaluation finding that a one-hour increase in PE between kindergarten and first grade did not reduce BMI for normal-weight children or for overweight boys. The increase did reduce BMI for overweight and obese girls, but only by 0.3 BMI points (Datar and Sturm 2004), about one tenth of the gap between obese and median-weight girls in first grade. (See Figure 4.1.)

In high school, the decline in high school PE has been offset to some degree by a long-term increase in participation in high-school sports. During the period of rising obesity, from 1980 to 2000, participation in high school athletics rose from 48% to 53% for boys and from 25% to 38% for girls.28 Girls’ athletic opportunities have steadily increased since the 1972 Title IX legislation banning gender discrimination in federally funded education programs, including sports. When increases in athletic participation are included in the balance, it may be that school-based physical activity has not declined at all, at least in high school. It may even be that some of the decline in high school PE is due to students being exempted because of participation in high school sports. We should acknowledge, however, that competitive sports are an unlikely form of activity for many obese children.

28 These figures were obtained by dividing high school sports participation by high school enrollments. Participation numbers are collected by the National Federation of State High School Associations (nfhs.org), and enrollment numbers are reported by the National Center for Education Statistics (nces.ed.gov).
A final way that schools may have affected obesity is by increases in homework. Critics sometimes allege that increases in homework have displaced other important activities, including leisure-time physical activity. It is true that homework increased in the early 1980s, around the time that child obesity began to rise. But the increase in homework was very slight, and throughout the period of obesity’s rise, homework loads remained considerably lower than they were in the 1960s (Gill and Schlossman 2003), when obesity prevalence was considerably lower.

This completes our review of obesity-related changes in the school environment. To sum up, we have found that regulated school meals have changed very little and account for essentially none of the increase in child obesity. Perhaps 1% of the increase in obesity can be explained by reductions in physical education and recess, and those reductions may be offset by increases in high-school sports participation, particularly by girls.

The only documented change in school practices that can explain much of the increase in child obesity is the increase in school-based marketing of unregulated competitive foods. These increases could explain perhaps 20% of the increase in obesity among high school students, and probably less among elementary and middle school students, who are less targeted by marketers in school.

The finding that in-school food marketing could account for 20% of the rise of child obesity is consistent with the broader finding that the rise in both adult and child obesity is primarily due to increases in food consumption (Swinburn et al. 2009; Cutler et al. 2003). This finding is also consistent with our earlier prediction that school practices would be responsible for only a small part of the increase in children’s weight. The
increasing promotion of competitive foods in school is only part of a much larger marketing effort to grow sales of high-profit (and usually high-calorie) processed foods (Nestle 2007; Pollan 2006; Cardello and Garr 2009). As we remarked earlier, school-based food marketing accounts for less than 20% of the money spent marketing food to children—to say nothing of the money spent marketing to their parents.

In short, the evidence so far suggests that the vast majority of excess consumption and weight gain takes place outside of school walls.

Data and Research Design

In the previous section, we reviewed plausible school sources of obesity, estimated their individual effects, and then added those effects to estimate the total effect of school on child obesity. We estimated that changes in school practices could account for about 20% of the increase in child obesity, so that the remainder, or 80%, must come from outside of school.

This approach is plausible, but not watertight. One possible problem is that we may have overlooked important school practices, or misestimated the effects of practices that we did recognize. Another problem is that we did not explicitly model non-school effects; instead, we assumed that anything left over after school effects were accounted for must be an effect of the non-school environment. This is unrealistic if important school variables are correlated with omitted non-school variables. For example, if the children who encounter high levels of food marketing in school also encounter high levels of marketing outside of school, then the effect of school-based marketing will be overestimated because of confounding with the effect of out-of-school marketing. And
the true effect of out-of-school marketing will be more than what is left over after in-
school marketing is accounted for.

One approach to remedying these problems is to fit a single model that
simultaneously incorporates well-measured variables representing all of the school and
non-school effects on obesity. This approach is unrealistic since we are not aware of a
single data set that includes all of the school and non-school variables that could
plausibly affect obesity. Even if we had such data, it could be argued that important
variables were missing or poorly measured. Remember that Coleman et al. (1966)
encountered exactly this criticism when they attempted to model the effect of school and
non-school variables on achievement.

A more direct way to gauge the relative importance of school and non-school
effects is through a seasonal design that allows us to compare rates of BMI gain when
children are in school during the academic year and when they are out of school during
summer vacation. The advantage of this approach is that it does not require us to identify
all—or any—of the school and non-school variables that might affect obesity.

Past research has examined seasonal patterns in BMI gain for a single cohort of
children (von Hippel et al. 2007). But if our goal is to explain the change in obesity levels
over time, it would be more informative to investigate whether seasonal patterns have
changed between earlier and later cohorts. If it is true that the increase in obesity has
come primarily from outside the school environment, then we would expect to find that
recent cohorts of children have faster summer gains than their predecessors, while
school-year gains have not increased at much.
Toward that end, we analyzed data from two studies that took seasonal BMI measurements on children in different eras. More specifically, we analyzed the Beginning School Study (BSS), which collected seasonal BMIs in 1987-1991, about midway through the rise of child obesity (Alexander and Entwisle 2003), and we analyzed the Early Childhood Longitudinal Study, Kindergarten cohort of 1998 (ECLS-K), which collected seasonal height and weight data near the peak of the obesity epidemic in 1998-2000 (National Center for Education Statistics, n.d.).

If era were the only difference between the two studies, then any difference in BMI patterns could only be due to secular change in school and non-school influences on obesity. Unfortunately, the two studies had several differences in addition to the time in history when they were carried out.

One difference is that the studies took seasonal BMI measurements at different stages in children’s school careers. The BSS collected seasonal BMI data when participants were in 6th to 9th grade, while the ECLS-K collected seasonal BMI data in kindergarten through first grade. Both surveys measured BMI at other times in children’s schools careers, but it is only during the stated period that BMI was measured on the twice-yearly fall-and-spring schedule needed to estimate growth during the school year and the summer. The difference in children’s ages is concerning, since children of different ages differ with respect to weight and weight-related behaviors. Compared to younger children, older children have higher BMI, faster BMI growth, greater autonomy from parental food choices, and greater exposure to food marketing in school. In addition, the differences in BMI gain between obese and normal-weight children are different at

---

29 Both the BSS and ECLS-K follow children for a considerably longer period, but only the years stated include the twice-yearly fall-and-spring height and weight measurements needed to estimate school-year and summer BMI growth.
different ages (see Figure 4.2). Despite these differences, broad trends suggest that younger and older children are subject to very similar influences when it comes to obesity. For both younger schoolchildren (ages 6-11) and older schoolchildren (ages 12-19), the prevalence of obesity has grown at almost exactly the same rate, from about 5% of the population before 1980 to about 15% of the population since 2000 (Ogden et al. 2007). The common trajectory for younger and older children provides some comfort about the validity of comparing results from the BSS and the ECLS-K.

