EdWorkingPaper No. 22-630

A Multi-State, Student-Level Analysis of the Effects of the Four-Day School Week on Student Achievement and Growth

Emily Morton
NWEA

Paul Thompson
Oregon State University

Megan Kuhfeld
NWEA

Four-day school weeks are becoming increasingly common in the U.S., but their effect on achievement is not well-understood. Using a difference-in-differences approach, we conduct the most representative student-level analysis to date of the effects of four-day weeks on student achievement and within-year growth using NWEA MAP Growth data. We estimate significant negative effects of the schedule on spring reading achievement (-0.07 SD) and fall-to-spring gains in math (-0.05 SD) and reading (-0.06 SD). The negative effects of the schedule are larger in non-rural schools and for female students, and they may grow over time. Policymakers and practitioners will need to weigh the schedule's varied effects across contexts in their decisions regarding how and if to implement the policy.

VERSION: July 2023

Suggested citation: Morton, Emily, Paul Thompson, and Megan Kuhfeld. (2023). A Multi-State, Student-Level Analysis of the Effects of the Four-Day School Week on Student Achievement and Growth. (EdWorkingPaper: 22-630). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/p96h-8a41

A Multi-State, Student-Level Analysis of the Effects of the Four-Day School Week on Student Achievement and Growth

Emily Morton,[1] Paul N. Thompson,[2] and Megan Kuhfeld[3]

1. Research Scientist
   Center for School and Student Progress, NWEA
   121 NW Everett St, Portland, OR 97209

2. Associate Professor of Economics
   Oregon State University
   340 Bexell Hall, Corvallis, OR 97331

3. Senior Research Scientist
   Collaborative for Student Growth, NWEA
   121 NW Everett St, Portland, OR 97209

* Send correspondence to ermorton36@gmail.com

July 2023

Abstract

Four-day school weeks are becoming increasingly common in the U.S., but their effect on achievement is not well-understood. Using a difference-in-differences approach, we conduct the most representative student-level analysis to date of the effects of four-day weeks on student achievement and within-year growth using NWEA MAP Growth data. We estimate significant negative effects of the schedule on spring reading achievement (-0.07 SD) and fall-to-spring gains in math (-0.05 SD) and reading (-0.06 SD). The negative effects of the schedule are larger in non-rural schools and for female students, and they may grow over time. Policymakers and practitioners will need to weigh the schedule's varied effects across contexts in their decisions regarding how and if to implement the policy.

Keywords: four-day school week, rural education, achievement; education policy

1. Introduction

Growing teacher burnout and an ongoing teacher shortage have plagued U.S. education in the wake of the COVID-19 pandemic (Fortin & Fawcett, 2022). A recent National Education Association (NEA) survey found that nearly two-thirds of educators reported burn out and general stress from the pandemic were very serious issues for teachers. As of fall 2022, teacher resignation and turnover rates were at historic highs, contributing to a growing teacher shortage nationwide (Bacher-Hicks, Chi, & Orellana, 2023; Camp, Zamarro, & McGee, 2023; Goldhaber & Theobald, 2023). Growing vacancies and the increasing use of unqualified teachers have concerning implications for instructional quality, curriculum content, and overall student learning (Clotfelter, Ladd, & Vigdor, 2010; Kuhfeld, Soland, & Lewis, 2022; Krueger & Whitmore, 2001; Ronfeldt, Loeb, and Wycoff, 2013; Wedel, 2021). Operating in resource-constrained environments, many school districts face challenges recruiting and retaining teachers. Increasing teacher salaries is often financially and politically infeasible, prompting districts to offer non-monetary benefits to attract and retain educators (Brown, 2017; Moored & Frank, 2013; Tennent, 2018).

One such policy school districts have increasingly turned to in recent years is the four-day school week. This alternative school schedule eliminates one day per week from the school calendar and extends the length of the remaining four school days. Prior to the pandemic, more than 1,600 schools in over 650 school districts nationwide employed four-day school weeks (Thompson, et al., 2021). Since the pandemic, however, four-day school week use has grown to encompass over 2,100 schools in more than 850 school districts – with many of these recent adoptions undertaken to address teacher retention and teacher burnout issues (Barnes, 2022; Markham and Bankert, 2022; Tamez-Robledo, 2022).

With this rapid growth in four-day school week use, a growing body of research around the academic implications of these school schedules has emerged. A majority of these studies (Anderson & Walker, 2015; Morton, 2021; Morton, 2023; Thompson, 2021b; Thompson, et al., 2022; Thompson, et al., 2023) have examined four-day school week achievement impacts in individual states, including Colorado, Oklahoma, and Oregon. These state-specific studies, with the exception of Anderson and Walker (2015), find negative achievement effects of four-day school week adoption. The remaining studies (Kilburn, et al., 2021; Thompson & Ward, 2022) assess four-day school week achievement effects across multiple states, also finding negative achievement effects. Overall, these studies suggest relatively small, negative average effects (~-0.02 to -0.09 standard deviations (SD)) of four-day school weeks on student achievement in math and reading.

Similar to this previous work, this study also examines the impact of the four-day school week on student achievement and compares these achievement impacts with other commonly used teacher retention and recruitment policies. We use a difference-in-differences analysis that combines a panel dataset of student-level test scores from NWEA's Measures of Academic Progress (MAP) Growth fall and spring assessments from 2009-2019 for students in kindergarten through 8^{th} grade across six states with four-day school week implementation data. In line with estimates from the previous literature using spring test scores, we find a non-significant 0.03 SD average decline in math test scores and a significant 0.07 SD decline in reading and larger negative effects (-0.08 SD in math and -0.11 SD in reading) of four-day weeks implemented at non-rural schools. When looking at within-school year (i.e., fall-to-spring) gains, we find that four-day school weeks significantly decrease students' math and reading gains by 0.05 SD and 0.06 SD, respectively, and greater declines among non-rural schools, where students' fall-to-spring gains dropped 0.08 SD in math and 0.09 SD in reading. We find suggestive evidence that the negative effects of the schedule grow in magnitude over time and these effects appear to vary by student race and gender. Finally, we compare the schedule's impact on achievement to the impact of potential policy alternatives, such as increased teacher turnover or vacancies. We show that the four-day week's negative average effects on test scores are slightly larger than those associated with small increases in class sizes (Krueger, 1999; Krueger and Whitmore, 2001) or teacher turnover (Ronfeldt, Loeb, and Wycoff, 2013; Wedel, 2021) but smaller than declines associated with teacher vacancies. The negative impacts stand in contrast to null or positive achievement impacts of costly teacher retention and recruitment strategies, such as induction and mentoring programs (Glazerman, et al., 2010), teacher incentive programs (Figlio & Kenny, 2007; Springer, Swain, and Rodriguez, 2016; Speroni, et al., 2020), and certification bonuses (Cowan and Goldhaber, 2018).

While our study posits a similar research question to many previous four-day school week achievement studies, our study makes several key contributions. In terms of the four-day school week literature, it is the first study to assess four-day school week achievement impacts using student-level data across multiple states. Previous research has either conducted student-level assessments within a particular state (e.g., Thompson (2021b) in Oregon) or examined the four-day school week policy across multiple states using district-level achievement estimates (e.g., Kilburn, et al. (2021) or Thompson and Ward (2022)). Not only do student-level data allow us to control for within-student variation and past educational inputs, the multi-state nature of the data allows us to reconcile the results of this study with these previous studies that generated either within-state student level estimates or across-state district-level estimates.

This study also contributes novel estimates on the impact of four-day weeks on within-year growth. The previous literature on the achievement effects of the four-day school week exclusively examines year over year (often spring-to-spring) achievement changes. As the NWEA assessments are conducted during both fall and spring we are able to compare estimates of year-over-year changes in achievement to estimates of within-year (e.g., fall-to-spring) achievement gains. These within-student school-year gains are a more proximal and valid outcome measure of the four-day school week policy, as they do not include unnecessary statistical noise from summer learning that is unrelated to the weekly school schedule during the school year. These within-year estimates are also less strongly associated with students' entering achievement levels and other socioeconomic inequalities (Atteberry & McEachin, 2020). Finally, this paper contextualizes the teacher retention-achievement tradeoff surrounding this school schedule to determine how four-day school week achievement impacts compare to increased class sizes, teacher turnover and shortages, and traditional teacher retention and incentive policies. The findings herein have implications for current debates around four-day school weeks in many state legislatures – most notably in Minnesota, New Mexico, Oklahoma, and Missouri. Given the sustained recent growth of four-day school weeks in response to and following COVID-19 and the relevant ongoing legislation seeking to rollback use of this school schedule, research on the implementation and effects of this policy is unprecedentedly salient and consequential.

2. Background on Four-Day School Weeks

Districts implementing four-day school weeks, which are primarily small, rural districts,[1] typically increase the length of the four remaining school days and have either Fridays or Mondays off. These adjustments typically result in fewer annual days and hours at school for four-day week students relative to five-day week students (Thompson et al., 2021). However, schools may reorganize daily schedules such that the resulting typical differences in subject-specific instructional time on a four-day week schedule are unknown (Kilburn et al., 2021). Some school districts may also offer "fifth day" remedial and/or experiential learning opportunities.[2]

Four-day school weeks are made possible by state policies that require districts to meet a minimum number of instructional hours without mandating a minimum number of instructional days. The number of schools operating on a four-day school week has increased sharply, by over 600%, over the past two decades, from 257 schools across 108 districts in 1999 to 1,607 schools across 662 school districts in 24 states[3] in 2019 (Thompson et al., 2021). This estimate is also

[1] There are a few districts that are exceptions to this pattern; for example, an urban district in Colorado that serves 18,000 students adopted a four-day week at the start of the 2018-2019 school year and two urban districts in Arizona that each serve over 12,000 students adopted a four-day week for the 2020-2021 school year in response to COVID-19 (Altavena, 2020).

[2] There is only limited information on what happens in districts on the "fifth day" that students are not in school. A study that conducted interviews and surveys in the winter of 2019 of 18 rural and town school districts (located across three states) operating on four-day school weeks found that district-run services or programming for students on the fifth day were rare (Kilburn et al., 2021). Indeed, students of all ages reported being primarily at home on the fifth day, spending their time on the fifth day primarily on free time, chores, playing sports, and working for their family or at a job. The study also found that parents and students were highly satisfied with the four-day week and overwhelmingly reported (85% of parents and 95% of students) that they would choose a four-day week over a five-day week. However, whether fifth day programming, students' experiences, and families' perceptions of four-day school weeks are similar in other districts and states is unknown.

[3] States with at least one school operating on a four-day school week schedule during the 2018-2019 school year were: Alaska, Arizona, California, Colorado, Georgia, Idaho, Iowa, Kansas, Louisiana, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Dakota, Utah, Texas, Washington, and Wyoming. The states where the four-day week is particularly popular include Colorado, Idaho, Missouri, Montana, Oklahoma, Oregon, New Mexico, and South Dakota (Thompson, et al. 2021; Kilburn et al., 2021). Colorado is leading the trend, with over 60% of its districts using a four-day week as of the spring of 2021.

likely conservative, as four-day school week implementation is systematically tracked in only six states (Heyward, 2018). A recent survey (Thompson et al., 2021) suggests that, historically, school districts have turned to the four-day school week for financial savings, attendance-related issues (e.g., low attendance rates, missing school for appointments or athletics), and issues related to being in a rural area (e.g., long bus rides, time to work on family farms and ranches, student/teacher retention), yet these financial and attendance motivations do not appear to come to fruition in many instances.[4]

2.1 Previous Literature and Conceptual Framework

The main aim of this study is to assess the achievement impacts of the four-day school week. While the expected achievement impacts of the four-day school week are largely ambiguous, a growing literature has generally found reductions in achievement following four-day school week adoption in single state (Anderson & Walker, 2015; Morton, 2021; Morton, 2023; Thompson, 2021b; Thompson et al., 2022; Thompson et al., 2023) and multi-state settings (Kilburn et al., 2021; Thompson & Ward, 2022). These studies, summarized in Table 1, generally leverage panel data and difference-in-differences research designs to study the causal effects of attending a school with a four-day week as opposed to school with a five-day week.

