



Maryland Population Research Center

WORKING PAPER

Equalizers or Enablers of Inequality? A Counterfactual Analysis of Racial and Residential Tet- Score Gaps in Year-Round and Nine-Month Schools

PWP-MPRC-2014-015

December 2014



Authors:

Odis Johnson, Jr.
University of
Maryland

Michael Wagner
University of
Maryland



EQUALIZERS OR ENABLERS OF INEQUALITY? A COUNTERFACTUAL ANALYSIS OF RACIAL AND RESIDENTIAL TEST-SCORE GAPS IN YEAR-ROUND AND NINE-MONTH SCHOOLS

Odis Johnson, Jr., PhD.
Chair & Associate Professor
African American Studies Department
University of Maryland

Michael Wagner, PhD.
Faculty Research Associate
African American Studies Department
University of Maryland

ABSTRACT

Persistent racial/ethnic and residential disparities in test-scores suggest schools fail to serve as society's great equalizers. In fact, studies show that racial inequalities in test-scores grow mostly while children are in school rather than when they are not. However, these studies rarely consider the qualities of children's neighborhoods as contributors to educational stratification, or use counterfactual modeling to strengthen causal inferences. Using ECLS-K data, an optimized matching algorithm and HLM, we pair children who attend year-round schools to those in 9-month schools. We then explore 1) whether there are mean differences in the reading and math performances of 1st graders attending year-round and 9-month schools; 2) if racial and residential differences in children's test-scores exist between the schooling types; and, 3) if neighborhood effects related to children's performances strengthen or weaken as their exposure to schooling increases. In contrast to previous claims that schooling increases test-score inequality, we find that year-round schooling is related to less racial inequality, most notably for African Americans, and that schools appear essential to the delivery of neighborhood influences on math and reading test-scores. The study implies that schools can be greater equalizers according to race than they are currently, but may also serve as enablers of neighborhoods' impressive stratifying effects.

GREAT EQUALIZERS?

Racial stratification in U.S. schools has a long and unclear history. Despite earlier studies that suggested schools functioned to equalize test-score differences among racial groups (Heyns 1978; Entwisle, Alexander and Olson 1992), recent evidence relying on ECLS-K data has suggested that schools fall short of this goal. Studies using these data have revealed a few commonalities. First, the overwhelming majority of studies, if not all of them, show that African Americans have a slower rate of growth in reading (reardon 2003; Downey, von Hippel and Broh 2004; Benson and Borman 2010; Johnson 2014) during the kindergarten school year. Their slower growth rate amounts to non-trivial losses in comparison to that of other children. As Benson and Borman (2010) have shown, these losses of about 2 reading test points per month totaled a loss of nearly a full month of academic year learning in both kindergarten and 1st grade. Johnson reported that the 1st grade reading loss for African Americans may have been as large as 1.3 months of academic year growth. Second, African Americans fell further behind in math as kindergarten

progressed until they trailed white children, according to Condon's (2009) estimation, by as much as 1.82 months of school time growth at the conclusion of 1st grade.

Test-score gaps are not unique to African Americans, however. Hispanics have been found to experience school-time setbacks of their own in kindergarten math (reardon 2003; Downey, von Hippel and Broh 2004; Benson and Borman 2010) and 1st grade reading (Johnson 2014; Downey, von Hippel and Broh 2004; Lee et al., 2004). Benson and Borman (2010) and Johnson (2014) found that these shortfalls amounted to approximately .8 months of academic year growth in math and reading. The research record is mixed for Asian/Pacific Islanders. Only Benson and Borman (2010) observed 1st grade losses in both subjects while Downey, von Hippel and Broh (2004) found that they gained in both subjects. Otherwise, Asian/Pacific Islander test-scores have appeared insignificantly different than those of white children.

Rather than equalizers, some stratification theories have suggested that schools function as conduits through which children's unequal backgrounds, abilities and cultural predispositions are projected in their education outcomes (Sorokin 1959; Parsons 1959; Bourdieu 1977). If this alternative functional explanation were plausible, racial/ethnic differences would be apparent in what children learned prior to the start of their educational careers and possibly grow as their summers are spent in their family and residential contexts. However, research has shown no significant reading gaps existed for African Americans before school's start (Entwisle, Alexander and Olson 1992; Johnson 2014; Benson and Borman 2010) and that none developed during the summer after controlling for other social background characteristics (Lee et al., 2004; Downey, von Hippel and Broh 2004; reardon 2003; Burkham et al., 2004; Johnson 2014; Benson and Borman 2010). Evidence related to the summer learning of Hispanic and Asian/Pacific Islander children is even more consistent across analyses of the ECLS-K, with studies having revealed no significant losses in either reading or math (reardon 2003; Lee et al., 2004; Downey, von Hippel and Broh 2004; Burkham et al. 2004; Benson and Borman 2010; Johnson 2014). A meta-analysis of summer effects produced no significant moderating effects for race after controlling for SES (Cooper et al., 1996). Having found stratification expands in schools after accounting for social background differences when it does not in non-school contexts, research has implied that U.S. schools have been neither great equalizers nor mere conduits, but have functioned instead like *great dividers* in regards to race/ethnicity.

Using ECLS-K studies as points of departure, we wondered if similar racial/ethnic differences would manifest within an alternative form of schooling, year-round education (YRE). While racial/ethnic stratification in year-round schools remained a primary concern of ours, we also sought to shed light on the importance of residential stratification to test-score differences since its effects have been shown in some studies to surpass the inequality found among races (Benson and Borman 2010; Johnson 2014). These issues are explored in this study within a counterfactual framework that contrasted year-round and traditional 9-month schools. Our most noteworthy results revealed that racial/ethnic test-score inequality was less prominent in year-round schools, most notably for African Americans, and that schools appeared essential to the delivery of residential influences on test performances. We close with a discussion of the study's theoretical and policy implications.

RACE/ETHNICITY AND YEAR-ROUND AND EXTENDED-YEAR SCHOOLS

If schools have functioned to increase racial/ethnic stratification, one might question how could giving children more time in them be a helpful policy approach? This is indeed an important question to ask since the percentage of the U.S. school-age population enrolled in year-round schools, while low, has increased steadily from 3.5 percent in 2005 to 4.1 percent in 2012 (U.S. Dept. of Education 2013), and has President Obama as one of its proponents. To this question, YRE advocates might answer that while they are often described as "year-round," these schools usually provide instruction for approximately 180 days just like traditional 9-month schools. The difference in a year-round schedule is that the days that children attend are spread out within the calendar year, typically with three cycles that include 60 days of instruction separated by 15-day intersessions. YRE proponents, such as the National Association for Year-Round

Education (see nayre.org) claim that these modified schedules have often led to less academic slippage than would have occurred during a typical summer recess and a reduced amount of time that teachers dedicate to review last year's lessons at the start of the traditional academic year (Cooper et al., 2003). Therefore, the potential benefits of YRE revolve around this "spacing effect" not just the amount of school exposure. The key question becomes whether spacing has altered the growth of racial/ethnic gaps in test-scores.

Unfortunately evidence about the impact of YRE and spacing-effects on learning and racial stratification is quite inconclusive. For example, Reece (2000) tested over 700 traditional and year-round school students at the beginning and end of the summer in reading, math, spelling and writing. While most of the estimates suggested summer knowledge retention was higher for year-round students, the analysis did not cover the school year to see if what children retained led to stronger academic year growth. In another study, McMillen (2001) compared 106 year-round schools to 1364 9-month schools and found no overall significant benefit for YRE. Similarly, von Hippel (2007) reported that after separating academic-year growth from summer growth, year-round children's rate of test-score growth was no better than their 9-month counterparts, offering little support for the spacing effect hypothesis. Although McMillen (2001) and von Hippel's (2007) research considered race, the samples they used lacked diversity. McMillen addressed this problem by using a white-non-white binary in his analysis, placing out of reach a fuller accounting of racial stratification. He concluded that inasmuch as race moderated YRE outcomes, white students in year-round schools had test performances about .04 standard deviation units stronger than their traditional school counterparts. There was no treatment dissimilarity for students of color. Von Hippel (2007) in contrast claimed that significant losses existed among the two school types for African American kindergarteners relative to Hispanics, who were the referent group in his analysis. However, von Hippel's year-round sample was only 5 percent African American which would yield an insufficient subgroup number no greater than 20 in each of the periods he evaluated.

Despite the official school schedule, some children in year-round schools do manage to experience more days of schooling. First, students who have performed less well during an instruction cycle often receive additional help during the intersession, in effect increasing the actual number of days that they attended school. Intersessions then give teachers an opportunity to meet the educational needs of children that would not have occurred during the traditional school year (Cooper et al., 2003). So while school dosage has been increased for students, the additional time is spent on remediation. Cooper's (2003) synthesis of evaluations suggested little can be concluded about remedial time, since most of them did not include controls for the number of days of instruction to assess remediation's impact on racial stratification.

Second, there are year-round schools that extend the school year to include over 200 days. Gandara and Fish (1994) examined extended-year effects for calendars that had approximately 223 days of school. Although the authors found extended-year benefits for children's learning and more positive parental perceptions of their schools, the benefits of increased school exposure are not easily distinguishable from the other significant changes that were implemented simultaneously, among them cooperative teaching, reduced class size and increased teacher salaries. Moreover, the description of the criteria and matching process leaves unknown to what extent the control and experimental schools and the children within them were similar.

