1 Online Appendix for *Rise and Shine: The Effect of School Start Times on Academic Performance from Childhood through Puberty*

1.1 Robustness checks for mover definition

Our identifying variation comes from students who move between schools in different time zones in the Florida panhandle. Most of these moves are quite long-distance; the median move is 83 miles. The disruption inherent in such a move may have an independent effect on achievement, which is important to control for in our context. To help identify the effect of moving, as well as the effect of other school-level covariates, we include in our sample students who move within a time zone. This requires defining what constitutes a move by setting a threshold distance between the schools the student attended. Otherwise, graduating from middle school to high school would constitute a move. A high threshold has the advantage of making the move more likely to match a cross-time zone move in terms of disruptiveness; a low threshold increases sample size and precision.

We settled on a threshold of 25 miles, but our results are robust to other threshold choices. Table A1 presents estimates for 15, 20, 25, and 30 mile thresholds for math and reading outcomes. We also consider defining a move as any move between different school districts, although this will include students who move less disruptive distances, such as when families move to a nearby suburb that happens to be in a different district. Across all definitions, the results are broadly consistent. In math, the effect for prepubescent children ranges from 0.009 to 0.037 SDs; the effect for adolescents ranges from 0.067 to 0.084 SDs. In reading, the range is 0.034 to 0.061 for younger children and 0.044 to 0.057 for adolescents. The effects statistically differ from zero for adolescents for both math and reading across all distances.

1.2 Specification robustness checks

We include two sets of control variable robustness checks. First, in Table A2, we consider different levels of aggregation for the demographic share controls (FRL, male, black, Asian, and Hispanic). Instead of aggregating at the school-year level, as we do in our main results, we consider district-year, district third graders-year,[1] school-year, and school-grade-year. All specifications include age-gender dummies and an individual fixed effect. For each level of aggregation, we present one specification with no other controls, one that adds urban dummies and log income controls, and a final model that includes school size and student/teacher ratio.

Comparing across the rows of Table A2, the results are largely unchanged. In Panel A, all specifications show an effect size in math of 0.003-0.037 SDs for prepubescents, and 0.062-0.096 for adolescents. The effect is statistically significant at the the 1% level or better for adolescents but null for younger students. In reading, the estimates are also similar across specifications: 0.046-0.087 SDs for prepubescents, and 0.044-0.074 SDs for adolescents. The prepubescent effect is occasionally significant at the 5% level; the adolescent effect has a p-value of about 1%.

For absences, the inclusion of demographics (but not the level of aggregation) makes a substantive difference in the results. Comparing Columns 1-3 with Columns 4-15, the inclusion of demographic controls (at any level of aggregation) reduces the size of the suspension effect from about 1.5 percentage points and significant at the 1% level to about 0.8 percentage points and significant at the 10% level for prepubescents. The adolescent effects are generally null once we control for demographics. Since there may be significant between-school differences in policies for counting absences (and these may be correlated with school demographics), we think that the results with demographic controls are more trustworthy. It is therefore reassuring that they are the same regardless of the level of demographic aggregation.

[1] District third graders-year is the demographic means for the third graders in the given district-year.

Table A3 contains sur second control robustness check. Columns 1 and 3 restate our baseline results for math and reading. Columns 2 and 5 include controls for latitude; average sunrise times over the school year vary by about a minute over the north-south range of the panhandle[2] and this could conceivably have some affect on sleep (in contrast, the east-west variation in sunrise times from longitude is nearly 20 minutes, excluding the time zone change). The addition of latitude has a moderately sized but statistically insignificant effect on the prepubescent coefficients. The change in the adolescent coefficients is smaller.

In Columns 3 and 6 of Table A3 we test whether the inclusion of third grade district test scores as control variables affects the results. Third grade test scores are appealing as a summary measure of district quality, but may be endogeneous if start times affect performance for children in kindergarten to third grade. For this reason we do not include them in our main specification, but it is reassuring that they have little effect on the results.

1.3 Changes in school characteristics over the move

A potential threat to our identification strategy is changes in school and peer characteristics as students move between time zones. If students moving from CT to ET move to significantly worse schools, while ET-CT movers moved to better schools, it would not be surprising that student achievement declined upon entering ET and rose upon exiting. Because, on average, there is less sunlight before school in ET than in CT, this could generate a spurious positive relationship between relative school start times and academic achievement.