[4] Previous research indicates that district expenditures decrease by only 1-2% on average as a result of the schedule change, and student attendance is not significantly affected (Anderson & Walker, 2015; Kilburn et al., 2021; Morton, 2021; Morton, 2023; Thompson, 2021a; Thompson, 2021b; Thompson et al., 2022). Some survey data suggest that the schedule does provide students with additional time at home with family and does reduce the weekly time students spend commuting to school by about 30 minutes relative to similar districts operating on a five-day week schedule (Kilburn et al., 2021). Whether improved student and teacher retention or any other intended consequences of four-day school weeks have been realized by the districts who have adopted them remain empirically unfounded.

Most closely aligned to the current study are two studies that pool achievement data across multiple states. Kilburn, et al. (2021) examined the achievement of 3rd to 8th graders across five states and found that four-day school week ELA achievement was between 0.145 and 0.229 SD lower and math achievement was between 0.144 and 0.189 SD lower than in comparable five-day school week districts eight years after initial four-day school week adoption. Thompson and Ward (2022) found smaller, but still negative (0.015 to 0.026 SD reductions) four-day school week achievement impacts, when pooling average school district achievement from 3rd to 8th graders across twelve states. The remaining four-day school week achievement literature assesses four-day school week implementation in individual state settings, finding mixed, but generally negative achievement effects across grade levels.

While intuition suggests many factors could be at play in driving these typically negative achievement effects, this literature has noted one key mechanism in particular – time in school. Thompson and Ward (2022) found that four-day school week districts with low amounts of time in school (i.e., in the bottom tercile of the time in school distribution, averaging about 30 hours of school per week) experience the most detrimental achievement effects, reductions on the order of 0.034 and 0.056 SD. Using changes in time in school as a result of the switch to the four-day school week in Oregon, Thompson (2021b) found that a one-hour decrease in time in school was related to a 0.018 SD reduction in math achievement and a 0.006 SD reduction in reading achievement.

Of note, the estimates from the studies that leverage district-level data are generally not statistically significant, whereas most of the estimates from studies using student-level data are statistically significant. This difference may be attributed to the relatively smaller sample sizes and lack of statistical power to detect small effects when using district-level data as opposed to student-level data. However, it is also possible that the significant negative effects estimated in the two studies using student-level data can be attributed to the specific implementation of four-day weeks in the state of Oregon, as both student-level studies used only data from Oregon (Thompson, 2021b; Thompson et al., 2021b). Therefore, the differences in the estimated effects in the existing research may be artifacts of the studies' level of data and/or represent meaningful state-based differences in the implementation and consequences of the four-day school week for student achievement. Our current study is able to reconcile this by using student-level achievement data across multiple states.

Additionally, the existing research on academic effects of the four-day week examines effects on standardized annual spring test scores (i.e., achievement status), which conflates effects of the policy during the school year with effects during the summer. Thus, a major contribution of the present study is its examination of the effects of the four-day school week on within-year fall-to-spring (school year) student test score gains, or growth, in addition to spring achievement. Because we would expect summer learning to be unrelated to the weekly school schedule during the school year, including summertime in the outcome measure theoretically includes unnecessary statistical noise in the outcome. This included additional noise would cause the standard errors of the estimate to be inflated, such that one would be less likely to detect a significant effect of a school-year policy on spring achievement as opposed to school year gains.

Furthermore, within-student school-year gains are not only a more proximal and valid outcome measure for the policy, but they also are less strongly associated with students' entering achievement levels and other socioeconomic inequalities (Atteberry & McEachin, 2020). Indeed, research shows that students' academic gains over time are a more effective and potentially less biased measure of school quality than students' performance on a test at one point in time (Chingos & West, 2015). Relatedly, growth is likely a better and more sensitive measure of the effect of the four-day school week compared to achievement because schools and districts have been shown to have more influence on growth than students' achievement level at any one point in time (Atteberry & McEachin, 2020; Reardon, 2019; Stiefel, Schwartz, & Rotenberg, 2011). Therefore, to better understand how four-day weeks affect students, it is important to parse the effects of the schedule on fall-to-spring gains. We compare these effects to the estimated effect on spring achievement in the same sample, as well as to the existing evidence on the effect of the schedule on spring achievement.

Finally, we consider heterogeneous effects of the four-day school week by school rurality and for different student subgroups. There is some empirical evidence that districts in more rural areas do not experience the same negative academic impacts of a four-day week as districts located in towns or suburbs. Using data from Oregon, Thompson et al. (2021) find that the negative effects of four-day weeks on achievement are explained entirely by the impacts in non-rural schools.[5] While there is no empirical evidence examining the mechanisms underlying this difference, there are some theoretical reasons that rural students may experience less negative academic impacts of the schedule. For example, rural schools who switch to a four-day week may recoup more class time for students who would otherwise miss class on Fridays due to travelling long distances for

[5] Thompson et al. (2022) finds that, among 11th grade students in Oregon, average negative achievement effects of four-day school weeks (math effect=-0.09 SD, *p*<.01; reading effect=-0.03 SD, *n.s.*) were explained entirely by the effects of the schedule among students in non-rural schools located in towns or suburbs (math effect=-0.13 SD, *p*<.01; reading effect=-.04 SD, *n.s.*). The average effect of the four-day week on achievement among 11th grade students attending rural four-day week schools was positive, though not statistically significant (math effect=0.08 SD, *n.s.*; reading effect=0.01 SD, *n.s.*).

athletics competitions, appointments, and other familial or extracurricular activities (Kilburn et al., 2021). It is also possible that four-day weeks are a more effective recruitment and retention strategy for teachers in rural areas relative to non-rural areas because rural districts often pay lower salaries, and the additional day off may allow teachers to pursue outside work to supplement their pay (Kilburn et al., 2021). This could subsequently impact rural student achievement if these policies keep higher quality teachers in these districts or enhance teacher effort on the remaining four school days. Estimating the impacts of the schedule separately for rural and non-rural districts across states is critical for informing ongoing policy debates regarding whether rural districts should be exempt from requirements that prohibit four-day weeks (e.g., Manley, 2023).

There are also empirical and theoretical reasons to expect that features of the four-day week, including longer school days and earlier start times, could differentially impact students based on their demographic characteristics, such as age, gender, and race. More specifically, middle school students may be more equipped to focus for longer periods of time than elementary students, but their achievement is also likely to be more negatively impacted by earlier start times (Heissel & Norris, 2018). Studies also suggest that female students may be better able to adapt to longer school days than male students, as elementary-aged girls score higher on average on measures of cognitive self-regulation than boys of the same age (e.g., Matthews, Marulis, & Williford, 2014). It is less clear how the features of the four-day school week may or may not disproportionately impact students from historically marginalized groups. While some foundational theories in education suggest that low-income students and students of color experience greater negative achievement impacts during time spent outside of school, more recent research on summer breaks and COVID-19 school closures indicate that disparities between groups within districts typically hold steady during these periods (Entwisle, Alexander, and Olson, 2000; Kuhfeld, 2019; Fahle et al., 2023).

The single Oregon-based study that has examined differences in academic impacts of the four-day school week by student characteristics finds greater negative impacts of the schedule on 8th grade students relative to 3rd grade students and finds no differences in effects by gender, race, or free- and reduced-price lunch eligibility. Whether these findings hold true in other states and whether there are larger negative impacts for the youngest students in kindergarten through second grade remain unknown (Thompson, 2021b).

3. Methods

3.1 Data

This study employs 11 years (2008-09 to 2018-19) of school-level data from a national database of school-level four-day school week adoption history (Thompson et al., 2021), school- and district-level demographic data from the National Center on Education Statistics (NCES) Common Core of Data (CCD), and seasonal student test event data from NWEA's anonymized longitudinal student achievement database.

The national four-day school week adoption dataset includes the year that each school adopted a four-day school week and whether they had a four-day school week each year from 2007-08 to 2018-19.[6] School- and district-level demographic data from the CCD include school enrollment, the percent of students in a school eligible for free and reduced price lunch (FRPL),

[6] These data were originally collected via direct correspondence with school districts and examination of historical lists of four-day school weeks from state departments of education. Based on the yearly lists of four-day school weeks and from responses of school districts an indicator for four-day school week use was generated for the years of this study. See Thompson, et al. (2021) for more details on the original data collection process.

the school student-teacher ratio, the percent of students in a district classified as English language learners (ELL), the percent of students in a district with an Individualized Education Program (IEP), a district's total current expenditures per pupil, and the urbanicity of the district.

3.1.1 NWEA MAP Growth Assessment Data

The test score data used in this study are from NWEA's Growth Research Database (GRD). School districts use NWEA's Measures of Academic Progress (MAP) Growth assessments to monitor elementary and secondary students' reading and mathematics growth throughout the school year, with assessments typically administered in the fall, winter, and spring. We use the fall and spring test scores, standardized by grade and term, to calculate fall-to-spring (school year) achievement gains. Districts choose the week(s) that they administer the assessment each term; for this reason, we control for test date or time between tests in each of our empirical analyses. For each student, we record fall-to-spring gains in the relevant spring term and drop the fall term data from the dataset. The NWEA data also include demographic information, including student race/ethnicity and gender, though student-level socioeconomic status is not available.

3.2 Sample

The sample includes all MAP Growth test events from the six states in the NWEA data that had at least one school adopt a four-day school week during the study period: Colorado, Iowa, Kansas, Montana, North Dakota, and Wyoming. Consistent with previous research on this topic (Morton, 2023; Thompson, 2021b), districts located in cities are excluded from the sample, as no districts that adopt four-day weeks are located in cities (as defined by NCES[7]), such that city

[7] NCES definitions of city, suburban, town, and rural classifications can be found at: https://nces.ed.gov/programs/edge/docs/locale_classifications.pdf.

districts would not provide a useful comparison for treated districts. The final sample includes over 6 million total fall and spring test scores (in math and in reading) of approximately 1 million kindergarten- to eighth-grade students in over 1,700 schools across 619 districts.

Of this sample, 35 schools and 20 districts are ever treated, meaning they ever adopted a four-day school week by the 2018-19 school year.[8] Just over 12,000 students were tested in these schools over the course of the study period. Across the six states, these 20 districts represent an average of 40% of the districts that adopt a four-day school week during the study period in each state. The never-treated group includes all remaining tested students across the six states who attended five-day week schools that never adopted a four-day week during the study period. Counts of test score events, students, schools, and districts in the sample are displayed by subject, grades, and treatment group in Table 2. The representativeness of the four-day week and five-day week schools in our analytic sample of all schools in the same states are presented in Appendix Table A1. Although we observe some small differences, the five-day and four-day week schools in the analytic sample are broadly representative of the respective groups across the six states.