Third, an extended school-year is available to most children in the U.S. through the provision of summer school. Summer school is offered to address losses that occurred in the previous academic-year, or to prevent anticipated losses in knowledge over the summer that Cooper et al., (2003) reported can reach as much as .10 standard deviation units. Summer school programs combat these dynamics by adding between three to six weeks of additional instruction, though programs typically have shorter school days and oftentimes 4-day weeks. Research on the effects of summer programs has been generally positive. Cooper et al., (2000) summarized the results of 93 studies of summer school and reported that summer programs, whether focused on remedial, accelerated, or enriched learning, had a positive

influence on children's learning outcomes. Those that focused on remediation and promotion included 121 separate independent samples, 95 of them having shown positive effects of summer school on all comparison dimensions. However, this synthesis summarized no evidence regarding racial/ethnic differences in summer school participation relative to non-participation.

NEIGHBORHOOD EFFECTS AND SCHOOL DOSAGE

Commentary that has weighed the importance of schools and neighborhoods to achievement ranges from suggestions that the two contexts are quite synonymous (Jencks and Mayer 1990) to claims that one is more important than the other (Arum 2000). Jencks and Mayer (1990) for example once suggested that the only difference between neighborhoods and schools was that one assigned grades and the other did not. While Jencks and Mayer (1990) argued that the effects of the two contexts were somewhat indistinguishable in research, Arum (2000) later claimed that features of the school's institutional community had much more to do with learning than its geographic one. Since opinions about the relationship of neighborhoods and schools varied so greatly, Johnson (2012a) reviewed the available neighborhood effects studies (83 in all) to catalogue these perspectives. Among those identified was a subset of frameworks that held assumptions similar to those of Jencks and Mayer's (1990) that Johnson branded "ecological correspondence theory" and others that he argued fit Arum's (2000) "autonomous institutional" perspective of exogenous school processes within neighborhoods.

Rendon (2014) sought to test the assumptions of these neighborhood-school models on adolescents' odds of dropping out of school. Although she found that neighborhood conditions increased the odds of dropping out of school, these odds became greater once school characteristics were considered and remained significantly higher among African Americans. Her observations not only implied that an appropriate understanding of neighborhood-school interactions would escape simple endogenous-exogenous binaries, but that neighborhoods might in fact rely on school mechanisms to deliver a neighborhood effect net of school effects. If this neighborhood-school interaction were true, then lengthened school years might actually lead to more residential stratification in education. Our analysis extends this line of thinking by observing how variation in children's amount of exposure to schooling or "dosage" corresponds to residential variation in test performances.

In yet another model of neighborhood-school processes, Johnson (2012a) observed that neighborhood effects research has often assumed that neighborhood processes are static within the calendar year. He then borrowed from seasonal learning research (Entwisle, Alexander and Olson 1992) in proposing a "faucet theory model" of neighborhood-school relationships to determine "whether schools stop the flow of neighborhood influences in educational production when in session (as suggested in the autonomous institution framework) or if the flow through an otherwise permeable educational system reduces to a trickle during cooler seasons due to a lull in the neighborhood's social activity" (p. 491). Extending from Johnson's (2012a) observation, we have identified seasonal change in the social organization of neighborhoods as a potential confounder of school calendar influences on racial group's test-score trajectories.

One study employing the faucet model estimated the impact of neighborhoods' economic and racial make-up in each season finding that economic segregation was the most salient social background determinant of math and reading inequality during the summer but not in the academic year (Benson and Borman 2010). However, Benson and Borman's analysis was unable to tell us if these significant summer neighborhood effects would retain their significance had children received some kind of summer instruction. Imposing variation in children's exposure to schooling during the summer, as our analysis has, clarifies whether summer test-score gains or losses extend from the absence of schooling or summer time changes in the neighborhood's social organization.

Testing neighborhood fluctuations is precisely why we added to the examination of neighborhood qualities the dimension of social disorganization, or the breakdown of social control indicated by criminal

behavior. This residential feature we thought would be best suited to detect temporal fluctuations in areas' social organization unlike their demographic compositions, such as race and SES, which are nearly static throughout the year. Summer increases in crime, violence and burglary have sadly become predictable (Lauritsen and White 2014), and with noted consequences for children. For instance, Sharkey and Sampson (2010) have demonstrated the occurrence of a homicide in African American's census block within a week of their assessment reduces their reading and vocabulary performances by at least a half standard deviation unit. Johnson's (2014) analysis found temporal fluctuations associated with the seasons in which neighborhood problems with burglary had their strongest connection with reading during the summer, and achieved a magnitude greater than any other social background characteristic. While others have investigated the temporal effects of neighborhood disadvantage in relation to dropping out and graduation within a counterfactual framework (Harding 2003; Wodtke, Harding and Elwert 2011), no counterfactual study has yet assessed the seasonal quality of neighborhood crime effects related to test-scores.

RESEARCH DESIGN AND METHODS

We start our analysis with the general question: do differences materialize in year-round and 9-month school students' test-score growth during the calendar year, 9-month academic term and summer? To pursue this question, we constructed a research design that features three key components. First, we utilized a counterfactual approach, shown in Table 1, in which year-round and 9-month schooling represent two treatment conditions experienced by similar children, the former we refer to as the experimental and the latter the control. Examining both schooling schedules strengthens the causal inferences that we hope to make about school exposure's influence on test-score performances because the counterfactual, or alternative educational experience, is also tested. Second, we carried out this counterfactual approach within a seasonal learning framework where test-score growth within both treatments was measured for the calendar year, summer and traditional 9-month academic year. The separate estimation of growth in distinct yearly, 9-month and summer periods allowed us to understand at what point in the calendar year were test-score gains likely and if the magnitude of growth implied that any summer slowdown or spacing effect had occurred.

--TABLE 1 NEAR HERE--

We also pose a second question regarding variation in how race/ethnicity related to test-scores across school schedules and seasons. More specifically, we ask as subset of related questions: Are there within-race/ethnicity test-score differences between the two school types, in which form of schooling is there less test-score inequality between racial/ethnic groups and does growth in these within- and between-race/ethnic gaps differ in the 9-month academic term and summer? Again, the separate estimation of racial/ethnic differences in distinct periods permitted the addition of covariates that accounted for the type of summer educational experience children in year-round school received (i.e. summer school or year-round school) and other social background factors.

Included among these other social background factors are residential dimensions. Hence, our third question reflects the nesting of the counterfactual design within residential areas (presented in Figure 1), and considers how might residential effects on children's test-outcomes correspond to the variation in school exposure found among schooling schedules and seasons. With this question, we addressed Johnson's (2012a) speculation regarding the degree to which the residential effects (shown at the left side of figure 1) related to racial/ethnic test-score gaps (right side of figure 1) are negotiated by children's exposure to the various schooling calendars (shown in the middle column of figure 1). In estimating the

neighborhood effect for the alternative educational treatment within the bottom middle block in particular, we were able to infer whether variation in children's scores is due to the absence of schooling or seasonal fluctuations in neighborhood activity. For example, in Table 1, summer test-score change is assessed in the presence and absence of schooling because 27.44% of experimental children's summer time is spent in school while control group children have no school days. Should summer losses occur for children that experienced no schooling rather than for those who did, we can then be more certain that those declines were due to the absence of schooling rather than summertime change in the social organization of neighborhoods, which is a potential confounder in existing studies of summer learning. Variation in the number of instructional days between year-round and 9-month schools is also exploited in estimating 9-month test-score growth, where experimental children are in school 22.28 days fewer than are their 9-month counterparts.

--FIGURE 1 NEAR HERE--

DATA

Data from the Early Childhood Longitudinal Study, Kindergarten Cohort 1998 - 1999 (ECLS-K) are ideal for this study because they enabled the measurement of summer change in test-scores. The National Center for Education Statistics (NCES) collected data about the families, schools, neighborhoods and activities of approximately 22,780 children, who were chosen at random from 1277 randomly selected public and private kindergarten programs.¹ Of this total sample, 30% were randomly selected to take assessments near the beginning and end of kindergarten and 1st grade. While this study's analysis was not limited to the children that participated in this subsample, only children in this subsample have a fall 1st grade test-score. Similar to the strategy employed by Downey, von Hippel and Broh (2004), we used the assessment dates and the beginning and end-dates of 9-month schools to calculate children's average rate of test-score growth between assessments.² Using this average rate of growth and under our valid assumption that growth between waves 3 and 4 was linear, we extrapolated for children who were missing scores what their test-scores would have been at wave 3, the start of 1st grade.

We focused on waves 2 – 4, which span from the end of kindergarten to the end of 1st grade, as the year-round schooling period for two reasons. First, over 35% of Hispanics did not take the fall kindergarten assessment in reading so defining year-round as covering the kindergarten year and following summer would result in a smaller sample. Second, studies have shown that more test-score growth happens during 1st grade than in kindergarten making it one of the more pivotal periods in which to gauge academic differentiation (Downey, von Hippel and Broh 2004; Benson and Borman 2010; Johnson 2014).