The studies also differ dramatically in size and scope. The BSS is a study of 838 children in 20 Baltimore City public schools, while the ECLS-K is a national study that started with 17,212 children in 992 public and private schools—although the part of the ECLS-K with seasonal BMI measurements is limited to a 30% random subsample of 5,380 students in 310 public and private schools. The BSS’s smaller sample reduces statistical power, so that our analysis is limited to larger and simpler effects than would be the case if we were examining the ECLS-K alone.

The geographic limitations of the BSS may also be important if BMI patterns are different in the Baltimore public schools than they are elsewhere. The Baltimore public schools do display less diversity than a national sample like the ECLS-K. The Baltimore public school population represented in the BSS is relatively impoverished, and all the children are either black or white; by contrast, the ECLS-K includes a substantial number of children from other racial and ethnic groups, especially Hispanics, as well as relatively affluent children and children in private schools. Beyond these characteristics, there is more uniformity among the public schools of a single city than there is the population of schools as a whole. In modeling the BSS, we will find that the between-school variation
is close to zero. It is worth pointing out, however, that Baltimore is near the national average in its obesity levels, at least for adults (Gregg et al. 2009).

In order to address these differences in sample size and geographic coverage we extracted a subsample from the ECLS-K that was more comparable to the children who attend the Baltimore public schools. Since Baltimore is a city at the border between the south and the northeast, our subsample was restricted to public schools in urban areas of the southern and northeastern census regions. This subsample of the ECLS-K contained 752 children in 41 schools—a sample size comparable to the BSS. The multi-city subsample does remain more diverse, however, than the single-city sample represented by the BSS.  

A final difference between the BSS and the ECLS-K is in the background variables that the surveys collected on children, families, and schools. In the BSS a handful of well-chosen background characteristics were recorded, while in the ECLS-K the available background characteristics were more extensive. This would be concerning if our analyses hinged on subtle background variables, but in fact our research is focused on the large and simple question of whether the school or non-school environment has changed more during the surge in child obesity. In addition, because of the BSS’s relatively small sample size, it is difficult to detect the effects of background variables even when they are available.

Table 4.2 summarizes the important design characteristics of the BSS, the ECLS-K, and the subsample of the ECLS-K that was chosen to be comparable to the BSS.

---

30 It would not be helpful to extract data representing only the city of Baltimore, since Baltimore, with 0.3% of the U.S. population, would be expected to account for only one of the 310 schools in the relevant part of the ECLS-K. In any case, it is not possible to identify an area as small as Baltimore in the public release of the ECLS-K.
Table 4.2. Samples and design.

<table>
<thead>
<tr>
<th></th>
<th>Beginning School Study (BSS)</th>
<th>Early Childhood Longitudinal Study, Kindergarten 1998 cohort (ECLS-K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Baltimore</td>
<td>US</td>
</tr>
<tr>
<td>Sector</td>
<td>Public</td>
<td>Public, private</td>
</tr>
<tr>
<td>Schools</td>
<td>20</td>
<td>310</td>
</tr>
<tr>
<td>Children</td>
<td>838</td>
<td>5380</td>
</tr>
<tr>
<td>Average birth date</td>
<td>1976</td>
<td>1993</td>
</tr>
<tr>
<td>Grades with seasonal BMI</td>
<td>Fall 6th - Spring 9th</td>
<td>Fall K - Spring 1st</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fall K - Spring 1st</td>
</tr>
</tbody>
</table>

Methods

Our analyses employed the three-level multilevel growth model described in our methodology chapter. In general, the issues encountered in modeling were the general issues encountered in fitting a multilevel growth model to seasonal data. However, we also had to solve three problems specific to working with BMI in these particular data sources.

Our first problem was to estimate measurement error. In the ECLS-K this was straightforward since the height and weight of each child was measured twice on each occasion. By comparing these paired measurements we estimated the standard deviation of the measurement error to be 0.1 BMI units.\textsuperscript{31} We assumed that the measurement error in the BSS, which had only one measurement per occasion, was the same as in the ECLS-K.

We suspect that comparing paired measurements leads to an underestimate of measurement error. In the ECLS-K, children were weighed twice, but the measurements were taken on the same scale by the same member of the survey staff on the same day. A

\textsuperscript{31} In comparing paired measurements we also discovered and corrected some data entry errors, although these corrections did not materially affect the analytic results.
full accounting of measurement error would also include variation from scale to scale, variation from staff member to staff member, and variation from day to day around the child’s stable level of BMI. To account for this possibility, we tried repeating our analyses with an error variance up to three times larger than the error suggested by comparing paired measurements. Our results were not materially affected, probably because even the inflated error variance was very small compared to the total variance in BMI.

Our second problem was how to calculate how long each child had been exposed to school and summer at the time of each measurement. To do this we need more than the measurement date, which is readily available in both the BSS and the ECLS-K. We also need the dates for the beginning and end of each school year, which are not in the public release of either data set. To get these dates for schools in the ECLS-K, we requested and obtained access to the relevant variables in the ECLS-K’s restricted data. To get these dates for the BSS, we corresponded with staff in the central office of the Baltimore City Schools, who told us that, at the time of the BSS, all of Baltimore public schools started on the day after Labor Day. Central staff did not know when the last day of school would have been, so we imputed a plausible date—June 17—by averaging the end dates across ECLS-K schools that started on the same day as the BSS schools. Our results are not sensitive to moving this end date a week or two in either direction.

An additional problem was that a large number of BSS children were missing measurement dates in Fall 1988. We initially filled these dates in using multiple imputation, but imputation typically assumes that the conditional distribution of the missing values is similar to that of the observed values (though not always; see Rubin
1987). From correspondence with BSS staff, we learned that the missing dates were actually quite unlike the observed dates; children with missing dates were typically measured before other children. BSS records indicated that the measurements with missing dates were taken in a narrow window between September 21 and September 30, 1988. We imputed a date of September 26, but any date in this narrow window would produce very similar results.