Descriptive statistics of the analytic sample are presented in Table 3. Students in the sample who attended schools that ever adopted a four-day week are less likely to be white, are more likely to be Native American, and have lower standardized scores on MAP Growth in math and reading than students who attended schools that never adopted a four-day week. As expected, the schools and districts that adopted four-day school weeks have smaller enrollments, have higher percentages of students eligible for FRPL, and are more rural and less suburban than five-day week

[8] Two of the 35 schools that adopt a four-day week during the study time period switch back to the five-day school week, or have a "transitory" four-day school week.

schools and districts. Differences in the characteristics of the students, schools, and districts in the pre- and post-four-day week adoption samples are generally small and not statistically significant. These descriptive differences are useful context for understanding the composition of the control and treatment samples, but they do not provide any insight into the causal effects of four-day school weeks.

3.3 Empirical Strategy

3.3.1 Difference-in-Differences

3.3.1.1 Average effects. To examine the causal impacts of the four-day week schedule on students' MAP Growth test scores, we use a quasi-experimental research design. More specifically, we leverage panel data and a quasi-experimental two-way fixed effects (TWFE) difference-in-differences (DID) research design to estimate the impact of four-day school weeks by comparing the changes over time in outcomes of students at schools with four-day school weeks to the contemporaneous changes in students at schools that never or did not yet have four-day school weeks. We first estimate variations of the following specification:

$$Y_{isdgt} = \pi_i + \delta_s + \theta_{gt} + \beta Fourday_{sdt} + \gamma TestDate_{isdgt} + \epsilon_{isdgt} \quad (1)$$

where Y_{isdgt} is the dependent variable of interest (i.e., grade-term standardized spring reading and math test scores or grade-term standardized fall-to-spring (school year) reading and math test score gains) for student *i* in term *t,* school *s,* district *d,* grade *g*. π_i are student fixed effects, δ_s are school fixed effects, and θ_{gt} are grade-term fixed effects for each spring term in the sample. $Fourday_{st}$ is an indicator variable that is equal to one each term that a student's school *s* has a four-day week schedule. $TestDate_{isdgt}$ represents the relevant date of a student's test (i.e., spring test date) or time between a student's two test dates (i.e., days between the fall and spring test in a school year) that corresponds to the specified dependent variable of interest. ϵ_{isdgt} is an idiosyncratic error term that accommodates clustering at the district level (Bertrand, Duflo, & Mullainathan, 2004).

We further examine the dynamic effects of the four-day school week to see if the effect of the schedule changes over time based on the number of years a school has had the schedule. We use the Granger causality test ("event study") to estimate the effect of the four-day school week on the specified outcome variables for the spring terms before and after a school adopts the schedule:

$$Y_{isdgt} = \pi_i + \delta_s + \theta_{gt} + \sum_{k=2}^{3} \beta_{-k} Fourday_{sd,t-k} + \sum_{k=0}^{3} \beta_{+k} Fourday_{sd,t+k} + \gamma TestDate_{isdgt} + \epsilon_{isdgt} (2)$$

where β_{-k} and β_{+k} respectively represent the "effect" of being k spring terms prior to or post adoption (k =0 for the first spring term the four-day week is adopted) relative to never adopting four-day weeks or being one spring term pre-adoption. Joint F-tests are employed to test the null hypothesis of a constant post-treatment "effect", $H_0\colon \beta_0 = \beta_1 = \beta_2 = \beta_3$. We present our results of this event study analysis conducted as an intent-to-treat (ITT) analysis that includes all schools that ever adopt a five-day week in the treatment group. The ITT analysis provides more conservative estimates of the effects of the schedule than a treatment-on-the-treated (TOT) analysis that excludes the two transitory four-day week schools from the analysis; nevertheless, in this case, the two approaches yield substantively similar results.[9]

3.3.1.2 Heterogeneous effects. We additionally investigate whether impacts of the four-day school week vary by school rurality and student characteristics. To test whether four-day school weeks vary by rurality, we delineate four-day school week adoption into two separate

[9] Estimates from the TOT analysis can be provided upon request.

treatments: (1) rural four-day school weeks and (2) non-rural[10] four-day school weeks. We then separately estimate Equation (1) for each of these definitions of *Fourday*. To examine differences in the effects of the four-day week by grade, we estimate Equation (1) separately by grade level. Finally, to examine heterogeneous effects of the four-day week based on student characteristics, we interact binary student-level demographic variables with the *Fourday* indicator in Equation (1) and report the interaction effect. The interaction effects can be interpreted as the extent to which four-day school weeks have more (or less) of an effect for the specified student subgroup relative to the specified omitted subgroup. We estimate interaction effects of the four-day school week with student gender and race.

3.3.1.3 Alternative Control Groups. Moreover, we conduct an additional robustness check that tests the sensitivity of our preferred DID estimates to the chosen control group (i.e., all students who took MAP Growth tests in a five-day week school in a state that had at least one school that administered MAP Growth and was operating on a four-day week). We estimate Equation (1) using an alternative control group restricted to students attending five-day week schools in rural locations. Though reduced in size and statistical power to detect an effect of the schedule, this control group provides a valuable comparison because treated schools are mostly (71%) located in rural areas; therefore, it is reasonable that rural five-day week districts could be a better counterfactual for districts that adopt four-day weeks than all non-city five-day week districts.

[10] Urban schools are omitted from our analytic sample, so non-rural four-day school week schools are located in either towns and suburban areas (as defined by NCES).

3.3.1.4 Validity of Empirical Design. Several important assumptions are embedded in the DID approach that can be tested using various robustness checks. We conduct tests to check for violation of the parallel trends assumption, evidence of selection into or out of treatment, and bias in our estimates due to variation in treatment timing. The rationale for each of these checks and the methodology we use to conduct each test are described in Appendix B. As detailed in the Results section herein, the results of the robustness tests do not indicate that any of the necessary assumptions for DID are violated in the present study.

4. Results

4.1 Difference-in-Differences

We examine the effects of the schedule on two outcomes: (1) spring achievement and (2) fall-to-spring (school year) achievement gains. We first examine average effects using the full sample of schools that adopt four-day school weeks, and then we examine heterogeneity in the effects based on school rurality and student characteristics.

4.1.1 Average Effects

The point estimates of the average effect of the four-day school week on students' math and reading spring achievement and school year gains are presented in Table 4. We include the fully unrestricted models in columns (1) and (4) and the school fixed effects models in columns (2) and (5) for completeness, but we focus our substantive interpretation on the point estimates produced by the preferred student fixed effects models, presented in columns (3) and (6). The results from the preferred model suggest that four-day school weeks have a negative average effect on students' spring test score performance and their school year gains.

Specifically, as presented in Panel A, when examining standardized spring test score outcomes, we find the student fixed effects model estimates a negative but non-statistically significant effect of the schedule on math scores (β=-0.03, *n.s.*) and a significant negative effect on reading scores (β=-0.07, p<.05). Effects of the four-day week on students' fall-to-spring (school year) test score gains are presented in Panel B of Table 4. The student fixed effects model estimates a 0.05 SD decrease (p<.10) in school year gains in math and a 0.06 SD decrease (p<.01) in school year gains in reading.

Taken together, the more precisely estimated negative effects of the four-day week on fall-to-spring gains relative to year-over-year spring achievement suggest that the effects of the four-day school week are likely concentrated during the school year. Indeed, the negative effects of the schedule on students' fall-to-spring gains are likely driving declines in spring achievement observed in the present study and previous literature.

We further examine the dynamic effects of four-day school weeks over time based on the event study post-adoption period estimates presented in Table 5. The results show that the magnitudes of the negative effects of the four-day school week generally increase (or at least hold steady) over time, with larger negative effects of the schedule observed in each of the years following implementation than the first year for all specifications. However, the differences between the effects each year are not statistically significant, as the joint F-tests conducted for each specification failed to reject a constant treatment effect over time.

To interrogate possible bias in our TWFE DID point estimates due to heterogeneous effects of treatment based on treatment timing (i.e., treatment cohort), we also re-estimate the event study presented in Table 5 using Sun and Abraham's (2021) interaction weighted estimator. These results, displayed in Appendix Table A2, are consistent with the results presented in Table 5, suggesting that variation in treatment timing is not strongly biasing the original estimates. Moreover, this finding aligns with that of other studies that have examined the influence of variation in treatment timing on DID analyses of four-day school weeks (Morton, 2023; Thompson & Ward, 2022).

Finally, we test the sensitivity of the student fixed effects estimates presented in Table 4 by repeating the analyses with an alternative, more restrictive control group. We conduct the same student fixed effects DID specified in Equation (1) but include only students attending five-day week schools in rural areas in the control group. Relative to the original estimates, we find (see Appendix Table A3) similar average effects of four-day weeks on spring achievement in math (β=-0.01, *n.s.*) and reading (β=-0.04, *n.s.*), and similar but larger negative average effects of four-day weeks on fall-to-spring gains in math (β=-0.08, p<.01) and reading (β=-0.09, p<.01).

4.1.2 Heterogeneous Effects

Examining how the effects of the four-day school week vary (or do not vary) across schools and students is important for further investigating what factors could be responsible for driving the observed negative average effects. We examine whether the estimated average effects of the four-day school week vary by school rurality and student characteristics.

4.1.2.1 Rurality. First, as presented in Table 6, we estimate the student fixed effects DID model specified in Equation (1) separately for rural schools that adopt four-day weeks (columns (2) and (5)) and non-rural schools that adopt the four-day week (columns (3) and (6)). For spring achievement, we find no detectable effect of rural four-day weeks on math (β=0.02, *n.s.*) or reading (β=-0.04, *n.s.*), but we find a significant negative effects of non-rural four-day school weeks: a

0.08 SD decrease (*p*<.05) in math scores and a 0.11 SD decrease (*p*<.01) in reading scores. When examining fall-to-spring gains, we find the rural four-day week decreases reading gains with trend-level significance by 0.04 SD (*p*<.10) but has no detectable effect on math gains (β=-0.01, *n.s.*); however, the non-rural four-day week significantly decreases students' math and reading gains, by 0.08 SD (*p*<.01) and 0.09 SD (*p*<.01) respectively.

We also test the effects of rural and non-rural four-day school weeks using an alternate, rural-only control group. In alignment with the original results presented in Table 6, we find (see Appendix Table A3) the negative effects of the schedule on spring achievement and fall-to-spring gains are larger for students attending non-rural four-day week schools (spring math β=-0.07, *p*<.05; spring reading β=-0.07, *p*<.05; fall-to-spring math β=-0.12, *p*<.01; fall-to-spring reading β=-0.11, *p*<.01) than for students attending rural four-day week schools (spring math β=0.05, *n.s.*; spring reading β=-0.01, *n.s.*; fall-to-spring math β=-0.04, *n.s.*; fall-to-spring reading β=-0.07, *p*<.01).

Separate event study specifications for rural and non-rural four-day week schools were conducted to further examine the dynamic effects of the schedule in each setting. As displayed in Appendix Tables A4 and A5, the pattern of increasingly negative effects of the four-day school week over time we observed in the overall sample is also observed, respectively, in the rural and non-rural four-day week school samples. Therefore, despite the smaller and less negative average effects of four-day school weeks in rural schools, the growing negative effects over time suggest that the average effect estimated herein may underestimate the potential effect of the schedule examined over a longer time period.