The success of the student matching procedures that we describe later depended on the completeness of the data. Therefore, in cases where information regarding children's background and educational experiences was missing, we added values from identical measures collected in earlier or later data collection waves when there was little reason to expect those qualities to have changed over time (e.g. gender, race, pre-school experience). However, some children did not have parent data or any test-scores and were subsequently removed from the sample. Once these eliminations were made, we removed Native Americans because fewer than 20 attended schools year-round, a number too small to yield meaningful analysis results.³ Finally, our consideration of academic calendars and the social organization of residential

¹ This analysis uses a panel weight (C24PW0) to compensate for the unequal probabilities of selection inherent in the ECLS-K's stratified sampling design.

² I used the beginning and end school dates supplied by school administrators, and when those dates were not provided, those given by parents.

³ NCES requires rounding to the nearest 10 when discussing the restricted use data sample sizes.

areas required that we remove from the sample children that changed schools or neighborhoods during the study period.

PROPSENSITY SCORE AND MATCHING PROCESS

Since we have not randomly assigned children to schools of any type, reasons for their enrollment in either year-round or 9-month schools remain unknown and possibly related to their test-scores. Hence, there may be many reasons why the test-scores of children vary between the two conditions that would make conclusions about the relevance of educational exposure quite premature and possibly incorrect. To address these challenges and strengthen the causal inferences we would like to make about school exposure, we matched children in 9-month schools to those in school year-round relying on propensity scores as child matching estimators. Propensity scores represent the predicted probability that individuals with certain qualities will experience a treatment when assignment is essentially nonrandom. Using individuals' characteristics, we estimated a propensity score to identify those with a similar probability of choosing the experimental or control condition, irrespective of their true condition assignment. Once individuals with a similar propensity score were identified and compared, there would be enough similarity in their pre-treatment characteristics to theoretically allow observed discrepancies between them to extend from the differing treatments they experienced.

To start, no theoretical basis exists on how to select the best or ideal number of pre-treatment dimensions for calculating matching estimators (Heckman et al. 1998). Rather than piling on numerous qualities to generate a matching estimator, we included qualities until balanced experimental and control group properties was achieved. We preferred this efficient approach because this study offers a multivariate analysis that includes pre-treatment characteristics in order to understand variability within each condition, and because we have taken additional steps to optimize our matching process.

We reasoned that the qualities used to estimate our logit modeled propensity scores should span the multiple units of analysis and contexts examined in this study. At the individual level, we included the traditional markers of social differentiation, namely children's gender and their parents' social class and marriage status. Regarding their school experiences, we included kindergarten repeaters since studies have shown that they score lower in reading and math (Benson and Borman 2010), and school sector to lessen the likelihood that year-round students would be paired with those in private schools given none of the year-round schools were private. Finally and as shown in Table 2, significant residential advantages existed for 9-month students relative to year-round students for all racial groups except African Americans. Asian-Pacific Islander and Hispanic children in 9-month schools for example have a median income that is on average 5100 and 4700 thousand dollars higher, respectively, than their year-round counterparts. We decided our propensity score should take into consideration residential qualities in order to address these treatment group discrepancies, more specifically the percentage of jobless males age 16 or over within the civilian labor force and the percentage of minorities within children's residential zip code. Pre-treatment characteristics' definition and measurement will be discussed in greater detail in the sections that follow.

YRE was defined in this study as children's enrollment in a school identified as such by their school administrators or as having experienced summer school. Rather than viewing start and end-dates to make assumptions about the type of school, we relied on the variable F4YRRND to identify schools with year-round calendars.⁴ Year-round school participation was determined for 1030 children, with nearly equal numbers attending summer school (520) or year-round schools (510). To pair these children with their counterparts in 9-month schools we express the general matching algorithm offered by Morgan and Winship (2007, p. 106) as:

⁴ Using school start and end-dates, von Hippel (2007) identified fewer year-round schools, but notes that including all year-round schools would not have changed his analysis results (p. 14).

$$\hat{\delta}_{TT,match} = \frac{1}{n^1} \sum_i \left[(y_i | d_i = 1) - \sum_j w_{i,j} (y_j | d_j = 0) \right],$$

where n^1 is the number of treatment cases, i is the index over experimental cases, j is the index over control cases, and $w_{i,j}$ represents the propensity scores that measure the distance between each 9-month school student (control case) and the target year-round student (experimental case).

There are numerous strategies that can be used in matching processes, and we have relied on several of them to increase the possibility of making successful matches and achieving balanced treatments. First, observational studies of YRE may control for race and other social background dimensions, but it often remains unclear whether balance among the two conditions applies to the racial groups within them. For example, von Hippel (2007) shows that there is a degree of positive self-selection among certain racial groups leading Hispanic and Asian-Pacific Islander children to be overrepresented in year-round schools while white and African American children are relatively underrepresented. In addition, Table 2 reveals social background differences between year-round and 9-month school children that appear to vary across racial groups. The greatest number of these significant treatment group differences emerge among Hispanic and white children. Therefore, we stratified our matching process according to race, that is, children attending year-round schools were matched, with the aid of propensity scores, to their co-ethnic/racial counterparts in 9-month schools. This maximized racial balance within the two school conditions while enhancing the likelihood that the schooling conditions are experienced by comparable racial/ethnic groups.

--TABLE 2 NEAR HERE--

A second adjustment to our algorithm was needed because our within-race matching approach would enhance the matching process only to the extent that within-race similarities are greater than they are between-race. Some inefficiency in the matching process could arise if it limited possible matches to those of the same racial group even when the true ideal match was of another race. To reduce the possibility that our matching strategy would yield less than ideal matches, our approach included a caliper match modification that is more stringent than has been used (Harding 2003), and bounds our matches to a maximum propensity score difference of .01 percent. For each year-round student w_i , the algorithm identified 9-month school students for whom $w_j = [w_i - .01, w_i + .01]$. If no 9-month match with a 1 percent or lower propensity score difference was found, the year-round case was eliminated from the analysis sample.

Third, we also carried out a one-to-one or “common-support” matching approach (Morgan 2001) using a lead-lag execution of our full-match algorithm (Hansen 2004). In this procedure, the algorithm identified a year-round case within the data file, considered the 9-month cases that preceded it (lead), and then considered those cases that followed it (lag) to identify the 9-month student with the proximate propensity score. The pair was subsequently flagged as matched and so not eligible to be matched to other records. Once the optimal match was identified for each year-round student, the $w_{i,j}$ is set equal to 1 for the matched control case and 0 for all other control cases. All of the unmatched 9-month students with 0 values and year-round students with no match were eliminated from the data set, leaving us with a final sample of just over 910 students in both year-round and 9-month schools, and nearly 1830 children in total. Inferences of the analysis are therefore limited to year-round students in 230 schools who have comparable counterparts in 251 traditional 9-month schools.

Fourth and finally, we also completed multiple-match procedures where any one year-round student was matched with multiple 9-month students of a propensity score within an acceptable caliper range. Morgan (2001) describes this process as one in which the 9-month school students that are matched to each year-round student are “stratified on the propensity score.” While this approach yields a

larger control group and overall sample size, it resulted in a larger percentage of experimental group members being unmatched and eliminated from the analysis. Rather than giving this sample equal attention in our analysis, we reserved its findings, and those of another common-support matched sample for a concluding discussion of the analytic matching process' sensitivity to unobserved characteristics.

ANALYTICAL CONSTRUCTS

We have estimated change in *reading* and *math* Item Response Theory (IRT) scale-scores, not only under the two conditions but also in time varying periods.⁵ We therefore report reading and math scores for the calendar year, the 9-month academic year and summer period. Regarding the latter two periods, assessments in these subjects occurred at times that did not coincide with the beginning and ending dates of the school year, resulting in the summer period having some days of schooling (that occurred after the last assessment of kindergarten and before the first assessment of first grade), and the exclusion of relevant days of instruction from 9-month test-score estimates. This feature of the data was corrected by first estimating the time that elapsed between the test dates and the start and end-dates of the 9-month school calendar, then calculating the amount of test-score growth that would have occurred during this time, and adding (to waves 2 and 4) or subtracting (from wave 3) the appropriate amount of test-score change. One dilemma in reapportioning test-score growth in this way was that year-round schools have start and end-dates that are much different than 9-month schools. For example, the average length of year-round children's summer vacation as reported by their parents is 54.13 days, nearly 24 days fewer than the 77.88 days reported for children in 9-month schools. In addition, 9-month schools tended to start several weeks later than year-round schools. To establish a similar summer timeframe in which to gauge test-score change between the two schooling types, we used the same elapsed-time approach described earlier to derive the test-scores that year-round children would have had at the start and end of 9-month schooling.

While the reliability of the tests exceeded 90% for both subjects (Rock and Pollack 2002), the reading IRT scores were rescaled during the 1st grade because the number of test items increased from 72 to 92. Children's kindergarten test-scores were then rescaled to reflect what their likely answers would have been had they taken the 92-item test. There is speculation that using 1st grade performances to rescale kindergarten scores might have resulted in their over or underestimation if the process failed to accurately account for what was learned or forgotten during the intervening summer (von Hippel 2007). This bias would apply to our analysis since we use the final kindergarten test-score. Despite this suspicion about reading scale-scores, our analysis yielded similar outcomes in both academic subjects.