We consider this question directly in Table A4. We take the years directly before and after each move, and term these pairs of years a *moving episode*.[3] We then regress school- and zip-level characteristics on moving episode fixed effects and move indicators for the four different types of movers: Eastern-Eastern, Central-Central, Eastern-Central, and Central-Eastern. Each coefficient is a measure of the change in characteristics over the move. As outcomes, we consider the five school-level demographic share controls included in our preferred specification (percent FRL, male, black, Asian, and Hispanic), as well as school student/teacher ratio and zipcode-level median income as a measure of school and community resources.

The first two rows of Table A4 show that peer quality changed slightly over the move for within-time zone movers. ET-ET movers had 4.5 percentage points fewer FRL classmates; CT-CT movers had 1.7 percentage points fewer. School quality as measured by the student/teacher ratio increased slightly for both groups. Median income rose by $1,000 for within-ET movers and fell by $430 for within-CT movers. These differences are statistically significant, but none are particularly large or striking.

The cross-time zone movers tell a slightly different story. Eastward movers generally ended up in a richer area — 4.5 percentage points fewer FRL classmates and $5,700 higher median income — and had 14.0 percentage points more black classmates and 0.5 percentage points more Hispanic classmates. School quality as measured by the student/teacher ratio was unchanged. ET-CT movers saw approximately the opposite changes in medium income and percent of black students. The economic and peer changes may work in opposite directions in this case, making it unclear in which direction the overall bias goes. However, neither the inclusion of demographic controls (in Table 2) or income controls (in Table A2) substantively changes our results, suggesting that changes in peer characteristics have only a moderate effect on outcomes over the move, and do not significantly affect our results.

[2]The average disguises some larger differences over the year; but it is never larger than three minutes.

[3]Since occasionally a student will move in consecutive years, a small number of observations are repeated.

1.4 Performance trend before move

In Section 4.1, we show that test score trends are similar for all groups of movers in the years before the move. However, math scores trend *up*, which is somewhat surprising since the disruption of the upcoming move would be expected to reduce scores. Figure A1 show results from a regression of scale scores on time-until-move dummies and a fixed effect for the period until the move. This is identical to the regression displayed in Figure 1, but without controls. The Figure confirms that unconditionally, test scores trend down in both math and reading before a move. This is largely a result of removing the age-gender fixed effects, which soak up any time trend. Comparing across different groups of movers, the trends are slightly further apart than in the version with controls, but are still generally statistically indistinguishable.

1.5 Robustness checks for puberty definition

One of our main interests in this paper is how the effect of relative school start times varies with pubertal status. This requires a working definition of puberty, and there are several defensible alternatives. Pubertal development is typically measured with the Tanner Scale. There are two versions; one that uses levels of pubic hair to define the stages and another that uses breast and genital development. We rely on the pubic hair version of the Scale, which Campbell et al. (2012) indicate is more closely associated with pubertal changes in sleep patterns. They also note that changes in sleep patterns begin during Stage 3, so we use the age of median attainment (by gender) of Stage 3 as the definition of puberty.

Table A5 shows our main results with three alternative definitions of puberty: pubic hair Stage 2, pubic hair Stage 4, and breast/genital Stage 3. These changes typically shift the age of puberty by at most a year, and not necessarily for both genders. The results are largely unchanged, although slightly attenuated in some specifications. Because this definition of puberty is a worse fit for the underlying biological processes, this is unsurprising.

1.6 Estimates without interactions

Table A6 displays a version of our baseline model without an interaction between relative start time and pubertal status. Allowing for heterogeneity by pubertal status is important, but for completeness we have included this specification.

Across the rows, the change in sunlight is about 30 minutes over the time zone border. For both math and reading, the effect of moving start times one hour later is about the average of the child and adolescent effects from Table 2. In math, the estimated effect is 0.043 SD per hour by the final column, and the estimates are only occasionally statistically significant. In reading, the effect is 0.059 SD per hour by the final column, and the effect sizes are all significant to at least the 5% level in all estimates. The attendance results vary, with a decrease of 0.7 percentage points in absence per hour of sunlight by the final column.

1.7 PSID data definitions

In this paper, we demonstrate that students treated with later relative start times have higher academic achievement. However, we do not directly observe sleep levels in the academic outcomes dataset. To more concretely link changes in start times to changes in sleep, we use the Child Development Supplement of the Panel Study of Income Dynamics (PSID) to estimate the effect of the time zone boundary on sleep. The survey collected time use diaries for students on a weekend day and a weekday in the years 1997, 2002, and 2007. We include all states with a single time zone,[4] and all children

[4]The CDS does not geocode individuals at a sub-state level in the publicly available version, which precludes analysis using observations in states with multiple time zones — including Florida.

who were 6-19 during the survey and within 400 miles of the ET-CT time zone boundary. Our aim is descriptive, so we regress daily hours of sleep on a fully interacted set of dummies for puberty, CT, and whether the night was a weekend. In our preferred specification, we also include controls for gender, black/non-black, and FRL status. We expect that children in CT will have more sleep on weekdays when they face earlier relative start times, and those in ET will compensate with more sleep on weekends.

