APPENDIX FOR ONLINE PUBLICATION

Appendix 1: Additional Tables and Figures

Table A1: Impact of attending school downwind, all students within 0.4 miles of a highway

	(1) Average FCAT	(2) Behavioral incident (0/1)	(3) Rate of absence
Downwind more than 60%	-0.0208	0.0486^{**}	0.0085^{**}
	(0.0202)	(0.0204)	(0.0039)
Observations	1,094,430	961,448	953,242
N Students	325,737	291,499	289,507

Notes: Results show estimates from our fixed effects model where downwind status is identified Each column shows results from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Standard errors clustered on zip and student are in parentheses. Coefficients of interest are an indicator variable for whether a student's school was downwind 60% or more of the time. Sample includes all students in any grade as long as they attend school within 0.4 miles of a highway.

$^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$

Table A2: Effects under various models

	(1) Controlling for only distance, number of roads, and changing schools	(2) Controlling for only distance, number of roads, changing schools, and school quality measures	(3) Main specification
	Panel A: Average FCAT		
Downwind more than 60%	-0.0393**	-0.0374*	-0.0398**
	(0.0192)	(0.0198)	(0.0192)
Observations	273,229	273,229	273,229
N students	107,463	107,463	107,463
	Panel B: Behavioral incident (0/1)		
Downwind more than 60%	0.0414**	0.0424**	0.0409**
	(0.0196)	(0.0198)	(0.0197)
Observations	237,684	237,684	237,684
N students	94,324	94,324	94,324
	Panel C: Rate of absence		
Downwind more than 60%	0.0050*	0.0052*	0.0053*
	(0.0029)	(0.0028)	(0.0029)
Observations	234,614	234,614	234,614
N students	93,249	93,249	93,249

Notes: Each row and column show results from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies. Column 2 additionally controls for other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. The final column is our primary specification and additionally controls for school demographic characteristics. Standard errors clustered on zip and id are in parentheses. Coefficients of interest are an indicator variable for whether a student's school was downwind 60% or more of the time. Sample includes only students in the years of policy-induced moves in grades 5, 6, 8, and 9.

$^{*}p < 0.1$ ** $p < 0.05$, *** $p < 0.01$

Table A3: Robustness to dropping zip codes without at least one downwind school

	(1) Average FCAT	(2) Behavioral incident (0/1)	(3) Rate of absence
Downwind more than 60%	-0.0447**	0.0381**	0.0052
	(0.0177)	(0.0193)	(0.0030)
Observations	16,805	14,899	14,675
N Students	7,039	6,288	6,211

Notes: Each column shows results from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Standard errors clustered on zip and id are in parentheses. Coefficients of interest are an indicator variable for whether a student's school was downwind 60% or more of the time. All models are estimated using only the "policy-induced" moves to middle/high school by dropping "K through 12" schools and limiting the sample only to students who change schools from fifth to sixth or eighth to ninth grade, and only zip-codes where there is at least one downwind school..

$^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$

Table A4: Alternative specifications: Distance of moves, cosine transformation of Distance, and wind intensity measures of treatment

	(1) Baseline	(2) Add control for distance of move	(3) Down 60, cosine distance dummies	(4) Downwind Intensity (0/1)	(5) Censored Intensity (0/1)
	Panel A: Downwind 60% of the time				
Downwind 60% of the time	-0.0398**	-0.0398**	-0.0499**		
	(0.0192)	(0.0193)	(0.0232)		
Distance of school move		0.0000			
		(0.0000)			
Move distance missing		-0.0067			
		(0.0059)			
	Panel B: Intensity Measures				
Downwind Intensity				-0.1395***	-0.1038*
				(0.0418)	(0.0546)
Observations	273,229	273,229	273,229	233,990	233,990

Each column shows results on test scores from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Column 1 replicates our preferred specification. Column 2 adds a control for distance of move from prior grade; it is equal to 0 for staying in the same school, 1 for a one-mile school move, and so on. Students with missing distance (because they lack a prior-year school) are imputed at the modal move distance of zero but do not differ from the overall sample. Column 3 replicates Column 1, but we replace the control for nearest road distance bin with distance bins calculated using the cosine of the angle of the wind direction and the nearest road location, to measure the approximate distance the pollution has to travel to get to the school if the road is parallel and pollution moves in a straight line. Column 4 replaces our downwind measure of treatment with a continuous intensity measure of wind exposure. Intensity is defined as the direction the wind is blowing in that hour normalized to be 1 when blowing directly at the school and 0 when blowing directly away. Column 5 shows results for a censored intensity measure where the intensity=0 if the wind blows parallel to the school or away. For more details on the construction of the intensity measure see appendix B. Standard errors clustered on zip and id are in parentheses. $^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$

Table A5: Impact of attending school downwind with different fixed effects

	(1) Average FCAT	(2) Behavioral incident (0/1)	(3) Rate of absence
Panel A: Grade-by-year FEs			
Downwind more than 60%	-0.0440**	0.0402**	0.0051*
	(0.0184)	(0.0197)	(0.0028)
Observations	273,229	237,684	234,614
N students	107,463	94,324	93,249
Mean of outcome	0.0206	0.3558	0.0618
Panel B: District-by-grade FEs			
Downwind more than 60%	-0.0790***	0.0388*	0.0009
	(0.0199)	(0.0201)	(0.0022)
Observations	273,229	237,684	234,614
N students	107,463	94,324	93,249
Mean of outcome	0.0206	0.3558	0.0618