In counterfactual modeling you would not typically include as model covariates the dimensions that were used to match children. However, our stratified matching strategy allowed us to use these characteristics to select matches within racial groups while assuming that these dimensions would also vary across racial groups when estimated within a conditional multivariate model. These characteristics related to children's social background and their amount of exposure to school environments. Social background variables were coded as 1 = yes, 0 = no for race/ethnicity (*Hispanic*, *white*, *Asian-Pacific Islander*, and *African American*) and *family social class*. In order to examine test-score differences between social classes, we used a composite measure of family social class that was segmented into equal-sized quintiles (e.g. *Low SES*, 1 = yes, 0 = no). This composite measure of family social class, provided by NCES, reflects the occupational status, educational level and total household income of parents. We also considered children's *gender* (1 = female, 0 = male), *single parent* family structure (1 = yes, 0 = no), and their *age* in months at the start of kindergarten.

⁵ IRT scale scores are designed to reduce ceiling and floor effects in studies of cognitive change. Reading assessments include concepts related to letter-case recognition; reading words in context; recognizing common words; and knowing letter sounds at the beginning and end of words. Assessments in math included count, number and shape concepts; numerical ordinality and sequences; addition and subtraction and simple multiplication and division (Rock and Pollack 2002).

Another group of variables accounted for differences in the amount of children's exposure to school and the kind of summer education they received. These measures included whether children attended a *full-day kindergarten* program (versus half-day), *attended pre-school*, and *repeated kindergarten*, all coded 1 for yes and 0 for no. Regarding possible differences between year-round and 9-month schools, it must be noted that the experimental treatment includes two forms of YRE: summer school and year-round school. We distinguished between these two arrangements by estimating differences in the *type of summer* instruction (1 = year-round, 0 = summer school) experimental students experienced.

In order to take advantage of the unique way both treatments inform questions of residential stratification, we used a NCES companion data file that linked ECLS-K children to characteristics of their zip-codes (Beveridge et al. 2004). While the imperfections of census data as proxies for residential areas have been noted (Jencks and Mayer 1990), they present an objective appraisal of areas that complemented the more subjective parent reports of neighborhood conditions that we used in this analysis. These variables included zip-codes' *median family income* and the *percentage minority*. The median income variable was created by first using a natural log transformation to achieve a more suitable distribution of incomes, then converting those values into z-scores. For the sake of interpretation, Table 3 reports the original values of this variable. We combined measures of the proportion of African American and Hispanic individuals to create the zip-code's *percentage minority* measure because those racial groups have the highest metropolitan segregation levels, and the largest proportion of their populations located in hyper-segregated areas (Logan, Stults and Farley 2004).

The ECLS-K provided a location type variable to identify children that resided in central cities and also asked parents their perceptions of their neighborhood, which we used to create a composite indicator of neighborhood social disorganization. The *city* variable (1 = yes, 0 = no) permitted us to account for the fact that year-round schools are present in and outside of central cities and to remain open to the ways in which city schools might differ from others. Regarding social disorganization, parents were asked: "*how much of a problem is burglary*", "*violent crime*" and "*selling/using drugs in the area*" (1 = big problem, 2 = somewhat a problem, 3 = no problem). Our *social disorganization* composite was coded 1 for yes and 0 for no if parents indicated that any of these factors were a big problem. Although it has been stated that areas with high minority compositions face more social problems (Jargowsky 1997), our diagnostics produced no concerns of multicollinearity among these neighborhood dimensions.

ESTIMATION

Using HLM version 6.08 (Raudenbush and Bryk 2002), we specified a 3-level model consisting of child test-scores at Level 1, between-child measures reflecting his or her social background and educational exposure at Level 2, and residential dimensions at Level 3. Given test-score change can happen in different periods of the year, test-score parameters are estimated separately for the 9-month school session, the summer and also for the entire calendar year, yielding the Level 1 equation:

$$Y_{tcn} = \pi_{0cn} + \pi_{1cn}(\text{Spring kindergarten assessment}_{tcn}) + \pi_{2cn}(\text{Fall 1}^{\text{st}} \text{ grade assessment}_{tcn}) + \pi_{3cn}(\text{Spring 1}^{\text{st}} \text{ grade assessment}_{tcn}) + e$$

Where test-scores Y_{tcn} is a function of an intercept representing reading and math for child c in neighborhood n , and her or his exposure to periods that span spring kindergarten test-score to fall 1st grade score (summer), the beginning and end-score of 1st grade (9-month), and the end of kindergarten to the end of 1st grade (year-round) at the time of test t . Test-scores during these time-spans are estimated for children that are enrolled in year-round and 9-month schools together in an unconditional analysis.

Within each time-span, our counterfactual approach assumed that every individual has a potential outcome in both treatment conditions, even if each child can be observed in only one school treatment at any one time (Morgan and Winship 2007). We express this assumption as:

$$\delta_i = Y_i^e - Y_i^c$$

where, Y_i = reading or math outcomes and e and c indicate whether test-scores are of the experimental or control condition. The matching procedures that we described earlier addressed the fact that we can only observe child i in one treatment and not both, allowing us to continue with the specification of the causal effect on child i as an expected value of difference between Y^e and Y^c :

$$\bar{\delta} = \bar{Y}^e - \bar{Y}^c$$

The average treatment effect is then the difference between these two estimated means as indicated by an *all-year* variable (1 = yes, 0 = no).

Other questions in our analysis concern variation in treatment effects across racial/ethnic groups, social background and school dimensions. Level 2 of the multilevel model specified social background, educational exposure and residential characteristics in all three time-spans for both schooling types separately. Each Level 2 parameter represents the adjustment in the area's average performance slope, β_{10n} . Test-score growth π_{1cn} is a function of children's age, gender, and single parent family structure; whether they repeated kindergarten; attended full-day kindergarten and a preschool program; the type of summer instruction; and, their family social class quintile (with the middle quintile excluded), race and city residency. The only way in which Level 2 differed across the three periods is that the variable, summer type, was withheld from the estimation of 9-month performances since there is no school exposure for control group children in this period. The full Level 2 equation is as follows:

$$\begin{aligned} \pi_{1cn} = & \beta_{10n} + \beta_{1n}(Age_{cn}) + \beta_{12n}(Gender_{cn}) + \beta_{13n}(Single\ parent_{cn}) + \beta_{14n}(Repeated\ kindergarten_{cn}) \\ & \beta_{15n}(Full\ day\ kindergarten_{cn}) + \beta_{16n}(Preschool\ program_{cn}) + \beta_{17n}(Summer\ type_{cn}) + \beta_{1,8-11n}(SES\ quintiles_{cn}) \\ & + \beta_{1,12-15n}(Race_{cn}) + \beta_{116n}(City_{cn}) + a_{cn} \end{aligned}$$

Recall that children in our analysis are nested within residential areas, 263 and 363 zip-codes for year-round and 9-month school children, respectively. We therefore model “neighborhood-to-neighborhood” variation in residential characteristics with random intercept models in all three periods and both school conditions. Hence, test-score change in each time-span, β_{10n} is a function of zip-codes' median family income and percentage African American and Hispanic, both segmented into equal thirds and a social disorganization composite. We express this Level 3 equation as:

$$\begin{aligned} \beta_{10n} = & \gamma_{100} + \gamma_{101,2n}(Median\ family\ income_n) + \gamma_{103,4n}(\% \text{ Minority}_n) + \\ & \gamma_{105n}(Social\ disorganization_n) + r_{10n} \end{aligned}$$

Where, the intercept γ_{100} , represents the average test performance of a specific residential area for all areas in the sample, γ_{101n} through γ_{104n} indicates the estimated deviation from the area mean test performance associated with a point increase among those characteristics, and γ_{105n} represents the average point change in children's mean test performance associated with a residential area's identification as having those problems.

--TABLE 3 NEAR HERE--

ANALYSIS

DESCRIPTIVE STATISTICS

Our sampling strategies achieved a sample more diverse than in previous research, giving us the freedom to provide a detailed analysis of racial/ethnic differences. Although von Hippel (2007) was one of the few studies that included separate racial/ethnic categories, the proportion of African Americans in his sample barely reached 5 percent whereas ours exceeded 12 percent. At the high end, our sample had more white (44%) than Hispanic (31%) children in contrast to other studies, such as von Hippel's, that had majority Hispanic samples.

While confirming that the treatment samples are balanced is difficult to do for the reasons outlined by Morgan and Winship (2007, p. 114), we argue that balance would be more apparent if the pre-treatment characteristic that seemed most highly correlated with the outcome, but was not a dimension on which children were stratified in our matching procedures, appeared sufficiently equal across groups after matching. Test-scores that were measured prior to treatment at the start of children's educational careers, for example, should appear similar across treatment groups under an assumption of balance. This is in fact the case; the mean-difference at kindergarten's start between the two treatment conditions was just .70 points (31.41 versus 32.11) in reading and .80 points in math (21.43 versus 22.23), both insignificantly higher for children in 9-month schools. We therefore have concluded that the variables on which we estimated propensity scores and our match optimization strategies yielded treatment groups without confounding pre-treatment test-score differences.