Table A7 contains the results. As discussed in Section 5.3, children in CT get 6 minutes more sleep per night during the week than children in ET; during puberty they get 17 minutes more. On the weekend, children in ET compensate for low levels of sleep during the week by sleeping 10 minutes more per night in the years before puberty and 19 minutes more while in puberty. We conservatively cluster at the state level. The coefficient for the difference in sleep between adolescents in CT and ET is significant at the 10% level; most others are not. Including student fixed effects suggests a slightly larger difference between the time zones: the decrease in sleep during puberty is 15 minutes smaller for adolescents in CT than in ET. This set of results corresponds to a pass-through rate of about 40-50% from school start times to sleep if Florida panhandle school start times are representative of the rest of the US near the ET-CT time zone boundary. This number is close to the 46% pass-through reported by Wahlstrom (1998).

1.8 Treatment bleed for schools near the time zone boundary

In the placebo analysis, we study how test scores change when students move east-west or west-east but *not* across the true time zone boundary. Ideally, we would examine within-time zone moves to and from the region directly adjacent to the boundary, to help test whether there are unobservable changes in the school or community environment that occur nearby, but not exactly at, the time zone boundary.

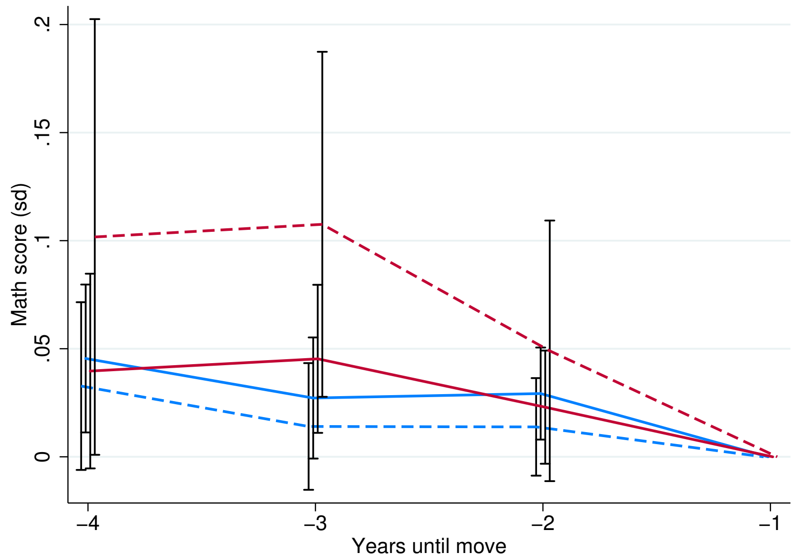
This approach will be problematic if there is an effect of being near the time zone boundary on school start times —- then, moving from directly beside the boundary in CT to a city fifty miles west could increase relative start times, directly increasing test scores. Figure A3 displays a nonparametric regression of relative start times on distance to the time zone boundary, estimated separately for each time zone. In the region directly adjacent to the boundary, start times veer towards the other time zone's norm, particularly for adolescents. We interpret this as the synchronization of start times across time zones, which allows parents to help their children prepare for school before going to work, whether or not they are commuting across time zones. This also means that start times are later for students moving west either from the region directly beside the boundary in CT, or *to* the region directly beside the boundary in ET.

In the main placebo results, we account for the treatment bleed across time zones by taking out a 25 mile "donut" around the time zone boundary. However, in the interest of completeness we include the unexcised version in Figure A4. The difference with Figure 5 is most stark in the puberty-time zone coefficient for math, where there is a consistent effect above the size of the true coefficient. Comparing between figures, removing the donut around the time zone boundary reduces the size of *all* placebo coefficients. The placebo effect is coming largely from individuals moving between the area close to the true time zone boundary and the rest of the study area, not individuals moving between areas far from the time zone boundary.