4.1.2.2 Student Characteristics. We further examine heterogeneous impacts of the four-day school week by student grade level, gender, and race. Results of the student fixed effects DID specification in Equation (1) conducted separately for each grade level are presented in Appendix Table A6. Overall, we do not observe any discernable pattern in the results that suggest the effects vary meaningfully for students in earlier versus later grades.

Results of the student fixed effects DID specifications from Equation (1) that additionally interact four-day school week status with student gender and race are presented in Appendix Table A7. We do not find statistically significant differences in the effects of four-day school weeks on spring achievement by gender or race. However, when examining fall-to-spring gains, we find that the four-day school week has a greater negative effect on female students' math (β=-0.04, p<0.01) and reading (β=-0.04, p<.05) gains than male students' gains.

We also find some significant differences by race: the four-day school week has a less negative effect on Native students' math gains (β=0.12, p<.01) and a greater negative effect on Hispanic students' math gains (β=-0.07, p<.01) relative to its effect on White students' math gains. For reading gains, our interaction effect estimates are also positive for Native students and negative for Hispanic students relative to White students, but the magnitudes of the interaction effects are smaller and not statistically significant.

5. Discussion

States, districts, and schools across the country are making policy decisions on the four-day school week with limited information about its effect on student achievement. The prior research fails to come to a consensus on the effect of the schedule, with the majority of rigorous studies estimating average effects on students' annual, spring state test scores that are not significantly different from zero. The present study addresses these limitations of the previous research by leveraging seasonal student test data from multiple states to provide a more granular estimate of the effect of four-day school weeks on student achievement and school-year growth. Specifically, this study uses panel student-level data for students in grades 3-8 across six states and a difference-in-differences research design to parse the effects of the four-day school week on annual spring student achievement and school-year growth and to examine differences in these outcomes by school and student characteristics.

5.1 Effects on Spring Achievement

Our findings on the average effect of the schedule on standardized spring test scores align with the existing research that estimates a +0.02 to -0.09 SD change in test scores from adopting the schedule. We find a non-significant 0.03 SD average decline in math test scores and a significant 0.07 SD decline in reading. Like Thompson et al. (2022), we find that these negative average effects are driven by the larger negative effects (-0.08 SD in math and -0.11 SD in reading) of four-day weeks implemented at non-rural schools (i.e., those located in a town or suburb), as opposed to those being implemented at rural schools. Putting the average effect sizes in perspective of other education interventions, the average effects of the four-day week may be considered "small" (<0.05 SD) for math achievement and "medium" (0.05 to <0.20 SD) for reading achievement (Kraft, 2020). However, the estimated effects on math and reading achievement in non-rural four-day week schools are "medium" and meaningful, approximately equal to a quarter of the estimated impact of a year of school on achievement in the fifth grade (Bloom et al., 2008).

At rural four-day week schools, the average estimated effects on spring test scores are small and not significantly different from zero.

However, our results also find some suggestive evidence that the effects become increasingly negative over time in both non-rural and rural districts, reaching "medium" sized effects in the third year of implementation. If the negative effects observed in the later post-treatment years hold or continue to grow larger, the average effects of rural four-day school weeks estimated herein may underestimate the average effect of the schedule for students exposed to it beyond four years. Thus, our findings provide further support for the argument that four-day school weeks may be implemented and/or experienced differently in rural areas such that they are less harmful for student achievement than in non-rural areas, but the schedule may still have substantial negative consequences for students' achievement and growth in rural districts over time.

The average effects of the schedule also mask differences in the effects of the schedule by student characteristics. More specifically, we find differences in effects of the schedule on school year gains by gender, such that the schedule negatively impacts female students more than male students, and by race, such that the schedule negatively impacts Hispanic students more and Native students less than White students. The magnitudes of the differences in the effects between female and male students are small, whereas, for math gains, the magnitudes of the differences in the effects for Hispanic and Native students are relatively larger. Again, however, the observed differential effects of the schedule by race may be explained entirely or in part by differences in the implementation of the schedule across schools and/or states as opposed to differences in the effect of the same implementation on students of different races. Examining the implementation factors that may explain the disproportionate impact of the schedule on different subgroups of students is a promising direction for future research.

5.2 Effects on Fall-to-Spring Gains

Estimating the effects of the policy on fall-to-spring gains in addition to spring achievement is a key contribution of this study, as the school year gains estimates provide a more proximal and valid measure of students' academic outcomes related to the four-day school week than spring achievement. As expected, when we estimate effects of the schedule on school year gains, the standard errors of the estimates are smaller, allowing us to estimate the impact of the schedule with more precision. We find that four-day school weeks significantly decrease students' math and reading gains by 0.05 SD and 0.06 SD respectively, both "medium" effect sizes (Kraft, 2020). The effect is again more negative among non-rural schools, where students' fall-to-spring gains dropped 0.08 SD in math and 0.09 SD in reading as an effect of their school adopting the schedule. Alternatively, the effects of the schedule on fall-to-spring gains of students at rural four-day week schools are less than half of those at non-rural four-day week schools: rural four-day week students' scores decline only by 0.01 SD in math (not statistically significant) and 0.04 SD in reading (significant at the statistical trend level).

A surprising finding in this study is that the declines in fall-to-spring gains and spring achievement are larger in reading than in math. These findings contrast Thompson's (2021b) and Thompson et al.'s (2022) results from Oregon as well as research that indicates school inputs have larger effects on math achievement than on reading (Jacob, 2005; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Various implementation factors could be related to the relative magnitudes of the effects on math and reading achievement and gains, such as the proportion of instructional time allocated to each subject when the schedule changed and how students spend their time on the "fifth day." Investigating the mechanisms and implementation factors underlying this finding, including how the four-day school week impacts subject-specific instructional time, is also an important direction for future research attempting to understand the effects of the schedule.

5.3 Contextualizing the Achievement Impacts Relative to Policy Alternatives

Finally, we compare the negative achievement effects of the four-day school week found in this study to the achievement impacts of other teacher recruitment and retention strategies. In the absence of the four-day school week, the most likely counterfactual for many school districts is maintaining the status quo of teaching vacancies – and associated class size increases that may result – or the use of unqualified or low-quality teachers to fill these vacant positions. Previous literature has suggested that both of these options, class size increases and unqualified or low-quality teachers, decrease student achievement. Another set of teacher recruitment and retention strategies target these outcomes through cost-increasing methods making them less comparable to the cost-reducing or cost-neutral four-day school week policy. These strategies, such as induction and mentoring programs, teacher incentive programs, and certification bonuses, have been found to increase teacher retention and student achievement.

We first consider the likely counterfactual options of increasing class size and hiring unqualified or low-quality teachers. The class size literature (see Jepsen, 2015 for a review of this literature) based on the Tennessee STAR experiment (e.g., Krueger, 1999) find that a one-student increase in class size decreases achievement by 0.048 SD. Research on class size using non-experimental techniques (see, Rivkin, Hanushek, & Kain, 2005; Jepsen & Rivkin, 2009), generally find smaller impacts on achievement, on the order of 0.01 SD, of a one-student increase in class size. A large literature (see, Rockoff, 2004; Koedel & Betts, 2007) has shown teacher quality is a key component in student achievement, as described through the educational production function. Teacher turnover has been shown to negatively impact a district's teacher quality by changing the composition of the teachers (Darling-Hammond & Sykes, 2003) who remain in the district over time (e.g., greater proportions of unqualified or low-quality teachers). Teacher turnover has also been linked to lower student achievement. Using simulated data, Hanushek and Rivkin (2010) find minimal impacts of replacing teachers with "rookie" (i.e., first-year teachers). Ronfeldt, Loeb, and Wyckoff (2013), however, find that a one standard deviation increase in teacher turnover decreases student achievement by 0.02 SD. They also find negative impacts for students with teachers that have remained in the district, providing evidence for the disruption effects of teacher turnover. Wedel (2021) finds the impact of an additional hour per week of instructional time on test scores is reduced by 0.019 SD when the teacher does not have a Bachelor's degree in the field they are teaching.

So how does the four-day school week stack up to these counterfactual options? Our estimates align with the previous research and indicate that the four-day school week negatively impacts achievement by about 0.03 to 0.07 SD. These results suggest that the four-day school week may be slightly worse for achievement than general teacher turnover effects. However, more work is needed to understand how the four-day school week impacts rates of teacher turnover and retention of existing teachers, as the four-day school week may be able to avoid the disruption effects of teacher turnover if it is an effective teacher retention tool. Given average U.S. class sizes are 16 to 21 students (NCES, 2018), one ongoing vacant teaching position would mean class sizes increasing by 8-10 students. Based on the estimates of the class size literature, this degree of class size change would result in achievement losses of between 0.08 and 0.48 SD. Thus, the four-day school week may be a desirable alternative to leaving a position vacant based on these estimates.

Many of the cost-increasing policy approaches to teacher retention have shown positive impacts on both teacher retention and student achievement. Glazerman, et al. (2010) finds that induction and mentoring programs have a non-significant, generally positive impact on teacher retention and increases student achievement by 0.11 to 0.2 SD. Cowan and Goldhaber (2018) find that providing a bonus for National Board certification increases teacher retention by 2.7 percentage points but has only a negligible impact on achievement. The most common forms of teacher retention policy are retention bonuses and performance incentive pay. Springer, Swain, and Rodriguez (2016) find that tested-subjects teachers receiving retention bonuses are 20 percentage points more likely to remain in their current school. These policies also lead to increases in student achievement on the order of 0.04 to 0.1 SD (Figlio & Kenny, 2007; Speroni, et al., 2020). To fully compare these policies with the four-day school week, however, more research is needed quantifying the impact of the four-day school week on teacher retention. Knowing this will better inform policymakers about the cost-retention-achievement tradeoffs that exist around the four-day school week and these other cost-inducing policies.

5.4 Limitations

The present study faces several key limitations. First, we are limited in our ability to generalize our findings to students and schools that do not administer NWEA assessments, and those schools may be fundamentally different from schools that do administer the assessments. However, the four-day school week schools in the analytic sample do represent 40% of the schools that adopted four-day weeks in the included states during the study period. Our results also may not generalize to the states beyond those included in this study, though the similarity of our findings and those of Thompson and Ward (2022), whose sample of 12 states includes ten states that are not in our sample, suggests that these negative effects of the schedule persist for certain implementations of the schedule (i.e., low time in school, non-rural areas) across states. Unlike Thompson and Ward (2022), we are not able to test differences in effects by time in school because the time in school data are not publicly available for some of the schools in the sample, and there is limited variation in the time school data across the districts for which time data are available. Our test score data are also limited because they do not capture the effects of the four-day week on students who move to another district that does not administer NWEA tests, and we cannot distinguish between a student no longer being enrolled at a district versus still being enrolled but not taking the NWEA test. Because we find no evidence that four-day weeks were differentially impacting school enrollment relative to five-day weeks in our sample, this missingness should not bias our estimates of the effect of the four-day week within the analytic sample. However, this missingness means that our results are not generalizable to mobile students.

5.5 Conclusion

For policymakers and practitioners, this study addresses previous uncertainty about the effects of four-day school weeks on academic outcomes and provides evidence supporting concerns about four-day school week effects on student achievement and growth, particularly for those implemented in non-rural areas. The estimated effects on math and reading gains during the school year are not “large” by the developing standards used to interpret effect sizes of education interventions (Kraft, 2020; What Works Clearinghouse, n.d.), but they are also not trivial. For the many districts and communities who have become very fond of the schedule, the evidence presented in this study suggests that how the four-day school week is implemented may be an important factor in its effects on students. As these schedules become increasingly popular across the country, understanding the key aspects of their implementation (e.g., annual subject-specific instructional time, daily start and end times, fifth-day opportunities) that enable the schedule to better support students' academic progress will be critical for informing future adoptions and continued use of the four-day school week.