UNCONDITIONAL ANALYSIS

Our first analysis question asked whether significant test-score differences existed between year-round and 9-month school children and if so, did they vary according to the season? Our analysis shown in Table 4 suggests the answer to both of these questions is "yes." In Table 4 are estimates for children's reading and math test-scores in all three time-spans for the full sample analysis. The first column provides the mean test-score while the second column labeled "difference" includes the all-year variable, which represents the estimated difference of year-round test-score performances from the mean. For the calendar year, the first row of Table 4 shows that experimental group children accumulated 4.097 ($p = .001$) fewer test-score points in reading, leading to a gap between them and control group children of just over a quarter of a standard deviation unit. YRE students also accumulated about 2.143 ($p = .002$) points less than control group students in math, equaling a gap of about .178 standard deviation units.

In the second row, the same models are specified for summer test-score growth. The reading analysis reveals no significant test-score growth or loss for students in general (-0.256 , $p = .52$) or year-round students in particular (1.007 , $p = .160$). However, the math analysis shows that all children experienced summer-time test-score gains (1.686 , $p = .001$), and that this growth did not appear to vary significantly according to school type ($.824$, $p = .137$).

--TABLE 4 NEAR HERE--

Estimates of 9-month academic year learning are detailed in the final row, and there we find the most notable analysis results. YRE children accumulated 4.747 ($p = .001$) points less than their 9-month counterparts in reading and 3.609 ($p = .001$) points fewer in math. While the finding that year-round children accumulated fewer points than their 9-month counterparts should have been expected given that they did so for the calendar year, we must explain why the relative loss is of a greater magnitude (29.48 and 29.05 standard deviation unit difference in reading and math, respectively). For children receiving YRE,

this lower 9-month accumulation of test-score points is expected due to there being nearly 13 fewer instructional days than what year-round children would have received within a calendar year and that children in 9-month schools receive. If this slower YRE 9-month growth is among summer school children, we reason that children might have been placed in summer school because they had performed less well during the 9-month school term. Moreover, these results cast doubt on the spacing effect hypothesis (i.e. less slippage due to shorter summer breaks or a quicker resumption of learning than children in traditional 9-month schools) because 9-month gains in year-round schools are .16 and .20 points per month lower in reading and math, respectively, than they are for the calendar year. While the seasonal dimension of our counterfactual framework has revealed that more instructional days led to test-score gains among year-round children, a .16 to .20 points per month gain suggest there are not enough months of the year in which to add instruction that would result in the elimination of the gap between them and 9-month children. For example, the summer would need to be 10.715 months long in order to totally offset the 9-month math shortfall of -3.61 points.

MIXED MULTIVARIATE LINEAR MODELS

We also asked was there variation in how race, social class and residency related to test-scores across seasons and school schedules. With this question we addressed longstanding speculation within research about schools as sites where academic differences grow (Downey, von Hippel and Broh 2004; Condrón 2009), and the school's possible mitigation of residential effects (Johnson 2012a; Rendon 2014). Pursuant to these interests, we specified conditional models for each treatment and time-span and reported the results in Tables 5 and 6. Addressing the racial concerns first, the reading results in Table 5 show that only the test-scores of Asian-Pacific Islander children differ from the average test-score growth of year-round school children. This stronger than average gain of 6.675 points ($p = .048$) occurred mostly during the 9-month period (5.251, $p = .021$). Moreover, stratification along the dimensions of race/ethnicity and social class appears surprisingly flat within year-round schools. So while year-round schools may not yield higher test-scores than 9-month schools, they appear to be institutions with more uniform benefits.

The story is quite different in 9-month schools where test-score gaps resemble those found in previous research (Downey, von Hippel and Broh 2004; Johnson 2014). Even as social class is considered in this model, African Americans still experienced a sizable setback (-8.231, $p = .001$) of approximately .458 standard deviation units. However, their negligible gains were offset by significant growth during the summer (2.064, $p = .029$) relative to white children, leading to a calendar year shortfall of 4.397 points ($p = .017$). It appears as though 9-month schools are not contexts where all status groups have kept pace with or exceeded mean test-score growth as they apparently have in year-round schools. Most important however, is that despite the higher mean achievement of children in 9-month schools (34.468 points), African American test-score growth in them totaled just 26.23 points, which is lower than the 9-month (30.119) and yearly (30.888) point-estimates of African Americans in year-round schools. That the 9-month African American estimate is lower than the 9-month YRE estimate is indeed surprising because the latter has fewer days of instruction in this period than the former. It is worth noting that 9-month schools were also especially effective in social class sorting since children in the lowest social class accumulated fewer points (-9.813 $p = .001$) with the majority of this slippage having occurred while they were in school (-8.596 $p = .001$). This large shortfall, equaling .546 standard deviation units, left their total accumulated points at 25.87, far lower than the mean test-score of children in the lowest social class in year-round schools.

--TABLE 5 NEAR HERE--

The estimates of the math analysis shown in Table 6 were more varied, but mirrored the reading results in important ways. In regards to race/ethnicity, Asian-Pacific Islander children were once again the only racial/ethnic group that exhibited significantly different math gains, but this time they had a lower rate

of growth over the calendar year (-3.990 , $p = .015$). In 9-month schools, Table 6 shows that racial/ethnic inequality is greater than it appeared in year-round schools. African Americans (-5.839 $p = .001$) and Hispanics (-3.086 $p = .026$) accumulated fewer points only while school was in session and experienced large calendar year setbacks equaling .463 standard deviation units for the former and .244 for the latter group. In fact, the 20.759 points that African Americans gained during the 9-month traditional academic year is less than African Americans gained with 9 months (21.356 points) and a full year (24.017 points) of YRE. Regarding social class differences, lower than average gains occurred for children in the lowest social class during the summer (-3.842 , $p = .001$), but they became insignificantly different than mean calendar-year growth once combined with their growth during the 9-month period. Year-round and 9-month children in the higher social classes experienced stronger than average gains over the calendar year. Consistent with the reading analysis, the math analysis has shown a greater degree of stratification in 9-month schools than year-round schools.

Turning our attention to the subject of residential effects, the reading analysis displayed in Table 5 reveals YRE children's growth in low income residential areas was about -6.165 points ($p = .002$) or .453 standard deviation units less than mean growth during the 9-month period. The fact that the slippage reduced to just .264 standard deviation units for the calendar year (-4.179 , $p = .032$) suggest that while summer instruction is not significant, it effectively offset 9-month losses. Equally large residential shortfalls occurred in reading in 9-month schools, but this time it was related to the neighborhood's social disorganization (-4.831 $p = .024$). Again, we see a decreased magnitude of the neighborhood effect once we considered 9-month school students' calendar year gains (-4.768 $p = .050$, 26.98 sd). Thus, in reading we have concluded that neighborhood effects on test-scores 1) were strongest during the 9-months of both school types where children had the maximum percentage (55.19 to 62.99) of school days, 2) modest in the calendar year where the percentage of instructional days ranged from 49.32 to 52.33, and 3) small to non-existent during the summer when children had the smallest percentage (0 to 27.44) of school days. In other words, neighborhood effects appeared strongest in periods with more school days, not fewer.

In math, the summer time appeared much more consequential to an understanding of neighborhood-based learning than it was in the reading analysis. For instance, children that resided in high income areas and attended year-round schools had stronger than average gains during the summer (2.767 $p = .022$). This positive association may have resulted from their receipt of summer instruction because the estimated effect of summer type showed a nearly significant advantage for YRE over summer school (1.621 $p = .077$). YRE includes approximately 7 more days of instruction during the summer than does summer school.

DISCUSSION

This study of race and residential effects in year-round and 9-month schools addressed a major void within research on an important question. The question of whether schools exacerbate or equalize disparities in children's test performance is an essential one to address in a society where education is the key to social mobility. There are voids in the literature regarding the usefulness of YRE as a viable policy option because we know little about what form it should take to secure the desired outcomes for children, and how it functions with regard to race/ethnicity and residency—two dimensions of social stratification that many give as much credit for the social reproduction of status hierarchies as is given to the family. About these inequalities, our analysis has led us to conclude that schools can be greater equalizers according to race than they are currently, but may simultaneously serve to enable neighborhoods' impressive stratifying effects.

Regarding the overall effects of YRE, this analysis has shown that children in year-round schools do less well and gain the least during the traditional 9 month period, casting doubt on the spacing effect hypothesis. Yet smaller relative losses as children's learning extended into the academic year suggested

that more time in them does improve test scores, but not significantly enough to make up the gap with traditional school children in just three months' time. Consequently, we conclude that test-score equality would not be achieved by simply lengthening the school year.

To the extent that year-round schools were a benefit during the summer, this analysis gave the edge to those with modified calendars over those that included summer school. At no time in this analysis of summer-time learning did summer school appear more beneficial than year-round school. Summer school effects have appeared positive with some consistency in research, but finding stronger effects for year-round schools is reasonable since year-round schools have more summer instructional days than do summer schools. We make this distinction cautiously because it would be erroneous to presume that the number of days is the only way in which these schools differ. Future research will need to illuminate to what extent their unique social organization contributes to differing impacts while accounting for variation in school dosage. Nonetheless, we assume that social organization differences are not spurious, and are in fact related to, if not caused by, the two treatments.