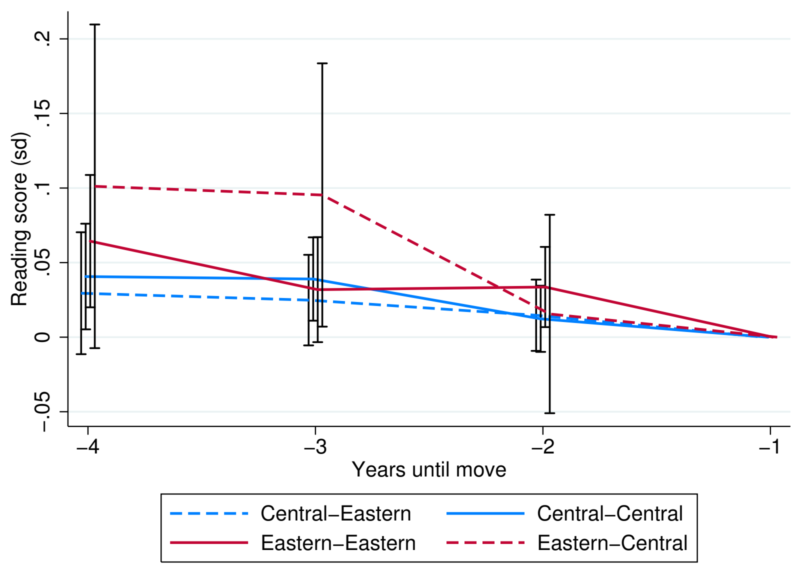
1.9 Online Appendix Figures

Figure A1: Pre-move trends in academic outcomes, by mover type without additional controls

(a) Math scores

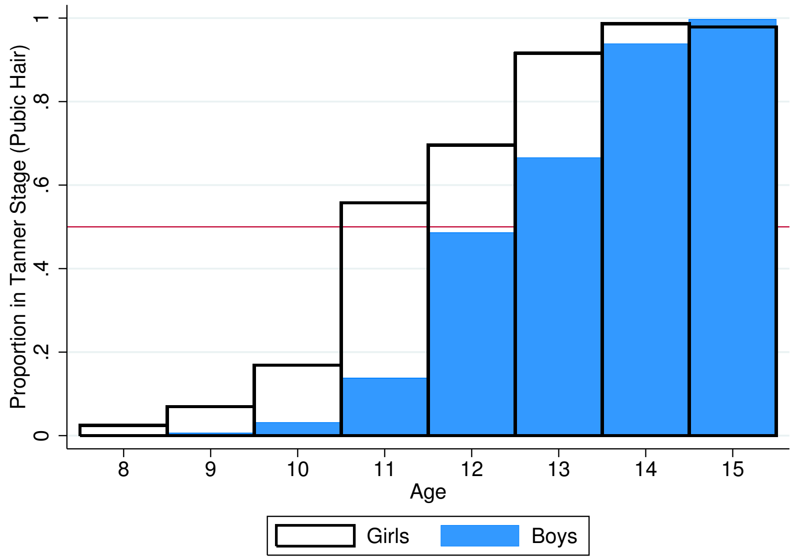


(b) Reading scores



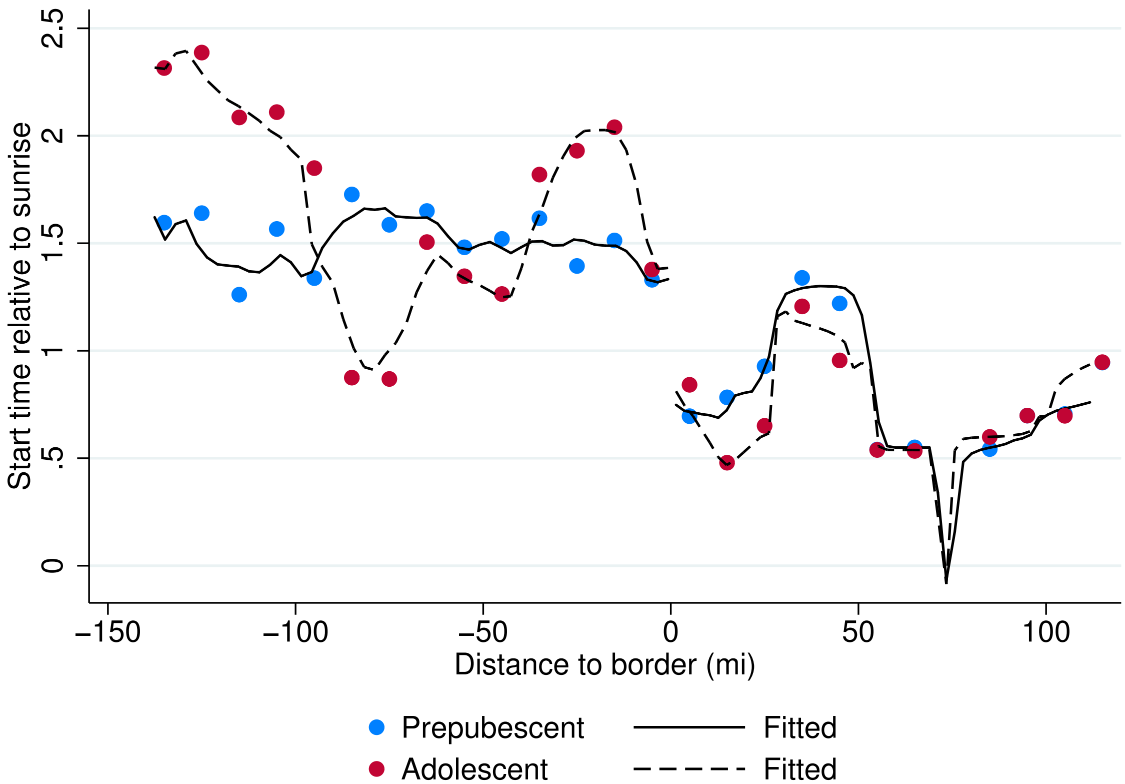
Displays the pre-move achievement trends for the four years leading up to a move of 25 miles. Results reported separately for four groups of movers: within CT, within ET, ET to CT, and CT to ET. Coefficients recovered from a regression of test scores on time-until-move dummies and a fixed effect for the period before the move. Standard errors are clustered at the individual level, and included as bars representing 95% confidence intervals.