Our final problem was that we could not fit a full three-level model to the BSS data, probably because there was inadequate variation at the school level. This is symptomatic of the BSS’s being a more homogeneous sample than the ECLS-K. We solved this problem by eliminating the top level of the three-level model, yielding a two-level model where measurements (level 1) were nested within children (level 2). Even in this two-level model, we had to reduce the parameters further by restricting the level 2 covariance matrix to fit a heteroskedastic first-order autoregressive model (ARH(1)).

Results

Basic Model

We begin by fitting a basic model that simply estimates average BMI growth for each school year and summer. The results are shown in Table 4.3 and Figure 4.3.
Table 4.3. Average BMI growth in two cohorts of schoolchildren.

BSS, 1987-1991

<table>
<thead>
<tr>
<th>Development</th>
<th>6th grade</th>
<th>Summer</th>
<th>7th grade</th>
<th>Summer</th>
<th>8th grade</th>
<th>Summer</th>
<th>9th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>All schools</td>
<td>19.13***</td>
<td>0.11***</td>
<td>-0.13***</td>
<td>0.11***</td>
<td>-0.01</td>
<td>0.06***</td>
<td>0.08*</td>
</tr>
<tr>
<td><strong>Contrasts:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer - avg (prev, next)</td>
<td>-0.25***</td>
<td>-0.10**</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ECLS-K, 1998-2000

<table>
<thead>
<tr>
<th>Development</th>
<th>Initial BMI</th>
<th>Kindergarten</th>
<th>Summer</th>
<th>1st grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>All schools</td>
<td>16.205***</td>
<td>0.020***</td>
<td>0.076***</td>
<td>0.033***</td>
</tr>
<tr>
<td><strong>Contrast:</strong> summer - avg (kind,1st)</td>
<td>0.050***</td>
<td>0.115***</td>
<td>0.033***</td>
<td></td>
</tr>
<tr>
<td>BSS-comparable schools</td>
<td>16.374***</td>
<td>0.003</td>
<td>0.097***</td>
<td></td>
</tr>
<tr>
<td><strong>Contrast:</strong> summer - avg (kind,1st)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.001
Figure 4.3. Average school-year and summer BMI growth in the ECLS-K and the BSS.
The patterns are generally consistent with the theory that increases in obesity come from outside the school environment. In the BSS, children gained BMI in each school year from 6th to 9th grade, with slower gains in the later school years. By contrast, summer growth increased from year to year. In the summer of 1988 children lost BMI; in the summer of 1989 children maintained BMI, neither losing nor gaining; and in the summer of 1990 children gained BMI at about the same rate as during the school years before and after. This pattern of low but increasing summer growth makes sense since these children were measured at a middle stage in the child obesity epidemic, when the risk of obesity was lower than it is now but increasing from one year to the next. The fact that the increase is observed in the summer rather than the school year suggests that the changes responsible for the increase in obesity were taking place primarily in the non-school environment.

If we extrapolate this pattern forward, we might expect that by the end of the 1990s, as obesity prevalence was nearing their peak, BMI growth would be considerably faster during the summer than during the school year. And in the ECLS-K, that is exactly what we do find. Specifically, in the summer of 2000, average summer BMI growth was nearly four times faster than during the school year before, and more than two times faster than in the school year after.

On the whole, it appears that, during the rise of child obesity, the risk of excessive BMI growth increased much faster during the summer vacation than during the school year. This suggests that the changes responsible for the child obesity epidemic took place primarily outside of the school environment—a pattern consistent with our earlier review of findings on school effects.
This is the central finding of this section, so it is worth spending some time exploring ways that it could be wrong. One possible problem stems from the fact that, as we pointed out earlier, children in the BSS and ECLS-K were not just measured in different eras; they were also measured at different ages. Children in the ECLS-K had seasonal BMIs recorded in kindergarten and first grade, while children in the BSS had seasonal BMIs recorded in 6th-9th grade. It could be argued that the differences we observe are not period effects but age effects, not differences between earlier and later periods but differences between younger and older children.

This criticism hinges on an already-remarked weakness of the research design, and it would be fairly credible if the results in Table 4.3 and Figure 4.3 were the only evidence available. In fact, though, there is considerable independent evidence that supports the period-effect interpretation:

1. We know that there was a rise in obesity among school-age children between the time of the BSS and the time of the ECLS-K. That means that either school-year or (as our results suggest) summer gains must have increased from one survey to the next.

2. We also know that the rise in obesity affected younger and older children about equally. Since common outcomes suggest common causes, it seems likely that if summer gains increased for one age group, they also increased for the other.

3. Even during the 4 years of the BSS, we can see gains increasing from one summer to the next. This suggests that the difference in summer gains
between the BSS and the ECLS-K is a continuation of a trend that was already visible during the earlier survey.

4. Finally, our earlier review suggests that the rise in obesity had much less to do with changes in school than it did with changes elsewhere. If this is true, then we would expect to see summer gain increase faster than gains during the school year.

Taken together, all these observations strongly suggest that the observed increase in summer gains is primarily a period effect, not an age effect.

Another potential criticism of our results begins by pointing out that while the ECLS-K is a national survey representing all young schoolchildren, the BSS is limited to public school students in the city of Baltimore. Perhaps the differences between the ECLS-K and the BSS result in part from the distinctiveness of Baltimore.

We can answer this criticism by pointing out that the growth pattern in the ECLS-K as a whole are very similar to the pattern in the ECLS-K’s BSS-comparable schools—i.e., the public schools from southern and northeastern cities, such as Baltimore. This comparison demonstrates that there is nothing terribly distinctive about the growth patterns in southern and northeastern cities such as Baltimore. Indeed, local surveillance data suggests that Baltimore’s obesity patterns are not terribly exceptional. Although figures for Baltimore’s children are not available, Baltimore’s adults have an obesity prevalence of 31.8% (Gregg et al. 2009), which is about the same as obesity prevalence in the nation as a whole (Ogden et al. 2006).

A more telling criticism of our results is that some of BSS patterns are hard to believe. First, the BSS results are inconsistent from year to year. The first two years
display a whipsaw pattern, with rapid BMI gains in 6th and 7th grade and rapid BMI losses in the summer between. It is difficult to believe that growth could change directions so quickly, especially since the pattern does not continue; in subsequent years, growth rates in both the school year and the summer are more moderate.