But this analysis did not estimate these general effects and then assume that they would apply equally to all racial/ethnic groups. Instead, we explored whether there was variation in how race/ethnicity related to test-scores across these forms of schooling and seasons. Regarding within-race/ethnicity differences between the two school types, African Americans were the only race/ethnic group that gained more in year-round schools than in 9-month schools in both subjects. Not only did African Americans in year-round school gain on co-ethnics in 9-month schools, their test-scores differed insignificantly from those of whites and Hispanics, while in comparison to Asian-Pacific Islanders they did slightly better in math and worse in reading. In contrast, the test-score growth of African Americans in 9-month schools lagged behind that of all racial groups in 9-month and year-round schools. This analysis therefore found that their greatest potential to serve as equalizers was for African Americans, the group that previous research suggested had disproportionately shouldered its stratifying effects. While this analysis had not found that African Americans in year-round schools had gains strong enough to equal the mean performance levels of white children in traditional schools, this was nearly accomplished in math while the reading gap was cut by over half (from 8.23 to 3.58 points). Apparently, year-round schools' compensatory capacity has been, until now, hidden away in existing research within the mean effects of unbalanced natural experiments and aggregated estimates of racial/ethnic groups.

The outcomes of this analysis identified for African Americans an educational policy option with some degree of transformative potential. However, we should be concerned about the gains of white children in YRE, and acknowledge that the lesser degree of racial stratification apparent in year-round schools comes at a cost to white children relative to those in traditional schools. So our task is to secure the gains of year-round schools for African Americans while maintaining the enrollment of other racial/ethnic groups in the traditional schools that seem to support their achievement best. The fact that schools in the U.S. are so racially segregated presents the ironic benefit of easily targeting African Americans for YRE reforms. But there are some obstacles in the way. As it stands, African Americans are relatively underserved by YRE in contrast to Asian American and Hispanic children, because they are underrepresented in the region where year-rounds schools are most popular (the West) and live in some states where YRE is generally unpopular. Of these states, Michigan, Mississippi, Virginia and Florida have instituted laws that restrict the start of the academic year to a period close to Labor Day while Alabama's law goes as far as to limit the number of school hours (i.e 1080) to that typical of a 180-day school-year. Arguments that reforms are good for African Americans rarely result in their adoption in a society that many people mistakenly claim to be post-racial and that prefers the appearance of race-neutral public policies. Federal incentives provided directly to schools would need to be significant enough to ease states' resistance to YRE.

This study also produced insight about how residential effects were related to test-outcomes given children's dose of school exposure. The pattern of relationships found in the reading analysis suggested

that schools did not offset neighborhood effects as much as they functioned to relay them. The largest residential effects of low SES and social disorganization occurred during the nine month periods of both school forms, the schools that offered the greatest number of instructional days. Within the summer when the percentage of instruction was lowest, neighborhood disadvantages were insignificantly related to test-scores. There are a few conclusions that we have drawn about these residential effects. First, the covariance of residential effects and school dosage is not entirely counterintuitive since the former arises in large part through human interaction and schools are possibly the primary medium that facilitates interaction for children. Second, the aforementioned results suggest that the artificial isolation of neighborhoods from schools as posited in the autonomous institutional framework is invalid. Third, there is little evidence presented in this study that supports the presence of temporal fluctuations in residential effects in the way we imagined they might occur; they were not stronger during the summer as they appeared in the results of Benson and Borman (2010) and Johnson (2014). This study implies the relevance of a faucet theory of neighborhood effects can instead be applied to the academic season where stronger effects were present. After all, Johnson (2012b) held out the possibility that “institutional effects may mediate neighborhood influences as much as they might inspire them” (p. 37). We therefore recommend that the field moves away from an ecological correspondence theory which presumes a simple resemblance of neighborhoods and schools in their function. Instead, we conclude from this study that without schools, neighborhoods could not function to impact children’s test-scores. In other words, the school is “an essential organ” of the neighborhood, without which residential disadvantages cease functioning in relation to young children’s test-performances.

But there are some cautions related to research of this kind that we must mention. One caution is that counterfactual models address the bias of only observed characteristics that would be found in inferential studies and not unobserved ones. While several social background dimensions were used to construct propensity scores, it remains possible that relevant pre-treatment characteristics have been excluded. We sought to test this proposition, and hence the vulnerability of this study’s analysis, by applying our pre-treatment match criteria to a smaller matched sample of 1094 (from the ECLS-K subsample) with a .01 caliper threshold, and a larger one-to-many matched sample of 3818 children using a more stringent .005 caliper threshold to simulate the influence of less and more efficient matching estimators.⁶ With both matched files, we were able to generate racial/ethnic and residential estimates that mirrored the patterns we shared in our main analysis. We are therefore optimistic that our findings are likely to withstand decreased propensity score differences among matches that could arise from the use of other matching characteristics. Another limitation is that our causal inferences pertain to year-round children that have 9-month matches. We highlighted the sample of the three that contained the greatest number of experimental children to enhance generalizability, but we make no claims that it is representative of ECLS-K children or children nationally in year-round schools. Of course, we encourage future studies to determine whether the findings of analyses like this one apply to children of other age groups.

REFERENCES

- Arum, Richard. 2000. “Schools and Communities: Ecological and Institutional Dimensions.” *Annual Review of Sociology* 26: 395-418.
- Benson, James and Geoffrey Borman. 2010. “Family, Neighborhood, and School Settings across Seasons: When do Socioeconomic Context and Racial Composition Matter for the Reading Achievement Growth of Young Children?” *Teachers College Press* 112(5): 1338-1390.

⁶ The one-to-many matching strategy required the use of a weight to achieve balance in the probabilities of selection across the two conditions.

- Beveridge, Andrew, Sophia Catsambis, Susan Weber, R. Sam Michalowski, Charis Ng, and Jerry West. 2004. Census Data and Geocoded location for the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) User's guide, (NCES 2004-116). Washington, DC: National Center for Education Statistics.
- Bourdieu, Pierre. 1977. "Cultural Reproduction and Social Reproduction." Pp. 487-511 in *Power and ideology in education*, edited by Jerome Karabel and A. Halsey. New York: Oxford University Press.
- Burkham, David, Douglas Ready, Valerie Lee, and Laura Logerfo. 2004. "Social Class Differences in Summer Learning between Kindergarten and First Grade: Model Specification and Estimation." *Sociology of Education* 77:1-31. doi:10.1177/003804070407700101
- Condrón, Dennis. 2009. "Social Class, School and Non-school Environments, and Black/White Inequalities in Children's Learning." *American Sociological Review* 74:683-708. doi:10.1177/000312240907400501
- Cooper, Harris, Kelly Charlton, Jeff C. Valentine, and Laura Muhlenbruck. 2000. "Making the Most of Summer School: A Meta-Analytic and Narrative Review." *Monographs of the Society for Research in Child Development* 65(1), entire issue.
- Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay and Scott Greathouse. 1996. "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review." *Review of Educational Research* 66:227-268. doi:10.3102/00346543066003227
- Cooper, Harris, Jeffrey C. Valentine, Kelly Charlton, and April Melson. 2003. "The Effects of Modified School Calendars on Student Achievement and on School and Community Attitudes." *Review of Educational Research* 73:1-52. doi:10.3102/00346543073001001
- Downey, Douglas, Paul von Hippel, and Beckett Broh. 2004. "Are Schools the Great Equalizer? Cognitive Inequality during The Summer Months and the School Year." *American Sociological Review* 69: 613-635. doi:10.1177/000312240406900501
- Entwisle, Doris R. and Karl L. Alexander. 1992. "Summer Setback: Race, Poverty, School Composition and Math Achievement in the First Two Years of School." *American Sociological Review* 57: 72-84.
- Gandara, Patricia and Judy Fish. 1994. "Year-Round Schooling as an Avenue to Major Structural Reform." *Educational Evaluation and Policy Analysis* 16: 67-85.
- Harding, David. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping out and Teenage Pregnancy." *American Journal of Sociology* 109:676-719. doi:0002-9602/2003/10903-0004
- Hansen, Ben. 2004. "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association* 99:609-618.
- Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd. 1998. "Characterizing Selection