References

Altavena, L. (June 10, 2020). 2 metro districts are moving to 4-day school weeks. How will working parents cope? *AZCENTRAL*. https://www.azcentral.com/story/news/local/arizona-education/2020/06/10/phoenix-glendale-cartwright-and-alhambra

Anderson, D. M., & Walker, M. B. (2015). Does shortening the school week impact student performance? Evidence from the four-day school week. *Education Finance and Policy*, *10*(3), 314-349.

Atteberry, A. C., & McEachin, A. J. (2020). Not where you start, but how much you grow: An addendum to the Coleman Report. *Educational Researcher*, *49*(9), 678-685.

Bacher-Hicks, A., Chi, O. L., & Orellana, A. (2023). Two years later: How COVID-19 has shaped the teacher workforce. *Educational Researcher*, *52*(4), 219-229.

Barnes, A. (2022, April 26). Can four-day school weeks salve teacher burnout? *The Hill*. https://thehill.com/changing-america/enrichment/education/3464149-can-4-day-school-weeks-salve-teacher-burnout/

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *The Quarterly Journal of Economics*, *119*(1), 249-275.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289-328.

Brown, E. (2017, May 27). With state budget in crisis, many Oklahoma schools hold classes four days a week. *Washington Post*. https://www.washingtonpost.com/local/education/with-state-budget-in-crisis-many-oklahoma-schools-hold-classes-four-days-a-week/2017/05/27/24f73288-3cb8-11e7-8854-21f359183e8c_story.html

Camp, A., Zamarro, G., & McGee, J. B. (2023). Movers, switchers, and exiters: Teacher turnover during COVID-19. University of Arkansas.

Chingos, M. M., & West, M. R. (2015). The uneven performance of Arizona's charter schools. *Educational Evaluation and Policy Analysis*, *37*(1_suppl), 120S-134S.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Cowan, J., & Goldhaber, D. (2018). Do bonuses affect teacher staffing and student achievement in high poverty schools? Evidence from an incentive for national board certified teachers in Washington State. *Economics of Education Review*, *65*, 138-152.

Darling-Hammond, L., & Sykes, G. (2003). Wanted, a national teacher supply policy for education: The right way to meet the" highly qualified teacher" challenge. *Education Policy Analysis Archives*, *11*, 33-33.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2000). Summer learning and home environment. *A notion at risk: Preserving public education as an engine for social mobility*, 9-30.

Fahle, E. M., Kane, N. T. J., Patterson, T., Reardon, S. F., Staiger, D. O., & Stuart, E. A. (2023). School District and Community Factors Associated With Learning Loss During the COVID-19 Pandemic.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*(5-6), 901-914.

Fortin J., & Fawcett, E. (2022, September 2). How Bad is the Teacher Shortage? Depends Where You Live. *The New York Times*. https://www.nytimes.com/2022/08/29/us/schools-teacher-shortages.html

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study. NCEE 2010-4027. *National Center for Education Evaluation and Regional Assistance*.

Goldhaber, D., & Theobald, R. (2023). Teacher turnover three years into the pandemic era: Evidence from Washington State. *CALDER Policy Brief No. 32*, *223*.

Hanushek, E. A., & Rivkin, S. G. (2010). *Constrained job matching: Does teacher job search harm disadvantaged urban schools?* (No. w15816). National Bureau of Economic Research.

Heissel, J. A., & Norris, S. (2018). Rise and shine the effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, *53*(4), 957-992.

Heyward, G. (2018). What Do We Actually Know about the Four-Day School Week?. *Center on Reinventing Public Education*. University of Washington Bothell, WA.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5-6), 761-796.

Jepsen, C. (2015). Class size: does it matter for student achievement?. *IZA World of Labor*.

Jepsen, C., & Rivkin, S. (2009). Class size reduction and student achievement the potential tradeoff between teacher quality and class size. *Journal of human resources*, *44*(1), 223-250.

Kilburn, M. R., Phillips, A., Gomez, C. J., Mariano, L. T., Doss, C. J., Troxel, W. M., Morton, E., & Estes, K. (2021). Does Four Equal Five? Implementation and Outcomes of the Four-Day School Week. RR-A373-1. *RAND Corporation*.

Koedel, C., & Betts, J. R. (2007). *Re-examining the role of teacher quality in the educational production function* (p. 49). National Center on Performance Incentives, Vanderbilt, Peabody College.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1-28.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, *114*(2), 497-532.

Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan*, *101*(1), 25-29.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, *51*(7), 500-506.

Manley, E. (April 14, 2023). Misouri Senate approves five-day school weeks, exempting rural districts. *FOX 2 Now*. https://fox2now.com/news/missouri/missouri-senate-approves-five-day-school-weeks-exempting-rural-districts/

Markham, D., & Bankert, A. (2022, September 20) Teacher shortage: Four-day school week incentive. *NewsNation*. https://www.newsnationnow.com/us-news/education/teacher-shortage-four-day-school-week/

Matthews, J. S., Marulis, L. M., & Williford, A. P. (2014). Gender processes in school functioning and the mediating role of cognitive self-regulation. *Journal of Applied Developmental Psychology*, *35*(3), 128-137.

Moored, Ginger, and Valerie Frank. (2013). *The Recession's Impact on Teacher Salaries. Washington DC: National Council on Teacher Quality*. https://www.nctq.org/dmsView/The_Recessions_Impact_On_Teacher_Salaries_NCTQ_Report

Morton, E. (2021). Effects of four-day school weeks on school finance and achievement: Evidence from Oklahoma. *Educational Researcher*, *50*(1), 30-40.

Morton, E. (2023). Effects of 4-Day School Weeks on Older Adolescents: Examining Impacts of the Schedule on Academic Achievement, Attendance, and Behavior in High School. *Educational Evaluation and Policy Analysis*, *45*(1), 52-78.

Nguyen, T., Lam, C., & Bruno, P. (2022). Is there a national teacher shortage? A systematic examination of reports of teacher shortages in the United States. EdWorkingPaper: 22-631.

Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(2), 40-68.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247-252.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Speroni, C., Wellington, A., Burkander, P., Chiang, H., Herrmann, M., & Hallgren, K. (2020). Do educator performance incentives help students? Evidence from the teacher incentive fund national evaluation. *Journal of Labor Economics*, *38*(3), 843-872.

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective teacher retention bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis*, *38*(2), 199-221.

Stiefel, L., Schwartz, A. E., & Rotenberg, A. (2011). What do AEFA Members say? Summary of results of an education finance and policy survey. *Education Finance and Policy*, *6*(2), 267-292.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.

Tamez-Robledo, N. (2022, June 22). Can Four-Day School Weeks Keep Teachers From Leaving? *EdSurge*. https://www.edsurge.com/news/2022-06-22-can-four-day-school-weeks-keep-teachers-from-leaving

Tennent, James. (2018, March 25). Are 4 Day Weeks the Answer to Problems Facing America's

School Districts? *Newsweek*. http://www.newsweek.com/four-day-weeks-school-districts-859419

Thompson, P. N. (2021a). Does a day lost equal dollars saved? The effects of four-day school weeks on school district expenditures. *National Tax Journal*, *74*(1).

Thompson, P. N. (2021b). Is four less than five? Effects of four-day school weeks on student achievement in Oregon. *Journal of Public Economics*, *193*, 104308.

Thompson, P. N., Gunter, K., Schuna, Jr, J. M., & Tomayko, E. J. (2021). Are all four-day school weeks created equal? A national assessment of four-day school week policy adoption and implementation. *Education Finance and Policy*, *16*(4), 558-583.

Thompson, P. N., Tomayko, E. J., Gunter, K. B., & Schuna Jr, J. (2022). Impacts of the four-day school week on high school achievement and educational engagement. *Education Economics*, 30(5), 527-539.

Thompson, P. N., Tomayko, E. J., Gunter, K. B., Schuna Jr, J., & McClelland, M. (2023). Impacts of the four-day school week on early elementary achievement. *Early Childhood Research Quarterly*, *63*, 264-277.

Thompson, P. N., & Ward, J. (2022). Only a matter of time? The role of time in school on four-day school week achievement impacts. *Economics of Education Review*, *86*, 102198.

Wedel, K. (2021). Instruction time and student achievement: The moderating role of teacher qualifications. *Economics of Education Review*, *85*, 102183.

What Works Clearinghouse. (n.d.). *What Works Clearninghouse procedures handbook, Version 4.0.* U.S. Department of Education, Institute of Education Science.

Table 1: Existing Literature on the Academic Effects of Four-Day School Weeks

Study	State(s)	Grades	Subject	Level	Effect size (SD)	Heterogeneous effects
Morton (2021)	OK	3-8	Math	District	-0.05	N/A
Morton (2021)	OK	3-8	ELA	District	-0.03	N/A
Thompson (2021b)	OR	3-8	Math	Student	-0.06**	Special education (+); 8th grade (-); No significant differences by race, gender, FRPL, ELL, or G/T designation
Thompson (2021b)	OR	3-8	Reading	Student	-0.04**	ELL (-); 8th grade (-); No significant differences by race, gender, FRPL, special education, or G/T designation
Kilburn et al. (2021)	ID, MO, NM, OK, SD	3-8	Math	District	-0.03	N/A
Kilburn et al. (2021)	ID, MO, NM, OK, SD	3-8	ELA	District	-0.04	N/A
Thompson & Ward (2022)	12 (AZ, GA, ID, KS, MN, MO, MT, NM, NV, OK, OR, SD)	3-8	Math	District	-0.03*	Time in school (+), significant only for lowest time in school group
Thompson & Ward (2022)	12 (AZ, GA, ID, KS, MN, MO, MT, NM, NV, OK, OR, SD)	3-8	ELA	District	-0.02	Time in school (+), significant only for lowest time in school group
Thompson et al. (2022)	OR	11	Math	Student	-0.09***	Rural (+)
Thompson et al. (2022)	OR	11	ELA	Student	-0.03	Rural (+)
Morton (2023)	OK	11	Math	District	+0.02	N/A
Morton (2023)	OK	11	ELA	District	-0.02	N/A

Notes. *p<.10, **p<.05, ***p<.01

Table 2: Math and Reading Sample Counts

Sample	All non-city five-day schools		Four-day week schools pre-adoption		Four-day week schools post-adoption	
	Math	Reading	Math	Reading	Math	Reading
Test events (fall and spring)						
K	303773	295989	1097	1097	3884	3506
1	412155	395808	1427	1430	4335	3975
2	624187	608760	2224	2230	6203	5298
3	763739	754253	2233	2238	5896	5590
4	778056	772339	1974	1973	5710	5107
5	783225	776379	1759	1739	5321	4683
6	809382	801314	1578	1566	5270	4906
7	811179	803905	1238	1239	3794	3785
8	790862	787714	1217	1206	3589	3588
Students						
K	74916	73402	750	750	1212	1135
1	79489	75983	781	817	982	924
2	110236	108252	1176	1169	1626	1448
3	119749	119874	1036	1003	1378	1318
4	111616	112111	840	867	1279	1173
5	113914	113445	849	836	1242	1107
6	129664	128795	903	884	1428	1358
7	143748	143448	721	734	1051	1037
8	185512	185500	860	862	1365	1395
Schools	1746	1743	35	35	35	35
Districts	597	598	20	20	20	20

Notes. Data are from NWEA's Growth Research Database (GRD) and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).