In addition, the BSS 6th and 7th grade results are not compatible with our reference growth rates derived from the CDC’s BMI-for-age charts (Figure 4.2)\(^2\). According to our BSS analyses, the average growth rates for 6th and 7th graders during the school year is 0.11 BMI points per month, but in the CDC data this would be faster than the growth rate for a child at the 95th percentile of BMI. The BSS average growth rate for the summer between 6th and 7th grade—a loss of 0.13 points per month—is even harder to believe. Results for the ECLS-K and the later years of the BSS are more plausible, with mean school-year growth rates approximating the 75th percentile on the CDC’s reference curves. There may be something wrong with the 6th and 7th grade measurements in the BSS.

What could have produced such anomalous results in the BSS? One possibility is that the equipment or procedures used for height and weight measurements were inconsistent from occasion to occasion. For example, perhaps the scale used to measure weight was running a little light in the fall or heavy in the spring. Or perhaps height was measured with shoes on in the fall and shoes off in the spring.

\(^2\) Raw data for these charts, giving BMI percentiles for each month of age from 2 years to 20, can be found on the CDC’s website at [http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm](http://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm). The distribution of monthly growth rates at each age was estimated by differencing. For example, the 50th percentile of BMI growth for 6-year-old (72-month-old) boys was estimated by subtracting the 50th percentile of BMI at 71.5 months from the same percentile at 72.5 months. Percentiles for both genders combined were estimated by averaging the percentiles for boys and girls.
Alternatively, it could be that the measurement dates were misrecorded. According to the BSS data files, fall weights and heights were typically measured in mid-October, and spring weights and heights were typically measured in mid-May. Although conversations with BSS staff indicate that these were truly the dates measurements were taken, it would not be surprising if certain fall measurement were actually copied from records of school physicals taken closer to the beginning of the academic year.

It is difficult to determine what happened, because the measurements were taken more than 20 years ago, and BSS investigators and staff are no longer available to answer questions. To test the sensitivity of the results to gross errors in the data, I re-ran the model under the assumption that fall measurements were taken on the first day of the school year, a month and a half before the fall measurement date recorded in the BSS, while spring measurements were taken on the last day of the school year, a month and a half after the spring measurement date recorded in the BSS. This change to the assumed measurement schedule had material effects on the results, changing some school-year growth rates by as much as 30% and some summer growth rates by more than 50%.

What these sensitivity analyses did not change, however, was the overall contour of the results. Even when I edited the measurement dates, the BSS still displayed BMI losses during the first summer, stable BMI levels during the second summer, and substantial BMI gains during the third summer. In short, the general result that each summer was more fattening than the last is robust to substantial errors in the measurement dates. And this is a crucial pattern in implicating the non-school environment for increases in obesity.
The Effect of Initial Obesity

We now turn our attention to children who are at increased risk for obesity. The most obvious such group consists of children who are already obese at the start of the observation period. The most obvious way to estimate the trajectory of such children is to use initial BMI as a regressor, but we cannot do this because we are already using initial BMI as a dependent variable. We did, however, find ways to estimate the effect of obesity in each survey.

In the BSS, twice-yearly BMI measurements began in 6th grade, but BMI was measured on less frequent occasions before that—for example, in the spring of 5th grade. There is no methodological difficulty in using 5th grade BMI to predict growth in 6th grade and beyond, so we added to our model an indicator of obesity derived from 5th grade BMI. If 5th grade BMI was above 22 for a boy or 23 for a girl—values that correspond to the 95th percentile for 11-year-olds—then we flagged the child as obese, and we used that flag is used to predict seasonal BMI growth during the 6th-9th grade school years and the summers between.

Table 4.4 displays the result. At the beginning of 6th grade, children who had been obese in the spring of 5th grade still outweighed other children by over 9 BMI points. During 6th and 7th grade, obese children grew BMI nearly twice as fast as their normal-weight contemporaries, though during 8th grade, obese children, unlike normal-weight children, did not gain BMI at all. During summer vacations, by contrast, there were no significant differences between the growth of obese and normal-weight children.

<table>
<thead>
<tr>
<th>Reference group</th>
<th>BMI, start of 6th grade</th>
<th>BMI growth per month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6th grade</td>
<td>Summer</td>
</tr>
<tr>
<td>Reference group</td>
<td>17.94***</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Obese at the end of 5th grade vs. reference group</td>
<td>9.34***</td>
<td>0.09**</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.0001. Standard errors in parentheses.

Taken at face value, the results suggest that, at the time of the BSS, the gaps between obese and normal children opened not during summer vacation, but during the school year—with the exception of 8th grade. That is, in 1987-91, about halfway through the rise in childhood obesity, the major sources of obesity lay primarily outside the school environment—though by 1990-91 that may have been starting to change.

In the ECLS-K, we had to estimate the effect of obesity in a different way. In the ECLS-K, there were no BMI measurements before the period of seasonal measurements began at the start of kindergarten. So using an out-of-period measurement in the regression was not an option. Instead, we inferred the effect of initial BMI by using the correlations between initial BMI and subsequent growth rates. These correlations are estimated as part of the random-effects component of our basic multilevel growth model. Table 4.5 shows the estimates.
Table 4.5. School and child-level variances and correlations from the ECLS-K, 1998-2000.

<table>
<thead>
<tr>
<th>Child-level</th>
<th>Variances</th>
<th>Monthly BMI gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Initial BMI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.248***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.112)</td>
</tr>
<tr>
<td>Correlations</td>
<td>Initial BMI</td>
<td>-0.288***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td></td>
<td>Kindergarten gain</td>
<td>-0.432***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td></td>
<td>Summer gain</td>
<td>-0.377***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.016)</td>
</tr>
<tr>
<td>School-level</td>
<td>Variances</td>
<td>0.403***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.042)</td>
</tr>
<tr>
<td>Correlations</td>
<td>Initial BMI</td>
<td>-0.699***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td>Kindergarten gain</td>
<td>-0.350***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.089)</td>
</tr>
<tr>
<td></td>
<td>Summer gain</td>
<td>-0.540***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.0001. Standard errors in parentheses.

At both the school level and the child level, Table 4.5 shows that initial BMI is negatively correlated with school-year growth, but positively correlated with summer growth. In other words, children with high initial BMI tend to gain faster than other children during summer vacation but slower than other children during the school year. Taken at face value, these results suggest that by the time of the ECLS-K, near the peak of the obesity epidemic in 1998-2000, the major sources of obesity lay outside the school environment.