Figure A2: Tanner stage 3 proportions by age and sex



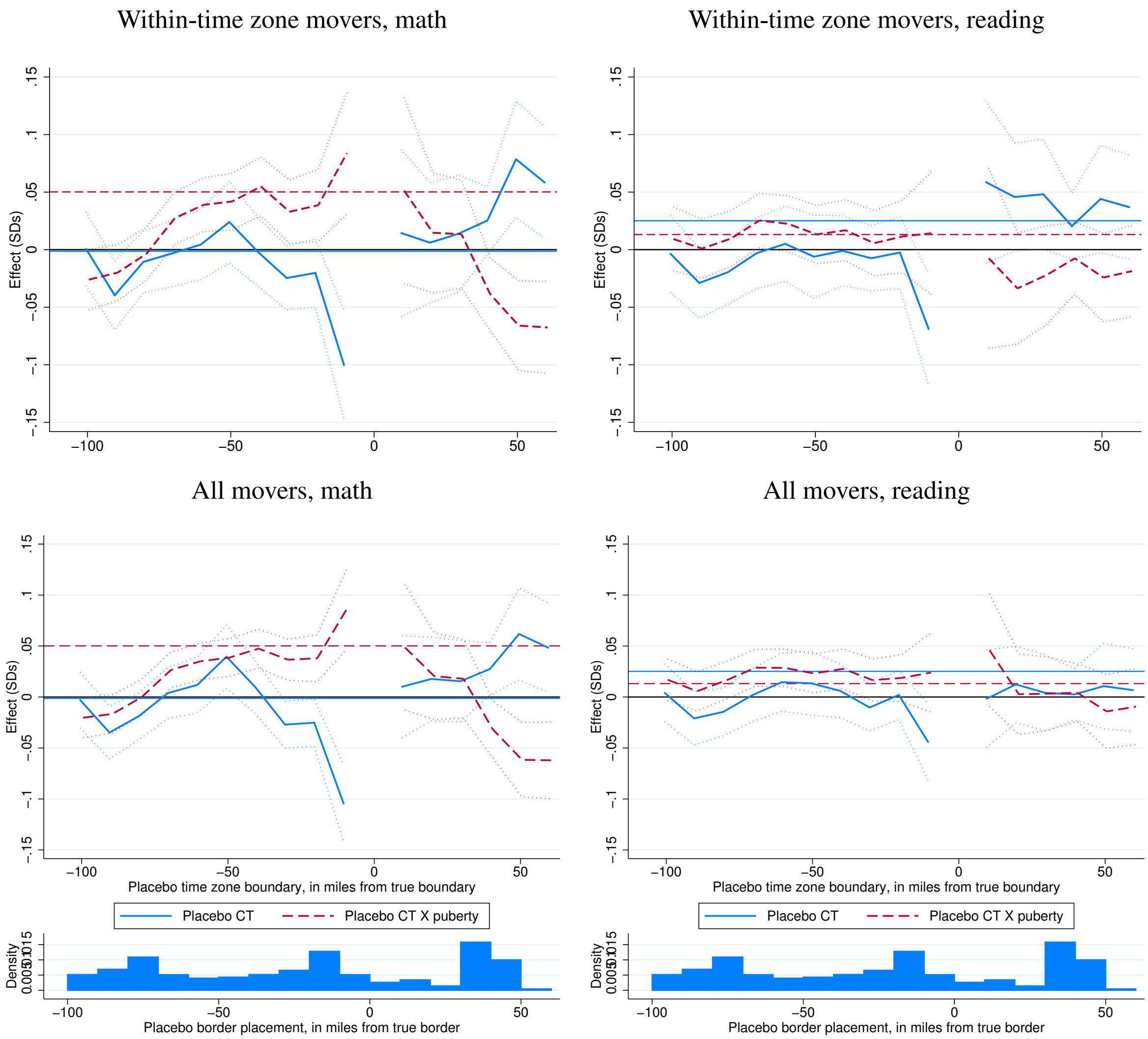
Displays proportion of children who had entered the Tanner Stage for pubic hair development at a given age for males and females. Horizontal line represents median child entering the stage.