- Bias Using Experimental Data." *Econometrica* 66(5):1017-1098.
- Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling*. New York: Academic Press.
- Jargowsky, Paul. 1997. *Poverty and Place: Ghettos, Barrios and the American City*. NY: Russell Sage Foundation.
- Jencks, Christopher and Susan Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." Pp. 111-86 in *Inner-city Poverty in the United States*, edited by Laurence E. Lynn, Jr., and Michael G. McGeary. Washington, DC: National Academy Press.
- Johnson, Jr., Odis. 2014. "Race–Gender Inequality Across Residential and School Contexts: What Can Policy Do?" Pp. 345-376 in *African American Males in PreK-12 schools: Informing Research, Practice, and Policy*, edited by James L. Moore and Chance W. Lewis. Emerald Publishing.
- Johnson, Jr., Odis. 2012a. "A Systematic Review of Neighborhood and Institutional Relationships Related to Education." *Education and Urban Society* 44(4):477-511.
- Johnson, Jr., Odis. 2012b. "Toward a Theory of Place: Social Mobility, Proximity and Proximal Capital." Pp. 29 – 46 in *Research on Schools, Neighborhoods and Communities: Toward Civic Responsibility, Presidential Volume* edited by William Tate. MD: Rowman & Littlefield and the American Educational Research Association.
- Lauritsen, Janet L. and Nicole White. 2014. Seasonal Patterns in Criminal Victimization Trends, Report NCJ245959. U.S. Department of Justice, Office of Justice Programs. Washington: Bureau of Justice Statistics.
- Lee, Valerie, David Burkam, Doug Ready and Laura LoGerfo. 2004. "Inequality Beyond the Starting Gate: Race and Class Differences in Children's Learning in Kindergarten and First Grade." Unpublished conference paper. Chicago: University of Chicago Harris School of Public Policy Studies.
- Logan, John R., Brian J. Stults and Reynolds Farley. 2004. "Segregation of Minorities in the Metropolis: Two Decades of Change." *Demography* 41:1–22.
- McMillen, Bradley J. 2001. "A Statewide Evaluation of Academic Achievement in Year-Round Schools." *Journal of Educational Research* 95(2):67-74.
- Morgan, Stephen L. 2001. "Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning." *Sociology of Education* 74:341 – 374.
- Morgan, Stephen and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- U.S. Department of Education. 2013. Digest of Education Statistics. Washington: National Center for Education Statistics. Accessed at http://nces.ed.gov/programs/digest/d13/tables/dt13_234.12.asp.
- Parson, Talcott. 1959. "The School Class as a Social System: Some of its Functions in American Society." *Harvard Educational Review* 29:297-318.

- Reardon, Sean. 2003. "Sources of Educational Inequality: The Growth of Racial/Ethnic and Socioeconomic Test Score Gaps in Kindergarten and First Grade." Unpublished manuscript. University Park: Pennsylvania State University.
- Reece, Jennifer, Carl Myers, Christy Nofsinger and Reagan Brown. 2000. Retention of Skills over the Summer Months in Alternative and Traditional Calendar Schools. *Journal of Research and Development in Education* 33(3):166-174.
- Rendon, Maria G. 2014. "Drop Out and 'Disconnected' Young Adults: Examining the Impact of Neighborhood and School Contexts." *Urban Review* 46:169-196.
- Rock, Donald A. and Judith M. Pollack. 2002. "Early Childhood Longitudinal Study—Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten through First Grade." Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Sharkey, Patrick and Robert J. Sampson. 2010. "The Acute Effect of Local Homicides on Children's Cognitive Performance." *Proceedings of the National Academy of Sciences of the United States of America* 107(26):11733-11738.
- Sorokin, Pitirim. 1959. *Social and Cultural Mobility*. New York, NY: The Free Press.
- von Hippel, Paul. 2007. What happens to Summer Learning in Year-round Schools? Unpublished Manuscript. Columbus: Ohio State University.
- Wodtke, Geoffrey T., David J. Harding and Felix Elwert. 2011. "Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation." *American Sociological Review* 76:713 – 736.

TABLE 1. Counterfactual Design

	IN SCHOOL YEAR-ROUND (Experimental)	IN SCHOOL 9-MONTH (Control)
YEARLY GROWTH Number of School days (% School days)	Yearly growth with year-round schooling 190.99 (52.33%)	Yearly growth with 9- month schooling 180 days (49.32%)
SUMMER GROWTH Number of School days (% School days)	Summer growth with year- round schooling 22.28 days (27.44%)	Summer growth with 9- month schooling 0 days (0%)
9-MONTH GROWTH Number of School days (% School days)	9-month growth in year-round schools 157.72 days (55.19%)	9-month growth in 9- month schools 180.00 days (62.99%)

FIGURE 1. Nested Seasonal-Counterfactual Design

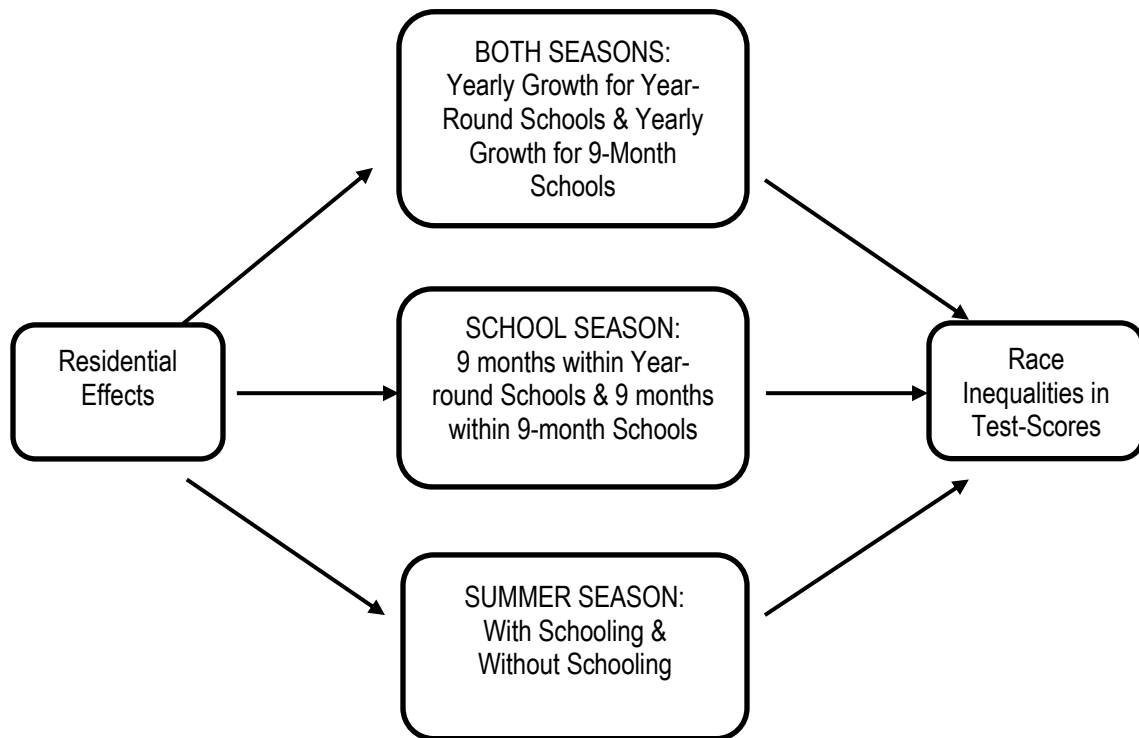


TABLE 2. Unmatched Mean Differences of Racial Groups in 9-Month Versus Year-Round Schools

Variables	Asian-PI		Black		Hispanic		White	
	Difference	SE	Difference	SE	Difference	SE	Difference	SE
Social class quintile (1 = low, 5 = high)	.13	.15	-.13	.14	.59***	.09	.08	.07
Gender (1 = female, 0 = male)	-.07	.05	-.01	.05	-.07+	.04	-.05+	.03
Single parent (1 = yes, 0 = no)	-.04	.04	.02	.05	.05+	.03	-.03	.02
Age in months (Months at kindergarten start)	.53	.43	.55	.45	1.28***	.30	.29	.25
Repeated kindergarten (1 = yes, 0 = no)	-.03*	.01	-.00	.02	.04*	.02	-.02*	.01
Full-day kindergarten (1 = yes, 0 = no)	.12*	.05	.12**	.04	.34***	.03	.06*	.03
Preschool (1=head start/center/day care, 0=no)	-.00	.03	.06	.04	-.02	.02	-.03**	.01
School sector (1 = public, 0 = private)	-.02	.04	-.02	.04	.12***	.02	.08***	.02
City location (1=yes, 0= no)	-.16***	.05	-.03	.05	.01	.03	.00	.02
Neighborhood disorganization (1=yes, 0=no)	.17***	.04	.01	.05	-.09*	.03	.01	.01
Zip code average percent male joblessness	-4.06***	1.11	-.50	1.33	-2.05*	.81	1.51**	.51
Zip code average median family income	5106.13*	2093.65	567.09	1418.65	4707.86***	1102.95	-1524.69	1095.40
Zip code mean percent minority	-8.13***	2.13	2.03	3.10	-14.61***	2.15	-4.42***	.83
Reading score kindergarten end to 1 st grade start	-1.53*	.74	-.48	.68	-1.30*	.57	-1.43***	.35
Reading score kindergarten end to 1 st grade end	.79	1.05	-1.23	1.04	2.15**	.81	2.07***	.56
Reading score grade 1 start to grade 1 end	2.20*	1.01	-.69	1.15	3.40***	.98	3.52***	.60
Math score kindergarten end to 1 st grade start	.48	.63	-.95	.63	-.75+	.44	-.26	.33
Math score kindergarten end to 1 st grade end	-.05	.73	-1.28+	.70	.95+	.55	.69+	.37
Math score grade 1 start to grade 1 end	-.51	.65	-.25	.79	1.66**	.53	1.10**	.41

TABLE 3. Descriptive Statistics, Matched Full (N = ~1830)