Figure 4.4 compares these patterns by plotting the BMI trajectories of obese and normal-weight children in both surveys.
Figure 4.4. Average school-year and summer BMI growth for obese and normal-weight children.

The growth patterns are different in the BSS and in the ECLS-K. In the BSS, the gap between obese and normal-weight children grew during the first two school years,
but shrank during the third school year and held constant during summer vacation. In the ECLS-K, by contrast, the gap between obese and normal-weight children grew during the summer vacation and shrank during the school year.

Taken at face value, these results are more or less consistent with the view that over time the sources of overweight have moved from inside to outside the school environment—from effects that are evident during the school year to effects that evident during summer vacation.

But again, the results from the BSS are on their face hard to believe. The BSS growth rates are internally inconsistent, suggesting that BMI grew twice as fast for overweight children as for normal-weight children during 6th and 7th grade, but did not grow at all during 8th grade. In addition, the BSS growth rates are inconsistent with reference growth rates derived from CDC data in Figure 4.2. According to the reference growth rates, for an 11- or 12-year-old child in 6th or 7th grade of median BMI the typical growth rate would be about 0.05 BMI points per month, and for a child at the 95th BMI percentile the typical growth rate would be about 0.08 BMI points per month. According to the BSS results, by contrast, the growth rates in 6th and 7th grade were much higher: for a normal-weight child the average growth rate was 0.10 BMI points per month, and for an obese child the average growth rate was 0.19 BMI points per month. These anomalies reinforce our concerns about the quality of the height and weight data in the BSS.

Ethnic Variation

In our next model we regressed BMI and BMI growth on children’s ethnicity. Ethnicity is an important variable since the gaps between black, Hispanic, and white children’s obesity levels have grown over time. We would therefore expect the gap
between black and white children to be smaller in the earlier BSS than it is in the later ECLS-K. We would have similar expectations for the Hispanic-white gap, but since Baltimore is almost entirely a black-and-white city there were no Hispanic children in the BSS.

The results, in Table 4.6 and Figure 4.5, are more or less consistent with our expectations. In the ECLS-K data from 1998-2000, black and Hispanic children started kindergarten with higher average BMIs than white children, and gain BMI faster than white children during summer vacation. During the school year, however, there is no significant difference between black, Hispanic, and white children’s growth. The seasonal pattern is entirely consistent with the view that the excess obesity risk of black and Hispanic children comes entirely from outside of school.

In the BSS data from 1987-91, by contrast, there were no statistically significant differences between the BMI or growth of black and white children. The effect sizes were similar to those in the ECLS-K, so the lack of statistical significance could be a result of the reduced power associated with the BSS’s small sample size. Indeed, when we extract an equally small sample of BSS-compatible schools from the ECLS-K, we also find no significant effect of race or ethnicity. In light of our limited statistical power, it is hard to say whether the lack of a significant black-white difference in the BSS is informative or not.
Table 4.6. BMI and BMI growth by ethnicity, in two cohorts of schoolchildren.

**BSS, 1987-1991**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>BMI, start of 6th grade</th>
<th>6th grade</th>
<th>Summer</th>
<th>7th grade</th>
<th>Summer</th>
<th>8th grade</th>
<th>Summer</th>
<th>9th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>White</strong></td>
<td>18.90***</td>
<td>0.13***</td>
<td>-0.13*</td>
<td>0.10***</td>
<td>-0.05</td>
<td>0.08***</td>
<td>-0.01</td>
<td>0.07***</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Black (vs. white)</strong></td>
<td>0.32</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.12^</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.07)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

**ECLS-K, 1998-2000**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>BMI, start of kindergarten</th>
<th>Kindergarten</th>
<th>Summer</th>
<th>First grade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>White</strong></td>
<td>16.085***</td>
<td>0.019***</td>
<td>0.055**</td>
<td>0.031***</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.005)</td>
<td>(0.014)</td>
<td>(0.005)</td>
</tr>
<tr>
<td><strong>Black (vs. white)</strong></td>
<td>0.370***</td>
<td>-0.006</td>
<td>0.068**</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.009)</td>
<td>(0.024)</td>
<td>(0.009)</td>
</tr>
<tr>
<td><strong>Hispanic (vs. white)</strong></td>
<td>0.464***</td>
<td>0.005</td>
<td>0.061**</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.009)</td>
<td>(0.023)</td>
<td>(0.009)</td>
</tr>
<tr>
<td><strong>Other (vs. white)</strong></td>
<td>-0.059</td>
<td>0.006</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.010)</td>
<td>(0.025)</td>
<td>(0.009)</td>
</tr>
</tbody>
</table>

**BSS-compatible ECLS-K, 1998-2000**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>BMI, start of kindergarten</th>
<th>Kindergarten</th>
<th>Summer</th>
<th>First grade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>White</strong></td>
<td>16.198***</td>
<td>0.004</td>
<td>0.108**</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.012)</td>
<td>(0.033)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Black (vs. white)</strong></td>
<td>0.118</td>
<td>-0.001</td>
<td>-0.021</td>
<td>0.057***</td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.016)</td>
<td>(0.046)</td>
<td>(0.017)</td>
</tr>
<tr>
<td><strong>Hispanic (vs. white)</strong></td>
<td>0.439^</td>
<td>0.000</td>
<td>0.028</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.017)</td>
<td>(0.048)</td>
<td>(0.018)</td>
</tr>
<tr>
<td><strong>Other (vs. white)</strong></td>
<td>0.242</td>
<td>-0.010</td>
<td>0.084</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.362)</td>
<td>(0.025)</td>
<td>(0.077)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>
Figure 4.5. The effect of ethnicity on BMI and BMI growth in two cohorts of schoolchildren.
Fuller Models

In our final model we added regressors representing age and gender as well as aspects of socioeconomic status such as income, eligibility for free or reduced price lunch, parental education, and family structure. The results are shown in Table 4.7.