Figure A3: Relative start time near the time zone boundary



Displays a nonparametric regression of relative start time (start time minus sunrise) on distance to the time zone boundary, estimated separately for each time zone. Scatter points are ten mile bin averages.

Figure A4: Effect of placebo time zones on academic achievement, no sample exclusion near true time zone boundary



Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Thin horizontal lines represent baseline coefficient estimates. We generate placebo time zones in ten mile increments from the true time zone boundary. Then, placebo coefficients are calculated from individual regressions of the outcome on the true time zone interacted with puberty, and the placebo time zone interacted with puberty. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors clustered at the individual level. We display results including and excluding cross-time zone movers.

1.10 Online Appendix Tables

Table A1: Academic outcomes on school start time for varying mover definitions, with student fixed effects

	Math (SDs)					Reading (SDs)				
	(1) dist	(2) 15 mi	(3) 20 mi	(4) 25 mi	(5) 30 mi	(6) dist	(7) 15 mi	(8) 20 mi	(9) 25 mi	(10) 30 mi
Start time - sunrise (h)	0.037	0.029	0.014	0.009	0.009	0.037	0.034	0.026	0.061*	0.053
	(0.034)	(0.033)	(0.031)	(0.035)	(0.037)	(0.036)	(0.034)	(0.032)	(0.036)	(0.038)
Start time X puberty	0.036**	0.038**	0.070***	0.073***	0.060***	0.007	0.011	0.018	-0.004	-0.008
	(0.018)	(0.017)	(0.018)	(0.019)	(0.022)	(0.019)	(0.018)	(0.018)	(0.020)	(0.023)
P(Start+Start X puberty=0)	0.001	0.002	0.000	0.001	0.004	0.029	0.025	0.033	0.014	0.049
Cragg-Donald F-stat	610.14	611.40	677.49	542.01	542.98	684.27	701.42	766.47	619.26	612.31
Number of students	33712	35744	28969	24768	21557	34144	36197	29393	25191	21957
Observations	143921	153462	120233	99835	84165	150800	160997	126110	104791	88408