Table 3: Descriptive Statistics

	Analytic sample					
	Students attending non-city five-day schools		Students attending four-day schools pre-adoption		Students attending four-day schools post-adoption	
	Mean	SD	Mean	SD	Mean	SD
Standardized achievement and growth outcomes[a]						
Spring math score	0.07	0.96	-0.12	0.95	-0.06	0.94
Spring reading score	0.09	0.96	-0.07	0.97	-0.02	0.95
Fall-to-spring math gains	0.00	0.53	0.03	0.59	0.03	0.57
Fall-to-spring reading gains	0.00	0.58	0.03	0.65	0.04	0.62
Student characteristics						
% Black	2.3	–	0.5	–	0.9	–
% White	70.2	–	55.0	–	52.2	–
% Native	2.4	–	13.3	–	10.5	–
% Hispanic	11.9	–	15.4	–	18.2	–
% Asian	2.0	–	0.3	–	0.33	–
% Female	51.2	–	52.0	–	51.0	–
School characteristics						
Enrollment	463.4	254.0	239.2	102.8	242.7	108.8
% FRPL	36.1	21.6	56.1	22.3	54.6	21.5
Student-teacher ratio	15.2	3.5	12.7	3.3	13.2	3.4
District characteristics						
% ELL	6.0	8.5	7.9	8.3	6.9	6.3
% IEP	11.7	5.6	7.6	8.4	9.2	8.0
Total current expenditures per pupil ($ thousands)	10.5	3.0	11.9	5.0	11.5	5.3
% Rural	33.2	–	53.8	–	60.5	–
% Town	31.5	–	42.0	–	30.5	–
% Suburb	35.3	–	4.2	–	9.0	–

[a]Achievement and growth outcomes in math and reading are calculated based on the math and reading grade-term standardized NWEA MAP Growth test scores. Scores were standardized based on the mean and standard deviation of the population of students who tested using MAP Growth across all states included in the present study in each grade-term.

Notes. FRPL = Free- or reduced-price lunch. ELL = English Language Learner. IEP = Individualized Education Program. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).

Table 4: Effects of the Four-Day School Week on Student Achievement and Growth

	Dependent variables: Standardized NWEA MAP Growth test scores					
		Math			Reading	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spring achievement						
Fourday	-0.153*	-0.010	-0.033	-0.122*	-0.025	-0.072**
	(0.085)	(0.043)	(0.037)	(0.073)	(0.020)	(0.034)
Observations	3082417	3082383	2815199	3029301	3029269	2762353
R-squared	0.000	0.132	0.861	0.000	0.108	0.832
Panel B: Fall-to-spring gains						
Fourday	0.012	-0.035*	-0.051*	0.020	-0.010	-0.062***
	(0.023)	(0.020)	(0.029)	(0.021)	(0.021)	(0.017)
Observations	2748457	2748431	2496288	2702248	2702227	2451835
R-squared	0.025	0.050	0.310	0.011	0.027	0.300
Grade-term FE		x	x		x	x
School FE		x	x		x	x
Student FE			x			x

Notes. Data are from NWEA's Growth Research Database (GRD) and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).

*p< 0.10, **p< 0.05, ***p< 0.01

Table 5: Effects of the Four-Day School Week on Achievement and Growth (Event Study)

	Spring achievement		Fall-to-spring gains	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
3 years before	-0.008	0.010	0.003	0.003
	(0.041)	(0.054)	(0.034)	(0.050)
2 years before	0.035	0.039	0.016	0.033
	(0.036)	(0.050)	(0.027)	(0.043)
Adoption year	-0.012	-0.032	-0.037	-0.032
	(0.030)	(0.021)	(0.030)	(0.033)
1 year after	-0.012	-0.076**	-0.052**	-0.065***
	(0.037)	(0.035)	(0.023)	(0.022)
2 years after	-0.056	-0.115*	-0.087**	-0.072
	(0.056)	(0.067)	(0.037)	(0.045)
3 years after	-0.053	-0.134**	-0.041	-0.077
	(0.068)	(0.063)	(0.052)	(0.049)
Observations	2815438	2762591	2498597	2452069
R-squared	0.861	0.832	0.310	0.300
p-value: (H_0: $\beta_{-3} = \beta_{-2} = 0$)	0.108	0.192	0.838	0.640
p-value: (H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3$)	0.195	0.483	0.100	0.346

Notes. All models include grade-term FE and student FE. Data are from NWEA's Growth Research Database (GRD) and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
*p< 0.10, **p< 0.05, ***p< 0.01

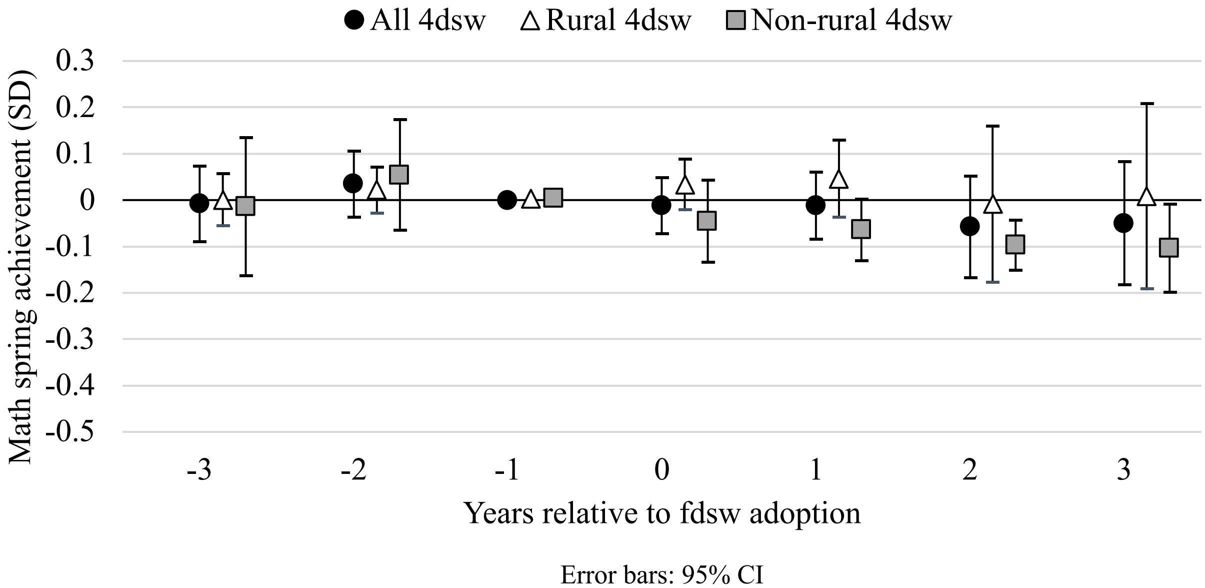
Table 6: Effects of the Four-Day School Week on Achievement and Growth by Rurality

	Dependent variables: Standardized NWEA MAP Growth test scores					
	Math			Reading		
	All four-day weeks	Rural four-day weeks	Non-rural four-day weeks	All four-day weeks	Rural four-day weeks	Non-rural four-day weeks
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spring achievement						
Fourday	-0.033	0.022	-0.081**	-0.072**	-0.040	-0.106***
	(0.037)	(0.042)	(0.031)	(0.034)	(0.040)	(0.036)
Observations	2815199	2801597	2800178	2762353	2750508	2747365
R-squared	0.861	0.861	0.861	0.832	0.832	0.832
Panel B: Fall-to-spring gains						
Fourday	-0.051*	-0.012	-0.083***	-0.062***	-0.038*	-0.086***
	(0.029)	(0.026)	(0.032)	(0.017)	(0.019)	(0.024)
Observations	2496288	2484033	2483325	2451835	2441504	2438885
R-squared	0.310	0.310	0.310	0.300	0.300	0.300
Grade-term FE	x	x	x	x	x	x
School FE	x	x	x	x	x	x
Student FE	x	x	x	x	x	x

Notes. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Figure 1: Effects of the Four-Day School Week on Spring Achievement (Event Study)

(a) Math



(b) Reading

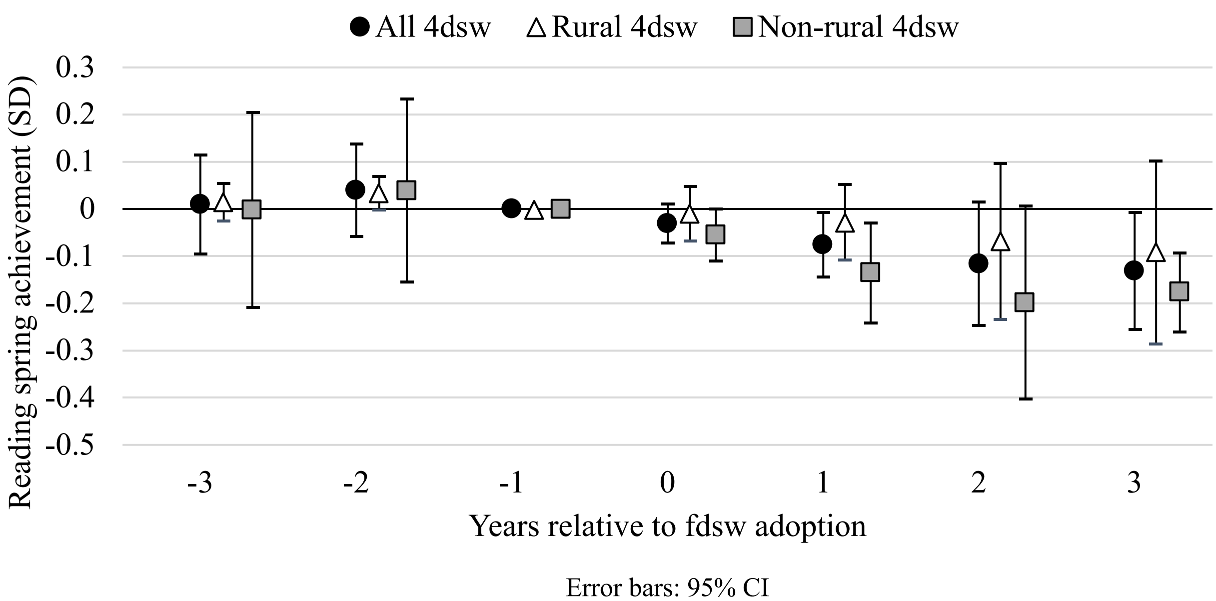
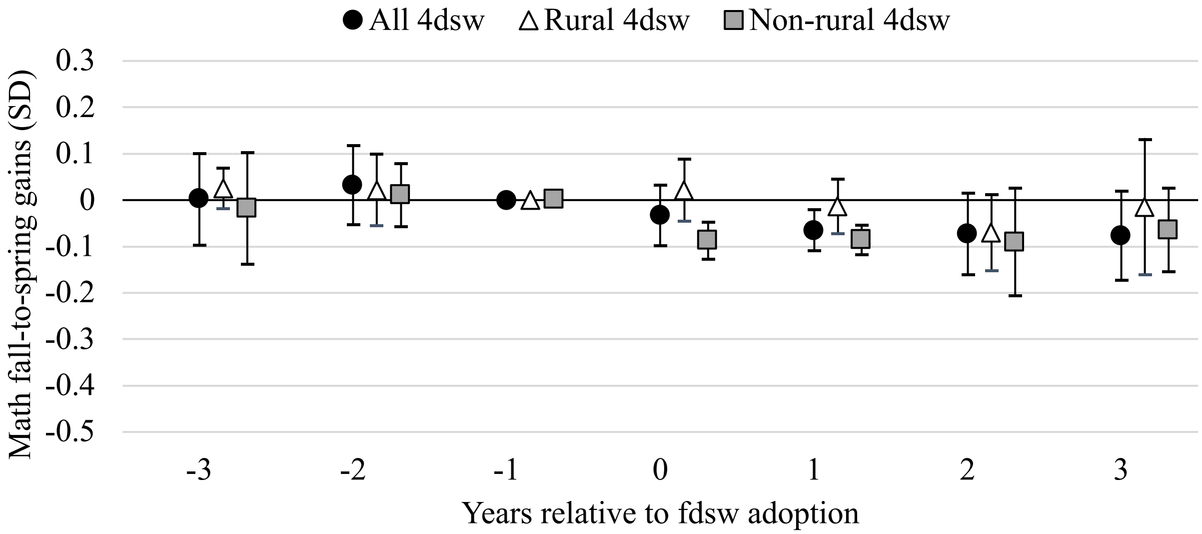