The set of regressors is somewhat smaller for the BSS than for the ECLS-K. This makes little difference since, aside from ethnicity, few of the regressors has a significant effect. In the BSS, there were no significant effects except for the effect of gender, which affects initial BMI and 6th grade BMI growth, but does not affect growth in subsequent periods. In the ECLS-K, ethnicity—being black or Hispanic as opposed to white—had the largest effect on initial BMI and on summer BMI growth, but had no effect on BMI growth during the school year. This is exactly the pattern that we observed when ethnicity was the only regressor in the model. The estimated effects of ethnicity are not reduced by the inclusion of socioeconomic variables representing parental education, family structure, and poverty, and those socioeconomic variables have little consistent effect on BMI. (Having a working mother affects starting BMI and BMI growth during first grade, but not during summer or kindergarten.) In this respect, the BMI patterns are quite different from the patterns typically observed for achievement and achievement growth—where parental education, family structure, and poverty have strong effects that often explain much of the apparent effect of ethnicity.

Since the inclusion of socioeconomic regressors has little effect on the results, our findings are little changed from what they were when ethnicity was the only regressor in the model. In the ECLS-K, which reflects recent trends, the effect of ethnicity is seen only in initial BMI and in summer BMI growth—findings that clearly implicate the non-
school environment. In the BSS, which reflects older trends, neither ethnicity nor any other variable has a consistently significant effect, but that could be partly a result of the BSS’s small sample size and homogeneous population.
Table 4.7. Fuller models of BMI growth in two cohorts of schoolchildren.

### BSS, 1987-91

<table>
<thead>
<tr>
<th>Reference group</th>
<th>BMI, start of 6th grade</th>
<th>6th grade</th>
<th>Summer</th>
<th>7th grade</th>
<th>Summer</th>
<th>8th grade</th>
<th>Summer</th>
<th>9th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference group</td>
<td>18.32***</td>
<td>0.10***</td>
<td>-0.07</td>
<td>0.06*</td>
<td>0.03</td>
<td>0.08**</td>
<td>-0.01</td>
<td>0.07**</td>
</tr>
<tr>
<td>Black</td>
<td>0.29</td>
<td>(0.56)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.02)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Female</td>
<td>1.19**</td>
<td>(0.44)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.07)</td>
<td>(0.02)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Female</td>
<td>0.29</td>
<td>(0.41)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Age, start of 6th grade</td>
<td>0.29</td>
<td>(0.49)</td>
<td>(0.02)</td>
<td>(0.07)</td>
<td>(0.03)</td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

**Note.** The lunch variable is coded quantitatively as follows: full price=0, reduced price=2/3, free=1. With this coding, the variable has an approximately linear relationship with BMI.

### ECLS-K, 1998-2000

<table>
<thead>
<tr>
<th>Reference group</th>
<th>Initial BMI</th>
<th>Kindergarten</th>
<th>Summer</th>
<th>First grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother has not finished high school</td>
<td>-0.102</td>
<td>0.015</td>
<td>-0.049</td>
<td>0.016</td>
</tr>
<tr>
<td>Mother finished high school but not college</td>
<td>0.104</td>
<td>0.001</td>
<td>-0.006</td>
<td>0.010</td>
</tr>
<tr>
<td>Single parent</td>
<td>-0.022</td>
<td>-0.001</td>
<td>-0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>Mother works</td>
<td>0.238**</td>
<td>0.004</td>
<td>-0.016</td>
<td>0.017**</td>
</tr>
<tr>
<td>Household income (thousands, square root)</td>
<td>-0.018</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>Age at start of kindergarten (months)</td>
<td>0.020*</td>
<td>0.002*</td>
<td>0.001</td>
<td>0.002***</td>
</tr>
<tr>
<td>Gender (female=1)</td>
<td>-0.215**</td>
<td>0.012*</td>
<td>0.012</td>
<td>0.003</td>
</tr>
</tbody>
</table>

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

The results above are generally consistent with our core hypothesis, suggesting that the increase in child obesity is due primarily to changes outside rather than inside the school environment. Our review of changes in school practices found that they could account for at most 20% of the increase in children’s weight since the early 1980s. And our comparison of seasonal growth patterns early and late in the obesity epidemic suggested that we have moved from being a country where children gained little if any BMI during summer vacation to being a country where children, especially those prone to obesity, gain BMI much faster during summer vacation than during the school year. In short, schools have changed relatively little during the rise in child obesity. It is the non-school environment that has grown dramatically more fattening.

Our confidence in saying this is tempered by the limited availability of data on seasonal BMI growth. We had to infer secular trends from two not entirely comparable studies. One study, the ECLS-K, was a national sample of 5,380 children whose BMI was measured seasonally from kindergarten through first grade in 1998-2000. The other study, the BSS, was limited to 838 children in Baltimore whose BMI was measured seasonally from 6th to 9th grade in 1987-91. And some of the BMI growth rates in the Baltimore study were so inconsistent with national standards that it seems likely some weights and heights were measured with an inconsistent protocol or miscalibrated equipment.

Our results are consistent with Japanese studies showing that obese children gain weight fastest during the summer, while children of lower weight gain weight fastest during the school year (Masako Kobayashi and Maiko Kobayashi 2006 and Japanese citations therein). Because Japanese obesity levels lag those in the U.S. (Matsushita et al.
most Japanese children today lose BMI in the summer. As Japanese obesity levels continue to rise, however, this pattern is likely to change.

Our results are harder to reconcile with a recent study of 865 middle schoolers in the Pacific Northwest, which found that those children tended to lose BMI during the summer (Daratha et al. 2009). However, our exploratory analyses of the ECLS-K (not shown) found that seasonal BMI patterns are different in the Western census region. Perhaps that accounts for recent results from the Pacific Northwest.

In general, the evidence in this chapter suggests that schools have relatively little to do with the rise in child obesity. This conclusion may seem surprising, but it is consistent with the view, pursued in this dissertation, that schools serve to stabilize and equalize society. Given that the obesity epidemic has emerged rapidly and affected some groups much more severely than others, it is not surprising that little of the obesity epidemic can be pinned on schools.

The finding that schools have not caused the obesity epidemic also resonates with some broader observations. The first observation is that the obesity epidemic is not limited to children; it has also affected adults. The rise in adult obesity began in the early 1980s, around the same time as the rise in child obesity (Ogden et al. 2007; Cutler et al. 2003). Given the concurrent rise of child and adult obesity, it seems likely that both have a common cause, and it seems unlikely that the cause is school. In fact, adults who are over 35 today, about 30% of whom are obese, attended school at a time when the child obesity rate was only 5% (Ogden et al. 2007). Something other than school has turned past generations of thin children into overweight adults.
A second observation is that, although the U.S. leads the developed world in obesity, most other developed countries also have quickly rising obesity levels, and some, such as the U.K. and Australia, are not far behind the U.S. (Bleich et al. 2008). Obesity is also a serious problem in some less developed countries, such as Saudi Arabia, and in the more affluent social strata of India. In short, rising obesity is a global problem that affects both adults and children. School practices that affect only the children of a single country can explain only a small part of the problem.