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic), and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Academic and behavioral outcomes on start time, with student fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Math Test Scores (SDs)															
Start time - sunrise (h)	0.012	0.011	0.009	0.020	0.028	0.028	0.031	0.036	0.037	0.009	0.014	0.014	0.003	0.012	0.012
	(0.041)	(0.035)	(0.035)	(0.036)	(0.037)	(0.036)	(0.035)	(0.036)	(0.036)	(0.035)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)
Start time X puberty	0.073***	0.051***	0.054***	0.057***	0.037*	0.039*	0.065***	0.041**	0.043**	0.073***	0.050***	0.053***	0.076***	0.050***	0.053***
	(0.020)	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
P(Start+Start X puberty=0)	0.002	0.005	0.005	0.001	0.003	0.002	0.000	0.001	0.000	0.001	0.003	0.002	0.001	0.004	0.003
Cragg-Donald F-stat	405.14	593.76	604.37	588.76	593.74	606.81	580.48	599.33	612.93	542.01	640.62	655.01	534.48	638.53	655.08
Number of students	24768	23516	23516	24768	23516	23516	24545	23294	23294	24768	23516	23516	24765	23514	23514
Observations	99835	91853	91853	99835	91853	91853	98751	90852	90852	99835	91853	91853	99823	91846	91846
Panel B: Reading Test Scores (SDs)															
Start time - sunrise (h)	0.087**	0.061*	0.061*	0.081**	0.075**	0.074**	0.071**	0.065*	0.065*	0.061*	0.049	0.048	0.051	0.046	0.046
	(0.041)	(0.035)	(0.035)	(0.037)	(0.037)	(0.036)	(0.035)	(0.036)	(0.035)	(0.036)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)
Start time X puberty	-0.013	-0.009	-0.008	-0.023	-0.022	-0.022	-0.011	-0.013	-0.013	-0.004	-0.003	-0.003	0.000	-0.002	-0.002
	(0.021)	(0.020)	(0.020)	(0.022)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
P(Start+Start X puberty=0)	0.004	0.015	0.015	0.008	0.014	0.014	0.008	0.015	0.014	0.014	0.027	0.027	0.025	0.030	0.030
Cragg-Donald F-stat	486.65	679.86	687.26	637.22	648.26	671.04	656.76	675.89	697.05	619.26	729.44	746.01	616.60	725.65	742.75
Number of students	25191	24048	24048	25191	24048	24048	24963	23823	23823	25191	24048	24048	25189	24045	24045
Observations	104791	96788	96788	104791	96788	96788	103547	95641	95641	104791	96788	96788	104776	96776	96776
Panel C: Absence Rates															
Start time - sunrise (h)	-1.860***	-1.463***	-1.431***	-0.718	-0.709	-0.695	-0.848*	-0.789*	-0.772*	-0.869*	-0.874*	-0.859*	-0.965**	-0.904*	-0.880*
	(0.590)	(0.505)	(0.502)	(0.474)	(0.483)	(0.479)	(0.460)	(0.471)	(0.467)	(0.485)	(0.467)	(0.464)	(0.492)	(0.470)	(0.466)
Start time X puberty	0.857***	0.677**	0.637**	0.395	0.330	0.304	0.439	0.353	0.320	0.469*	0.384	0.365	0.491*	0.396	0.367
	(0.294)	(0.278)	(0.275)	(0.285)	(0.286)	(0.283)	(0.274)	(0.278)	(0.275)	(0.268)	(0.268)	(0.265)	(0.269)	(0.270)	(0.266)
P(Start+Start X puberty=0)	0.010	0.012	0.011	0.274	0.182	0.166	0.156	0.117	0.103	0.219	0.091	0.087	0.151	0.081	0.077
Cragg-Donald F-stat	274.18	413.70	416.25	425.38	431.70	439.86	453.02	458.47	467.24	383.62	451.74	458.44	373.38	447.12	454.86
Number of students	15906	15130	15130	15906	15130	15130	15906	15130	15130	15906	15130	15130	15903	15128	15128
Observations	66263	61128	61128	66263	61128	61128	66263	61128	61128	66263	61128	61128	66252	61122	61122
Urban and log income	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Size and S/T ratio	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
District controls	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
District grade 3 controls	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No
School controls	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No
School-grade controls	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Start time and its interaction with puberty are instrumented by time zone. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude, and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Outcomes on school start time, with latitude and school test grade scores

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Start time - sunrise (h)	0.009	-0.035	0.015	0.061*	0.035	0.051
	(0.035)	(0.033)	(0.037)	(0.036)	(0.034)	(0.037)
Start time X puberty	0.073***	0.085***	0.073***	-0.004	0.004	-0.001
	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
Latitude controls	No	Yes	No	No	Yes	No
Third grade district scores	No	No	Yes	No	No	Yes
P(Start+Start X puberty=0)	0.001	0.029	0.001	0.014	0.069	0.035
Cragg-Donald F-stat	542.01	631.95	508.46	619.26	715.55	589.27
Number of students	24768	24768	24288	25191	25191	24730
Observations	99835	99835	97483	104791	104791	102276

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic), and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Florida school and peer characteristics on move

	% FRL	% male	% black	% Hispanic	% Asian	S/T	Med income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Move, ET-ET	-4.494***	-0.452***	0.186	-0.100	0.263***	0.258***	1010.277*
	(0.726)	(0.118)	(0.801)	(0.224)	(0.059)	(0.081)	(601.359)
Move, CT-CT	-1.681***	-0.316***	-0.582**	0.110***	-0.011	0.190***	-429.606***
	(0.280)	(0.054)	(0.227)	(0.037)	(0.025)	(0.038)	(162.849)
Move, ET-CT	0.115	-0.009	-15.350***	0.025	0.426***	0.124	-4778.338***
	(0.923)	(0.162)	(1.015)	(0.183)	(0.084)	(0.103)	(731.901)
Move, CT-ET	-4.513***	-0.557***	13.965***	0.495***	0.023	0.113	5729.001***
	(0.939)	(0.163)	(1.010)	(0.166)	(0.088)	(0.101)	(752.117)
P(ET-CT=CT-ET)	0.002	0.029	0.000	0.105	0.003	0.944	0.000
Observations	31763	31763	31763	31763	31763	31763	27747