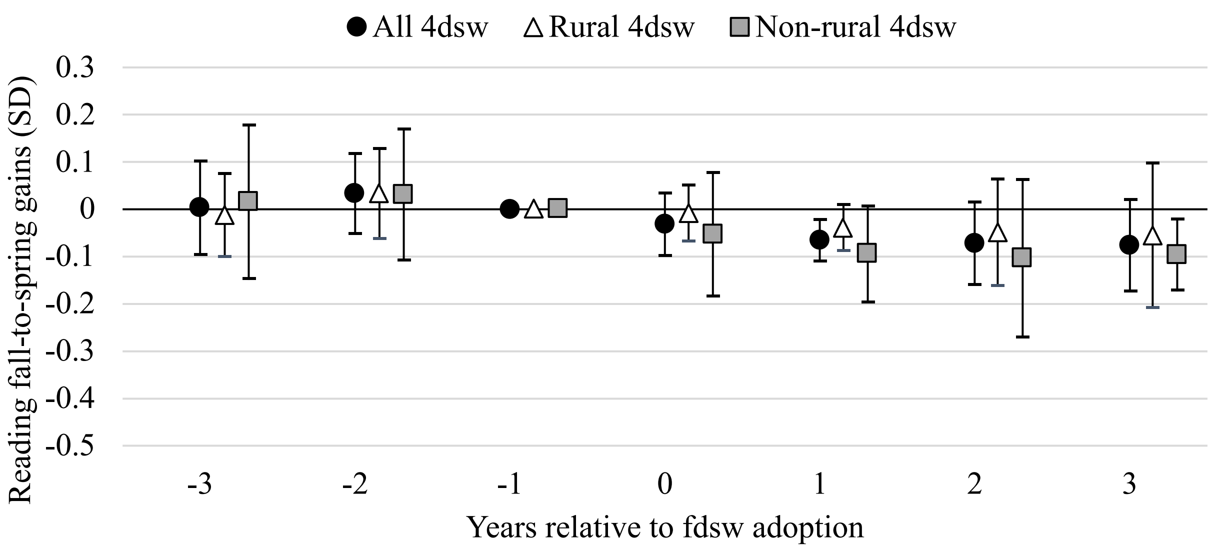
Figure 2: Effects of the Four-Day School Week on Fall-to-Spring Gains (Event Study)

(a) Math



Error bars: 95% CI

(b) Reading



Error bars: 95% CI

Appendices for "A Multi-State, Student-Level Analysis of the Effects of the Four-Day School Week on Student Achievement and Growth"

Appendix A

Appendix Table A1: Representativeness of Schools in the Analytic Sample

	Always five-day				Ever four-day[a]			
	Schools in the analytic sample		Population[b]		Schools in the analytic sample		Population[b]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Student characteristics								
% Black	1.9	–	0.6	–	4.4	–	0.9	–
% White	78.7	–	66.3	–	71.7	–	71.4	–
% Native	2.9	–	8.8	–	2.6	–	3.9	–
% Hispanic	12.3	–	21.4	–	16.4	–	21.2	–
% Asian	1.6	–	0.4	–	1.8	–	0.6	–
School characteristics								
Enrollment	342.4	211.4	196.6	112.5	325.3	242.3	167.5	176.6
% FRPL	38.2	20.5	54.5	19.8	44.1	23.9	48.1	22.4
Student-teacher ratio	14.4	3.3	12.0	3.3	15.8	14.5	12.0	4.5
District characteristics								
% ELL	4.8	7.6	7.5	7.3	7.1	10.1	4.7	6.9
% IEP	10.5	6.6	6.7	7.8	10.6	8.2	6.4	8.0
Total current expenditures per pupil ($ thousands)	11.0	3.1	13.4	5.4	10.9	3.8	13.3	5.8
% Rural	41.7	–	70.6	–	35.8	–	76.1	–
% Town	27.0	–	27.3	–	22.4	–	11.4	–
% Suburb	31.2	–	2.1	–	16.7	–	6.2	–
% City[c]	0.0	–	0.0	–	25.1	–	6.3	–

[a]The school ever implemented a four-day school week by the 2018-19 school year.
[b]The full population includes all public schools in the six states included in the study: Colorado, Iowa, Kansas, Montana, North Dakota, and Wyoming.
[c]City districts were excluded from the analytic sample because no city district was observed adopting a four-day school week.
Notes. School-level data are from the CCD. FRPL = Free- or reduced-price lunch. ELL = English Language Learner. IEP = Individualized Education Program. School-level data from the Common Core of Data (CCD) were averaged for each school from the period of 2008-09 to 2018-19, such that each school is represented once in the data.

Appendix Table A2: Effects of the Four-Day School Week on Student Achievement and Growth (Event Study), Reweighted to Account for Variation in Treatment Timing

	Spring achievement		Fall-to-spring gains	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
3 years before	-0.001	0.008	0.005	0.007
	(0.033)	(0.039)	(0.021)	(0.034)
2 years before	0.038	0.038	0.014	0.038*
	(0.028)	(0.028)	(0.015)	(0.022)
Adoption year	-0.012	-0.027	-0.035*	-0.035*
	(0.025)	(0.019)	(0.018)	(0.021)
1 year after	-0.021	-0.076***	-0.046***	-0.051***
	(0.018)	(0.027)	(0.016)	(0.017)
2 years after	-0.054*	-0.113**	-0.078***	-0.056***
	(0.028)	(0.046)	(0.021)	(0.014)
3 years after	-0.050	-0.134***	-0.032	-0.069***
	(0.031)	(0.048)	(0.033)	(0.025)
Observations	2817887	2762591	2496523	2452069
R-squared	0.8594	0.8306	0.3090	0.2996
p-value: $(H_0: \beta_{-3} = \beta_{-2} = 0)$	0.107	0.102	0.620	0.238
p-value: $(H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3)$	0.257	0.121	0.130	0.670

Notes. All models include grade-term FE and student FE. Data are from NWEA's Growth Research Database (GRD) and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
**p*< 0.10, ***p*< 0.05, ****p*< 0.01

Appendix Table A3: Effects of the Four-Day School Week on Achievement and Growth Using a Rural-Only Control Group

	Sample					
	Math			Reading		
	All four-day weeks	Rural four-day weeks	Non-rural four-day weeks	All four-day weeks	Rural four-day weeks	Non-rural four-day weeks
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spring achievement						
Fourday	-0.012	0.054	-0.065**	-0.039	-0.005	-0.071**
	(0.038)	(0.041)	(0.027)	(0.033)	(0.043)	(0.031)
Observations	772006	758609	757448	758260	746595	743725
R-squared	0.841	0.841	0.841	0.817	0.817	0.817
Panel B: Fall-to-spring gains						
Fourday	-0.084***	-0.041	-0.116***	-0.087***	-0.065***	-0.107***
	(0.030)	(0.028)	(0.034)	(0.016)	(0.023)	(0.021)
Observations	680821	668708	668228	668575	658362	655986
R-squared	0.316	0.317	0.315	0.304	0.303	0.303
Grade-term FE	x	x	x	x	x	x
School FE	x	x	x	x	x	x
Student FE	x	x	x	x	x	x

Notes. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
**p*<.10, ***p*<.05, ****p*<.01

Appendix Table A4: Effects of the Rural Four-Day School Week on Achievement and Growth

(Event Study)

	Spring achievement		Fall-to-spring gains	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
3 years before	-0.003	0.016	0.025	-0.013
	(0.029)	(0.021)	(0.022)	(0.045)
2 years before	0.018	0.036*	0.021	0.033
	(0.026)	(0.019)	(0.040)	(0.049)
Adoption year	0.030	-0.010	0.021	-0.009
	(0.028)	(0.030)	(0.035)	(0.030)
1 year after	0.044	-0.029	-0.051	-0.040
	(0.043)	(0.042)	(0.030)	(0.025)
2 years after	-0.010	-0.067	-0.073*	-0.049
	(0.087)	(0.086)	(0.042)	(0.059)
3 years after	0.000	-0.097	-0.016	-0.057
	(0.104)	(0.101)	(0.076)	(0.079)
Observations	2801836	2750746	2484268	2441738
R-squared	0.861	0.832	0.310	0.300
p-value: (H_0: $\beta_{-3} = \beta_{-2} = 0$)	0.520	0.162	0.457	0.730
p-value: (H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3$)	0.012	0.779	0.019	0.685

Notes. All models include grade-term FE and student FE. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
**p*< 0.10, ***p*< 0.05, ****p*< 0.01

Appendix Table A5: Effects of the Non-Rural Four-Day School Week on Achievement and Growth (Event Study)

	Spring achievement		Fall-to-spring gains	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
3 years before	-0.018	-0.001	-0.021	0.015
	(0.077)	(0.107)	(0.062)	(0.084)
2 years before	0.051	0.041	0.009	0.031
	(0.062)	(0.101)	(0.035)	(0.072)
Adoption year	-0.050	-0.055*	-0.091***	-0.055
	(0.046)	(0.028)	(0.021)	(0.067)
1 year after	-0.069**	-0.137**	-0.090***	-0.096*
	(0.034)	(0.055)	(0.016)	(0.053)
2 years after	-0.103***	-0.202*	-0.095	-0.106
	(0.028)	(0.106)	(0.060)	(0.086)
3 years after	-0.110**	-0.179***	-0.068	-0.098**
	(0.049)	(0.044)	(0.047)	(0.039)
Observations	2800178	2747365	2483325	2438885
R-squared	0.861	0.832	0.310	0.300
p-value: $(H_0: \beta_{-3} = \beta_{-2} = 0)$	0.050	0.207	0.792	0.849
p-value: $(H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3)$	0.592	0.000	0.070	0.058

Notes. All models include grade-term FE and student FE. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
*$p< 0.10$, **$p< 0.05$, ***$p< 0.01$

Appendix Table A6: Effects of the Four-Day School Week on Achievement and Growth by Student Grade Level