Policy Implications

The findings in this chapter have either pessimistic or optimistic implications for policy, depending on how they are read. The pessimistic reading is that, given that schools did relatively little to start the child obesity epidemic, it is unlikely they can do much to reverse it. Even if school practices were rolled back to the 1970s—with fewer school vending machines and more time allocated to recess and physical education—the non-school environment would remain so fattening that only a fifth of the increase in child obesity would be reversed.

A more optimistic reader, on the other hand, would notice that we have in fact discovered an intervention that restrains children’s weight gain. That intervention is school. When children today are in school, they are much less vulnerable to excessive BMI gain than they are elsewhere. In fact, children’s school environments are only slightly more fattening today than they were before the obesity epidemic.

The effectiveness of school as an obesity preventive is limited by the fact that U.S. children spend only 6.5 hours per day, 180 days per year, in their school environments. It stands to reason that lengthening the amount of time that children spend
in school would reduce the problem of obesity. Indeed, our analyses suggest that today about one-half of young schoolchildren’s annual BMI gain, and nearly all of the excess BMI gain by obese children, takes place during the two-and-a-half months of summer. A brief calculation suggests that if the school year were lengthened and the summer vacation were shortened by one month, about 40% of obese children’s annual weight gain would not take place. A straightforward way to test this conjecture is to test whether the schools that have considerably more than 180 school days per year reduce their student’s weight gain. Such extended-year schools are rare in general (there are none in the ECLS-K), but they do exist in the charter sector. Obesity prevention may not be a sufficient reason to extend the school year, but it could be an additional benefit.

Extending the school year is probably not the only way to prevent summer weight gain. If preventing obesity means reducing children’s exposure to their usual summer environments, then increasing the availability of summer camps, summer schools, or other summer programs may be an effective intervention as well. A number of studies have shown that camps specially designed for weight loss can help obese children (e.g., Walker et al. 2003). It may be that ordinary summer camps would help as well. In our analysis of the ECLS-K, we looked briefly at children who attended summer camps or summer schools between kindergarten and first grade. We found negligible effects, but the summer camps and summer schools attended by very young schoolchildren are very short, both in days per summer and in hours per day. It would be more informative to look at older children who enroll in all-day summer school or attend overnight summer camp for a month or more.
Our analysis also helps us to predict which school-based interventions are worth pursuing. Although our analysis suggests that schools have caused only 20% of the rise in child obesity, 20% is not a trivial contribution, and its source is well-understood. Schools’ contribution to obesity does not come from regulated breakfasts and lunches, nor does it come from exercise programs such as physical education, recess, and organized sports. Instead, our analysis suggests that schools’ only substantial contribution to the child obesity epidemic comes from the sale of unregulated competitive foods—a practice that has been increasing because food vendors offer schools a share of the revenue. A simple intervention would be to offer schools replacement revenue if they stop selling competitive foods. This intervention could be tested in a randomized field trial.

These policy recommendations emerge from the sociological principles that underlie this dissertation—namely that schools are relatively uniform, relatively stable, and quite different from the rest of children’s lives. In evaluating influences on children’s weight, the difference between one school and another do not matter that much, and neither to the differences between schools’ practices today and schools’ practices a generation ago. The difference that matters most is the difference between what children do when they are in school and what they do after the bell rings.
This dissertation is part of a larger research agenda whose purpose, broadly speaking, is to identify the major sources of inequality and of secular change in children’s physical and intellectual growth. Rather than trying to identify all the specific variables that might be influential, I take a broader view and try to distinguish the total contributions of school and non-school influences. To separate school from non-school influences, I rely primarily on seasonal comparisons between school-year and summer growth. I also rely secondarily on analysis of historical change and on evidence from genetic research.

Like the rest of my research agenda, the results in this dissertation are consistent with the two generalizations that I elaborated in the introduction. First, the inequalities that schools visit on children today are minor compared to the inequalities that children experience in other parts of their lives. Second, the pace of change in schools today is glacial compared to the pace of change in the rest of society.

These generalizations have not always been true. To the contrary, schools were once deeply unequal, and schools have experienced periods of rapid change. Probably the greatest inequalities and the fastest changes have been experienced by black children in the South. But since the near equalization of black and white Southern schools by 1950, schools have been relatively equal, accounting for little of the inequality that is observed
in children’s achievement and attainment. And since the integration of Southern schools plumed in the early 1970s, schools have been relatively slow to change. Probably the biggest change in the U.S. school system since the early 1970s has been the change in many states’ school funding formulas. But this change has been gradual, proceeding state by state, and it has been a change in the direction of greater equality. What this means is that from the early 1970s to the present—a period covering nearly all modern education research since the Coleman report—U.S. schools have constituted a system that is relatively stable, relatively fair, and growing slowly fairer to the degree that it is changing at all.

Results

The results in this dissertation are consistent with the view that schools are relatively stable and relatively fair. In my chapter on child obesity, I find that schools have not changed enough since 1980 to explain more than 20% of the increase in obesity among school-age children. And I report that the school practice that has contributed the most to child obesity—the marketing of junk food outside of the regulated school lunch and breakfast programs—is most prevalent in relatively affluent schools, and so probably can’t explain why obesity has grown fastest among black and Hispanic children. In short, neither the rapid change in obesity prevalence nor the unequal distribution of obesity across racial and ethnic groups has much to do with schools.

In my chapter on year-round schools, I find that even a reform as minor as rearranging the school calendar can provoke organized resistance, and that such a reform has a negligible effect on increasing the achievement of disadvantaged students. The finding is consistent with a long history of studies suggesting that the reforms typically
implemented by schools tend to be minor and have at best small effects on achievement (Hanushek 1997)—not small by every scholar’s standards (e.g., Hedges, Laine, and Greenwald 1994), but certainly small in comparison to the achievement gaps between racial, ethnic, and socioeconomic groups.

**Implications**

The results and perspective presented in this dissertation have implications both for sociological research and for education policy.