Variables	Mean	STDV
Asian/Pacific Islander (1 = yes, 0 = no)	.13	.34
Black (1 = yes, 0 = no)	.12	.32
Hispanic (1 = yes, 0 = no)	.31	.46
White (1 = yes, 0 = no)	.44	.50
Social class quintile (1 = low, 5 = high)	2.84	1.45
Low social class (1 = yes, 0 = no)	.26	.44
Middle low social class (1 = yes, 0 = no)	.19	.39
Middle social class (1 = yes, 0 = no)	.18	.38
Middle high social class (1 = yes, 0 = no)	.19	.39
High social class (1 = yes, 0 = no)	.18	.38
Gender (1 = female, 0 = male)	.54	.50
Single parent (1 = yes, 0 = no)	.19	.40
Age in months (months at kindergarten start)	65.11	4.38
Sector (1 = public, 0 = private)	.19	.39
Repeated kindergarten (1 = yes, 0 = no)	.04	.20
Full-day kindergarten (1 = yes, 0 = no)	.49	.50
Preschool (1=head start/center/day care, 0=no)	.09	.29
Year-round school (1=yes, 0=no)	.23	.42
All-year (1= summer school/year-round, 0= 9-month)	.50	.50
City location (1=yes, 0= no)	.45	.50
Zip code average median family income	51052.06	23217.06
Zip code median family income – Lower third	.39	.49
Zip code median family income – Middle third	.32	.47
Zip code median family income – Upper third	.29	.46
Zip code mean percent minority	34.30	29.75
Zip code percent minority – Lower third	.22	.41
Zip code percent minority – Middle third	.34	.48
Zip code percent minority – Upper third	.44	.50
Zip code percent male jobless in civilian labor force	36.17	11.89
Neighborhood disorganization (1=yes, 0=no)	.15	.35
Months from kindergarten end to grade 1 start	2.62	.28
Months from test 2 to kindergarten end	1.08	.49
Months from grade 1 start to test 3	1.43	.52
Months from grade 1 start to test 4	8.30	.57
Months from grade 1 start to grade 1 end	9.45	.36
Months from test 4 to grade 1 end	1.18	.51
Reading score kindergarten start	31.77	11.92
Reading score kindergarten end to 1 st grade start	-.11	9.28
Reading score kindergarten end to 1 st grade end	32.64	16.88
Reading score grade 1 start to grade 1 end	32.58	16.10
Math score kindergarten start	21.84	9.52
Math score kindergarten end to 1 st grade start	1.64	8.64
Math score kindergarten end to 1 st grade end	25.59	12.03
Math score grade 1 start to grade 1 end	23.93	12.42

Table 4. Unconditional Models of Reading and Math Test-Score Mean Growth and Year-Round/9-Month Differences, Full Sample (N= ~1830)

	Reading				Math			
	Mean		Difference		Mean		Difference	
	Growth	SE	Growth	SE	Growth	SE	Growth	SE
Yearly Growth								
Intercept	32.479***	0.594	32.299***	0.591	25.762***	0.373	25.730***	.374
All-Year	--	--	-4.097***	1.166	--	--	-2.143**	.693
Level 1 & 2 τ /STDV	216.10***	14.70	213.09***	14.59	106.17***	10.30	104.96***	10.25
Level 3 τ /STDV	37.864***	6.153	37.891***	6.155	11.242***	3.353	11.731***	3.425
Summer Growth								
Intercept	-0.256	0.399	-0.217	0.396	1.686***	0.320	1.701***	0.314
All-Year	--	--	1.007	0.599	--	--	.824	0.554
Level 1 & 2 τ /STDV	59.656***	7.724	59.717***	7.727	51.972***	7.209	52.163***	7.222
Level 3 τ /STDV	10.101***	3.178	9.750***	3.122	8.018***	2.831	7.516***	2.742
9-month Growth								
Intercept	32.466***	0.569	32.343***	0.551	23.846***	0.472	23.764***	.451
All-Year	--	--	-4.747***	1.060	--	--	-3.609***	.799
Level 1 & 2 τ /STDV	185.28***	13.61	181.85***	13.49	101.50***	10.07	100.51***	10.03
Level 3 τ /STDV	35.045***	5.919	33.991***	5.830	25.916***	5.091	23.459***	4.844

*** = $p < .000$, ** = $p < .01$, * = $p < .05$, + = $p < .10$

TABLE 5. Multivariate Models of Reading Growth, Year-Round & 9-Month

<i>Reading</i>	Year-Round Enrollment			9-Month Enrollment		
	Yearly	Summer	9-month	Yearly	Summer	9-month
Intercept	30.888***	0.134	30.119***	34.153***	- 0.395	34.468***
Age	0.324	0.105	0.157	-0.096	0.117	-0.207
Gender	-2.803	-0.999	-1.879+	-1.801	-1.212*	-0.370
Single parent	0.647	-0.395	2.432*	-1.697	-0.356	-1.184+
Repeated kindergarten	-8.443*	-3.412	-5.136+	-9.529*	-2.563	6.900+
Full day kindergarten	-1.171	-0.549	0.198	-0.988	0.248	-0.587
Pre-school program	-0.474	2.122	-0.739	-0.245	0.798	-0.622
Summer type	1.209	0.909	--	--	--	--
Low social class	-3.429	-2.704	1.367	-9.813***	-1.059	-8.596***
Mid low social class	2.713	0.702	1.142	-2.401	-0.067	-1.931
Mid high social class	0.585	0.295	-0.200	4.271+	1.040	3.679
High social class	2.450	1.777	2.619	3.535	2.457+	-0.203
Asian/Pacific Islander	6.676*	-1.109	5.251*	-2.052	1.972	-2.372
Black	2.233	0.049	1.413	-4.397*	2.064*	-8.231***
Hispanic	1.559	-1.638	2.127	0.271	1.139	-2.902
City	1.441	0.809	1.368	1.802	0.684	1.172
Low area income	-4.179*	-1.827	-6.165**	1.663	-1.367	3.371+
High area income	-0.017	-1.081	-1.795	-0.760	1.066	-0.843
Low minority percent	4.590+	1.290	0.974	0.024	0.370	-0.705
High minority percent	0.048	2.638*	- 2.806	0.317	0.887	0.031
Social disorganization	-3.012	2.302	-1.123	-4.768*	0.377	-4.831*
Level 1 & 2 variance	177.41***	72.189***	73.374***	215.529***	36.688***	231.64***
Standard deviation	13.320	8.496	8.566	14.681	6.057	15.219
Level 3 variance	29.370***	10.956***	120.76***	37.070***	7.510***	19.749*
Standard deviation	5.419	3.309	10.989	6.088	2.740	4.444

*** = $p < .000$, ** = $p < .01$, * = $p < .05$, + = $p < .10$

TABLE 6. Multivariate Models of Math Growth, Year-Round & 9-Month

<i>Math</i>	Year-Round Enrollment			9-Month Enrollment		
	Yearly	Summer	9-month	Yearly	Summer	9-month
Intercept	24.017***	2.342***	21.356***	26.846***	1.226***	25.668***
Age	0.107	0.262**	- 0.062	0.171+	0.075	0.030
Gender	2.531**	0.929	0.286	1.599+	0.068	1.704+
Single parent	-1.385	-1.591+	0.354	0.358	-0.996	1.164
Repeated kindergarten	-3.846+	-0.046	-2.619	-9.857***	-1.684	-8.328***
Full day kindergarten	-2.398*	-0.829	-1.874	-1.030	0.484	-2.442*
Pre-school program	-0.128	-0.032	-0.473	3.060	- 0.514	3.738
Summer type	0.506	1.621+	--	--	--	--
Low social class	-1.608	-3.842***	2.122	-2.081	-1.039	-3.507*
Mid low social class	1.106	-0.924	2.104	3.007+	-0.954	2.592
Mid high social class	3.008*	-0.786	3.943*	4.061**	-0.649	1.210
High social class	2.956+	-1.179	4.177*	3.589*	-0.024	0.949
Asian/Pacific Islander	-3.990*	3.019	-3.663	-2.403	1.227	- 4.471*
Black	-0.705	1.372	-0.103	-5.839***	0.778	- 4.909*
Hispanic	-1.819	0.362	-0.553+	-3.086*	0.254	-2.246
City	1.870+	-0.418	1.193	0.709	-0.388	1.261
Low area income	0.851	0.933	-0.493	1.883	-0.471	2.672
Upper area income	0.015	2.767*	-3.002	2.653*	1.530+	2.286
Low minority percent	0.361	-1.133	2.251	-1.338	-0.063	-1.426
High minority percent	0.990	1.677	-0.159	1.308	1.153	0.177
Social disorganization	-0.625	0.648	-0.002	-0.477	2.226*	-2.937+
Level 1 & 2 variance	86.623***	54.043***	56.566***	97.746***	35.967***	135.067***
Standard deviation	9.307	7.351	7.521	9.887	5.997	11.622
Level 3 variance	5.974**	3.380*	42.759***	6.918*	12.280***	16.747**
Standard deviation	2.444	1.839	6.539	2.630	3.504	4.092

*** = $p < .000$, ** = $p < .01$, * = $p < .05$, + = $p < .10$