Dependent variable as noted in panel heading. Regression is of school/zip summary stat on move, with student X moving event FE. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Alternative definitions of puberty

	Math (SDs)				Reading (SDs)			
	(1) Preferred	(2) Stage 2	(3) Stage 4	(4) BG	(5) Preferred	(6) Stage 2	(7) Stage 4	(8) BG
Start time - sunrise (h)	0.009	0.011	0.032	0.025	0.061*	0.057	0.056	0.058
	(0.035)	(0.036)	(0.035)	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)
Start time X puberty	0.073***	0.064***	0.029	0.040**	-0.004	0.003	0.006	0.002
	(0.019)	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)	(0.021)	(0.020)
P(Start+Start X puberty=0)	0.001	0.003	0.005	0.008	0.014	0.012	0.002	0.010
Cragg-Donald F-stat	542.01	566.32	444.15	542.35	619.26	655.35	487.58	615.52
Number of students	24768	24768	24768	24768	25191	25191	25191	25191
Observations	99835	99835	99835	99835	104791	104791	104791	104791

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Academic and behavioral outcomes on start time, with student fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: First stage, relative start time (hours)							
CT (=1)	0.598***	0.475***	0.585***	0.547***	0.475***	0.584***	0.547***
	(0.015)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Observations	115778	115778	115778	115778	115778	115778	115778
Panel B: IV estimates, math test scores (SDs) on relative start time							
Start time - sunrise (h)	-0.005	0.047	0.048*	0.044	0.045	0.048*	0.043
	(0.019)	(0.034)	(0.028)	(0.029)	(0.034)	(0.028)	(0.029)
Cragg-Donald F-stat	2254.173	744.796	1120.532	1002.330	746.364	1120.434	1003.020
Panel C: IV estimates, reading test scores (SDs) on relative start times							
Start time - sunrise (h)	0.061***	0.081**	0.069**	0.059**	0.080**	0.069**	0.059**
	(0.019)	(0.032)	(0.028)	(0.028)	(0.032)	(0.028)	(0.028)
Cragg-Donald F-stat	2587.05	911.72	1209.23	1151.57	913.31	1209.80	1152.03
Panel D: IV estimates, absence rate (%) on relative start times							
Start time - sunrise (h)	-0.664**	-1.539***	-0.549	-0.670*	-1.510***	-0.559	-0.672*
	(0.275)	(0.501)	(0.391)	(0.407)	(0.499)	(0.389)	(0.405)
Longitude	No	Yes	Yes	Yes	Yes	Yes	Yes
District quality	No	No	Yes	No	No	Yes	No
School quality	No	No	No	Yes	No	No	Yes
Time since move	No	No	No	No	Yes	Yes	Yes
Cragg-Donald F-stat	1394.52	475.67	721.91	669.77	476.44	722.82	669.98

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Relative start time instrumented by time zone. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies and individual fixed effects. Sample size is fixed within panels: 34018 students and 115778 student-years in Panel A, 24768 students and 99835 student-years in Panel b, 25191 students and 104791 student-years in Panel C, and 15906 students and 66263 student-years in Panel D. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Hours of sleep by time zone

	(1)	(2)	(3)
Central	0.081	0.103	
	(0.088)	(0.131)	
Puberty	-0.451***	-0.804***	-0.676***
	(0.055)	(0.122)	(0.134)
Weekend	1.421***	1.192***	1.229***
	(0.102)	(0.158)	(0.158)
Central X weekend	-0.107	-0.166	-0.102
	(0.156)	(0.194)	(0.188)
Central X puberty	0.218	0.183	0.257
	(0.139)	(0.185)	(0.195)
Weekend X puberty	0.384***	0.616***	0.586***
	(0.087)	(0.161)	(0.150)
Central X wkend X puberty	-0.215	-0.149	-0.229
	(0.168)	(0.239)	(0.224)
P(Central + Central X weekend = 0)	0.830	0.566	
P(Central + Central X puberty = 0)	0.074	0.085	
Demographic controls	No	Yes	No
Student fixed effects	No	No	Yes
Observations	6,084	3,737	6,084

Dependent variable is hours of sleep per night. Sample is all children 6-19 in the Child Development Supplement of the Panel Study of Income Dynamics within 400 miles of the ET-CT time zone boundary in a state with a single time zone. Demographic controls in Column 2 include gender, race, and FRL status. Standard errors in parentheses and clustered at the state level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.