	Grade							
	1st	2nd	3rd	4th	5th	6th	7th	8th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Math spring achievement								
Fourday	-0.061	0.016	-0.140*	-0.022	-0.051	0.143**	-0.021	0.026
	(0.054)	(0.115)	(0.077)	(0.082)	(0.072)	(0.073)	(0.063)	(0.039)
Observations	261365	325233	483831	584577	592039	559062	579838	605689
R-squared	0.885	0.879	0.887	0.915	0.927	0.931	0.940	0.947
Panel B: Math fall-to-spring gains								
Fourday	-0.098	0.076	-0.281***	0.009	-0.071	0.089	-0.046	0.054
	(0.089)	(0.095)	(0.078)	(0.100)	(0.122)	(0.082)	(0.073)	(0.051)
Observations	180543	274653	405493	532859	538304	500092	510981	532126
R-squared	0.513	0.536	0.535	0.530	0.523	0.522	0.518	0.520
Panel C: Reading spring achievement								
Fourday	-0.074	0.035	-0.086	-0.040	-0.048	0.058	-0.074	-0.044
	(0.062)	(0.094)	(0.056)	(0.061)	(0.065)	(0.064)	(0.078)	(0.056)
Observations	73036	250893	308342	469891	575366	584127	548494	566583
R-squared	0.863	0.874	0.880	0.893	0.905	0.907	0.904	0.906
Panel D: Reading fall-to-spring gains								
Fourday	-0.055	0.089	-0.194*	-0.040	-0.081	0.023	-0.100	0.030
	(0.111)	(0.065)	(0.110)	(0.097)	(0.072)	(0.084)	(0.124)	(0.088)
Observations	50379	174781	260386	393697	523706	532150	490408	498706
R-squared	0.531	0.517	0.530	0.534	0.526	0.513	0.508	0.501

Notes. Data are from NWEA's Growth Research Database (GRD) and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).

**p*< 0.10, ***p*< 0.05, ****p*< 0.01

Appendix Table A7: Effects of the Four-Day School Week on Achievement and Growth by Student Gender and Race

	Dependent variables: Standardized NWEA MAP Growth test scores					
	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spring achievement						
Fourday	-0.033	-0.026	-0.020	-0.072**	-0.058	-0.048
	(0.037)	(0.041)	(0.039)	(0.034)	(0.039)	(0.032)
Fourday*Female		-0.013			-0.028	
		(0.017)			(0.022)	
Fourday*Native			0.028			-0.047
			(0.077)			(0.065)
Fourday*Hispanic			-0.053			-0.068
			(0.051)			(0.044)
Fourday*Other race			-0.041			-0.029
			(0.039)			(0.051)
Omitted subgroup	None	Male	White	None	Male	White
Observations	2815199	2814598	2814725	2762353	2761782	2761879
R-squared	0.861	0.861	0.861	0.832	0.832	0.832
Panel B: Fall-to-spring gains						
Fourday	-0.051*	-0.028	-0.054***	-0.062***	-0.041***	-0.062***
	(0.029)	(0.028)	(0.019)	(0.017)	(0.015)	(0.021)
Fourday*Female		-0.044***			-0.041**	
		(0.014)			(0.019)	
Fourday*Native			0.124***			0.054*
			(0.020)			(0.033)
Fourday*Hispanic			-0.069***			-0.038
			(0.023)			(0.027)
Fourday*Other race			0.008			0.011
			(0.022)			(0.061)
Omitted subgroup	None	Male	White	None	Male	White
Observations	2496288	2495824	2495824	2451835	2451380	2451367
R-squared	0.310	0.310	0.310	0.300	0.300	0.300

Notes. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
**p*< 0.10, ***p*< 0.05, ****p*< 0.01

Appendix B

Methodology for Robustness Checks

Parallel Trends

DID models assume that there would be “parallel trends” in the outcomes of the control units and treated units over time if the treated units had never received treatment. It is of course impossible to test this assumption in the post-treatment period, as we cannot observe any untreated treated students, but we can test this assumption during the pre-treatment period. The assumption would be violated in this study if, for example, students who ever attend a school with a four-day school week during the study period had decreasing spring test scores before their school adopted the four-day week relative to students at schools that never adopted the four-day week. In that case, one would not be able to determine whether any observed declines in four-day week students’ test scores should be attributed to the four-day week as opposed to the pre-existing negative trend for those students. Therefore, we use the event study equation specified in Equation (2) to test the parallel trends assumption during the pre-treatment period. To uphold the parallel trends assumption during the pre-treatment period, the estimated “effect” of attending a four-day school week school in the future (i.e., β_{-k}) should not be significantly different from zero. Joint F-tests are employed to test the null hypothesis of a constant pre-treatment “effect” equal to zero, $H_0: \beta_{-3} = \beta_{-2} = 0$.

The pre-treatment period point estimates in the event studies are presented in Table 4 and Appendix Tables A5 and A6, and depicted in Figures 1 and 2. These estimates are used to evaluate whether, conditional on student and grade-term fixed effects, outcomes of students attending schools that eventually adopt a four-day school week but had not yet adopted it (“3 years before,” "2 years before," etc.) were trending differently in the pre-treatment period relative to the outcomes of students at schools that never adopt a four-day week. The results fail to reject the null hypothesis that the students in the treatment and control groups are trending similarly during the pre-treatment period. The joint F-tests presented in Table 4 also fail to reject the null hypothesis that the combined pre-treatment "effects" of attending a school that would adopt a four-day week were constant and equal to zero, providing further support for the parallel trends assumption.

Selection Into and Out of Treatment

The validity of DID estimates also hinge on the assumption that selection into and out of treatment is random. The inclusion of fixed effects in the DID model addresses concerns about selection into treatment based on baseline differences in the treatment and control group. The fixed effects effectively enable comparisons of the changes *within* the treatment group over time to changes *within* the control group over time (i.e., the difference in the differences), so the average difference between the two groups is not a concern. The inclusion of student fixed effects in the DID model, therefore, enable us to evade most concerns about nonrandom school-level selection into treatment, as we examine changes in academic outcomes within students, and we do not observe any students who both switch schools and switch treatment status during the study period.[11] We likely do not observe any students who switch schools *and* switch treatment status in part because NWEA testing is a school or district decision, so students who leave their NWEA partner school for another school that does not use NWEA testing would not continue to be observed in our data.

[11] We likely do not observe any students who switch schools *and* switch treatment status in part because NWEA testing is a school or district decision, so students who leave their NWEA partner school for another school that does not use NWEA testing would not continue to be observed in our data.

Relatedly, another potential threat to the validity of the DID approach is if the composition of a school changes in response to adopting the four-day week such that students are systematically selecting into or out of four-day week schools. For example, if students were systematically disenrolling in a school following its adoption of the four-day week because they or their family disliked the four-day week, school enrollment would decrease in the year(s) following adoption. The composition of the students would also necessarily change such that a greater percentage of students and/or families would be in-favor of the four-day week. In such a case, it would again be impossible to know if any observed effects of the four-day week on students should be attributed to the four-day week as opposed to the changes in the population at their school. However, substantial spikes in enrollment and/or disenrollment are generally unlikely at schools that adopt four-day school weeks due to the rurality of the communities in which they are located. Nevertheless, we test for evidence of changing school composition at the school-level in response to four-day week adoption using the following specification:

$$\mathrm{Z}_{sdt} = \delta_s + \theta_t + \beta Fourday_{sdt} + \epsilon_{sdt} \tag{3}$$

where Z_{sdt} represents school, and district demographic characteristics related to school composition, including school enrollment, the percent of white students, the percent of Native American students, the percent of Hispanic students, the percent of students who are FRPL-eligible, the percent of students with an IEP, the percent of students classified as ELL, and the student-teacher ratio. All of the point estimates in Appendix Table B1 are substantively small and statistically insignificant, indicating that we fail to reject the null hypothesis that school composition is not changing during the post-treatment period as an effect of the four-day school week. Therefore, we find no strong evidence for selection into or out of treatment based on observables using this method.

Variation in Treatment Timing

A final robustness check is required to address an assumption embedded in TWFE DID specifications with variation in treatment timing, such as the specification used in this study. When there is variation in treatment timing and the effects of treatment vary over time, the TWFE DID estimator represents a weighted average of all two-group[12] by two-period (i.e., year) DID estimators (Goodman-Bacon, 2021). The weights on each 2x2 comparison are determined by the proportion of students in the treatment versus control group and the variance of the treatment dummy in the full sample each year. Whereas the proportion of treated students will be highest in comparisons made toward the end of the study period, the variance of treatment status will be largest in comparisons made in the middle of the study period. Therefore, a school adopting the four-day week during the study period could cause the estimated effect of the schedule to be underrepresented or overrepresented in the overall TWFE DID estimator. A developing body of literature shows that these weighted average fixed effects estimators can poorly represent the average treatment effect, and they are more likely to poorly represent the average treatment effect if there are heterogeneous treatment effects by treatment timing (Goodman-Bacon, 2021; de Chaisemartin & D'Haultfoeuille, 2020; Callaway & Sant'Anna, 2021; Sun & Abraham, 2021). To address this issue, we compare our event study point estimates to reweighted estimates calculated

[12] The analytic subsamples in this study include up to 11 "groups": never-treated districts, always-treated districts (only for the attendance analyses sample), and the nine cohorts of districts that adopted four-day weeks each year from 2011-2019.

using Sun's (2021) *eventstudyinteract* Stata package. This package uses an interaction weighted estimator to reweight the event study estimates in the following three steps: (1) estimating the interactions between relative time indicators (i.e., years pre- and post- four-day school week adoption) and cohort indicators (i.e., calendar year of four-day school week adoption), (2) estimating the cohort shares in each relative time group, and (3) estimating the weighted average of the estimated interactions with weights proportional to the estimated cohort shares.

References

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230.

de Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964-96.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.

Sun, L. (2021). eventstudyinteract: interaction weighted estimator for event study. https://github.com/lsun20/eventstudyinteract.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.

Appendix Table B1: Effects of the Four-Day School Week on School Composition

	Dependent variables: School characteristics							
	School-level variables						District-level variables	
	Enroll-ment	% White	% Native American	% Hispanic	% FRPL	Student-teacher ratio	% IEP	% ELL
Math sample								
Fourday	1.026	0.632	-0.488	-0.331	-0.118	0.425	-0.652*	-0.491
	(4.778)	(1.094)	(0.766)	(1.124)	(1.397)	(0.307)	(0.384)	(0.853)
Observations	12925	12895	11915	12798	12900	12550	10984	11903
R-squared	0.966	0.967	0.988	0.974	0.933	0.736	0.910	0.946
Reading sample								
Fourday	1.548	0.695	-0.482	-0.351	0.181	0.367	-0.615	-0.393
	(4.733)	(1.138)	(0.760)	(1.170)	(1.464)	(0.314)	(0.386)	(0.845)
Observations	12765	12734	11768	12635	12739	12393	10837	11749
R-squared	0.966	0.968	0.988	0.974	0.931	0.737	0.909	0.946
Grade-term FE	x	x	x	x	x	x	x	x
School FE	x	x	x	x	x	x	x	x

Notes. FRPL = Free- or reduced-price lunch. ELL = English Language Learner. IEP = Individualized Education Program. Data are from NWEA's Growth Research Database (GRD), the NCES Common Core of Data (CCD), and a national database of school-level four-day school week adoption history from 2008-09 to 2018-19 (Thompson et al., 2021).
**p*<.10, ***p*<.05, ****p*<.01.