**Implications for Sociology**

Although sociologists have produced some of the strongest evidence for schools’ relative fairness (Coleman et al. 1966; Alexander 1997; Alexander and Entwisle 1996; Heyns 1978; Downey et al. 2004), many sociologists are influenced by the idea that schools are unfair. Influential sociological theories hold that the education system is used to “maintain inequality” (Raftery and Hout 1993) or to “reproduce the class structure” (Bowles 1971; see also Bourdieu and Passeron 1970; Willis 1977; Bowles and Gintis 1976). These theories are reinforced by touchstone narratives that are commonly assigned to orient undergraduates to the sociological perspective on schools (Rist 1970, 1973; Kozol 1991, 2006). While most practicing sociologists hold more nuanced views, many of the educational topics that sociologists choose repeatedly to study—such as tracking, ability grouping, teacher expectations, and the distribution of material resources across schools—hold interest largely because of the background narrative that schools are unfair.

A focus on exposing unfairness can be healthy if it drives institutions to become fairer, but it can unbalance our interpretation of schools’ role in society. We may be
hyper-vigilant to examples of unfairness in the school system, and less alert to evidence that schools are used to remedy inequality. To return to the introduction, one example of a story that sociologists have largely overlooked is the near-equalization of resources between black and white schools in the South from 1910 to 1950—before the *Brown* decision and the Coleman report. Another example is the growing use of state budgets to level resources between rich and poor school districts, starting with the *Serrano* decision in 1971. These are stories of social movements, distributive justice, race, and class—stories that should be attractive to sociologists. Yet perhaps because these stories show schools becoming fairer and fairer over time, the pleasure of telling them has fallen to economists (Margo 1990; Hoxby 2001; Card and Abigail Payne 2002; Card and Krueger 1996; Corcoran et al. 2004).

*Implications for Policy*

The implications for policy and intervention are even stronger. For decades, educational stakeholders have assumed that the school system is a major source of inequality, and that inequality can be substantially reduced by increasing resources or changing incentives in the schools. This assumption has informed government policy for decades, from the Lyndon Johnson administration’s Title I programs to the Obama administration’s Blueprint for Reform. The same assumption motivates nonprofits such as Teach For America, which believes that the educational opportunities of poor children constitute “a national injustice,” and seeks to remedy that injustice by recruiting elite college graduates to teach in high-poverty classrooms (Teach For America 2010).

These programs are well-intended, but the results and perspective of this dissertation suggests that they will have limited effects. If the school system has become
increasingly fair, if the remaining sources of inequality lie primarily outside of school walls, then attempts to make schools still more equal are likely to have reached a point of diminishing returns. Consistent with this prediction, evaluations of programs from Title I to Teach For America have generally had disappointing results (Decker et al. 2004; Carter 1984). The programs do have positive effects, but those effects are small compared to the inequalities that the programs seek to remedy.

Yet the perspective in this dissertation should not be taken as an excuse for pessimism or inaction. Although the perspective explains the disappointing results of past reforms, and suggest pessimism about most of the reforms being tried today, the same perspective also suggest the kinds of reforms that might be expected to work. If schools are fairer than the rest of children’s lives, then reductions in inequality can best be achieved either by increasing the amount of time that children spend in school, or by intervening in other areas of children’s lives. The most successful charter schools pursue one or both of these avenues. For example, one of the five pillars of charter schools in the Knowledge is Power Program (KIPP) is “more time”—more time in the form of an extended school day that lasts until 5 p.m., an extended school week that includes two Saturday mornings per month, and an extended school year that runs three weeks into the summer (Mathews 2009). Another well-known charter school, the Harlem Promise Academy, also has an extended school year; in addition, the Promise Academy is part of the wider Harlem Children’s Zone, a broad net of programs that intervenes in aspects of children’s non-school lives, including a Baby College that helps parents learn to enrich their children’s home lives years before school begins (Tough 2009). Randomized evaluations indicate that both the KIPP schools and the Harlem Promise Academy, unlike
most charter schools, have had large and positive effects on their students’ achievement (Dobbie and Fryer 2009; Angrist et al. 2010). Our finding that obesity increases primarily when children are out of school (Chapter 3) suggests that extended-year schools might help children maintain a healthy weight as well.

Given the resistance that parent and industry groups have mounted against alternative school calendars (Chapter 3), it seems likely that top-down mandates to extend the year of all public schools in a state or city will encounter strong resistance. Middle class parents see no reason to extend the school year; teachers unions wonder who will pay their members for the extra work; and summer camp and amusement park operators see a threat to their business. Powerful politicians including the governor of Ohio and the mayor of Chicago have expressed a desire to extend the school year (Strickland 2009; City of Chicago 2006), yet so far neither Ohio nor Chicago has added a day to its standard school calendar.

Rather than mandating that all schools lengthen their year, a more promising approach is to use the charter system to offer interested families, particularly in poor areas, the choice of enrolling in a school with an extended academic year. The KIPP schools, for example, do exactly this, and the fact that KIPP’s enrollment lotteries are regularly oversubscribed suggests that poor neighborhoods have considerable unmet demand for extended-year schools. Continued expansion of the KIPP network, which currently runs 86 schools, would be desirable. Future research should examine whether KIPP’s effectiveness comes primarily from its giving students “more time,” or whether other pillars of KIPP’s approach are equally important.
To answer the question in the title of this dissertation, if we are concerned about inequality or unhealthy social change, schools are not the problem. Instead schools are a potential solution, and we could make more of that solution by increasing the time that the neediest students spend in school. Instead of trying to squeeze the remaining unfairness out of the school system, it would be more productive to use extended-year charter schools to counter the unfairness in the rest of children’s lives.
References


174


City of Chicago. 2006. “Mayor Daley Announces Full-Day Summer Programs For More Than 1,000 Students at 11 Englewood Schools.” http://egov.cityofchicago.org/city/webportal/portalContentItemAction.do?blockName=Mayors+Office%2FJune%2FI+Want+To&deptMainCategoryOID=-536882034&channelId=0&programId=0&entityName=Mayors+Office&topChannelName=Dept&contentOID=536941191&Failed_Reason=Invalid+timestamp,+engine+has+been+restarted&contentTypeName=COC_EDITORIAL&com.broadvision.session.new=Yes&Failed_Page=%2Fwebportal%2FportalContentItemAction.do&context=dept (Accessed June 8, 2010).


Swinburn, Boyd, Gary Sacks, and Eric Ravussin. 2009. “Increased food energy supply is more than sufficient to explain the US epidemic of obesity.” *Am J Clin Nutr*
Teach For America. 2010. “A national injustice.”