Notes: Each row and column shows results from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Panel A additionally adds Grade-by-year fixed effects to the main model. Panel B additionally adds school-district by grade fixed effects to the main model. Standard errors clustered on zip and id are in parentheses. Coefficients of interest are an indicator variable for whether a student's school was downwind 60%. All models are estimated using only the "policy-induced" moves to middle/high school by dropping "K through 12" schools and limiting the sample only to students who change schools from fifth to sixth or eighth to ninth grade.
$^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$

Table A6: Robustness to clustering: Average FCAT

Clustering Scheme:	(1) School	(2) Student	(3) School & Student	(4) Zip code	(5) Zip code & Student
Downwind more than 60%	-0.0398** (0.0155)	-0.0398*** (0.0086)	-0.0398*** (0.0133)	-0.0398* (0.0241)	-0.0398** (0.0192)
Cluster on School	Yes	No	Yes	No	No
Cluster on Student	No	Yes	Yes	No	Yes
Cluster on Zip	No	No	No	Yes	Yes

Notes: Each column shows results from a different regression. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Standard errors clustered as noted are in parentheses. Coefficients of interest are an indicator variable for whether a student's school was downwind 60% or more of the time. All models are estimated using only the "policy-induced" moves to middle/high school by dropping "K through 12" schools and limiting the sample only to students who change schools from fifth to sixth or eighth to ninth grade.

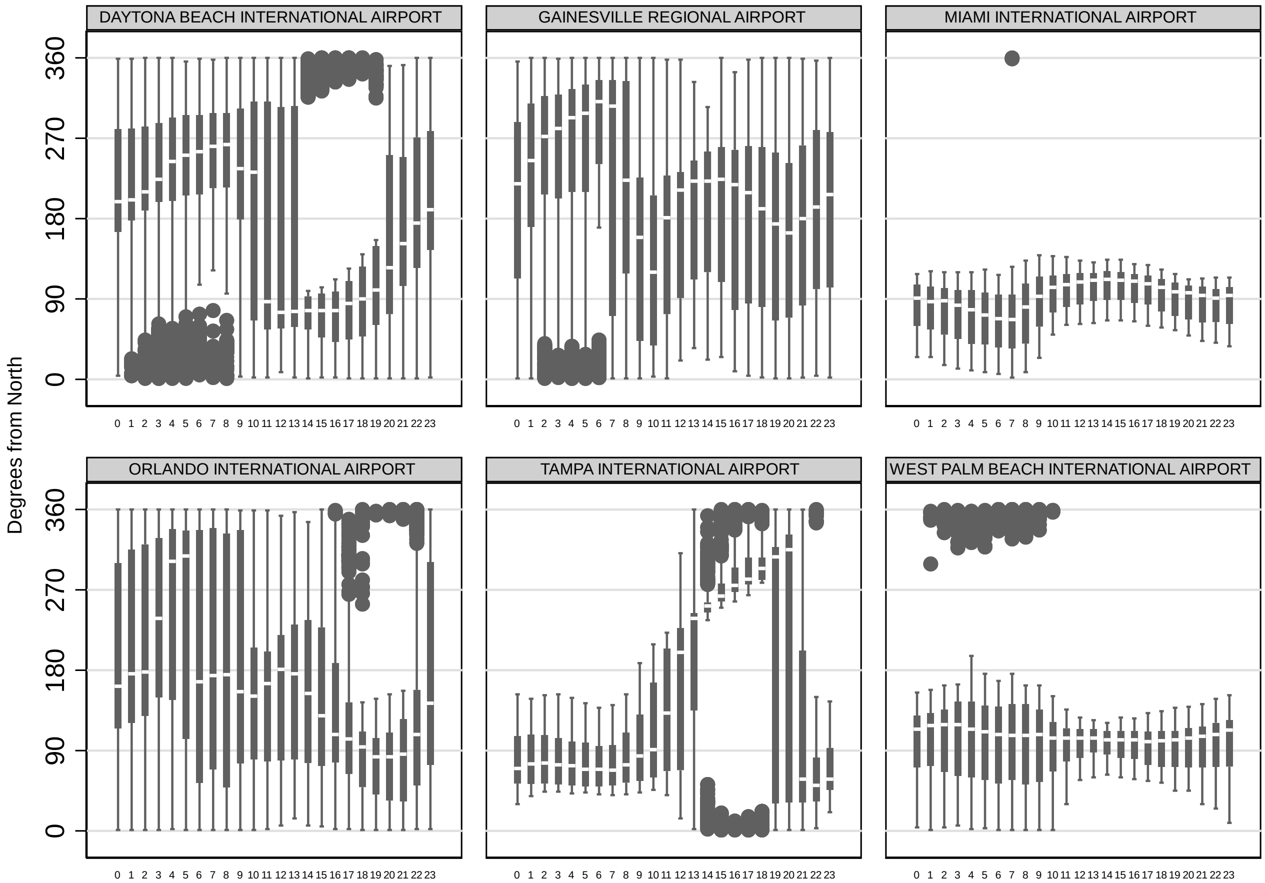
$^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$

Table A7: Annual versus testing week variation on average FCAT

	(1) Annual variation	(2) Test-week variation	(3) Annual and test-week variation
Annual	-0.0424*	NA	-0.0353
	(0.0214)		(0.0221)
Testing week	NA	-0.0181**	-0.0140*
		(0.0059)	(0.0064)
Observations	218,397	218,397	218,397
N Students	86,882	86,882	86,882

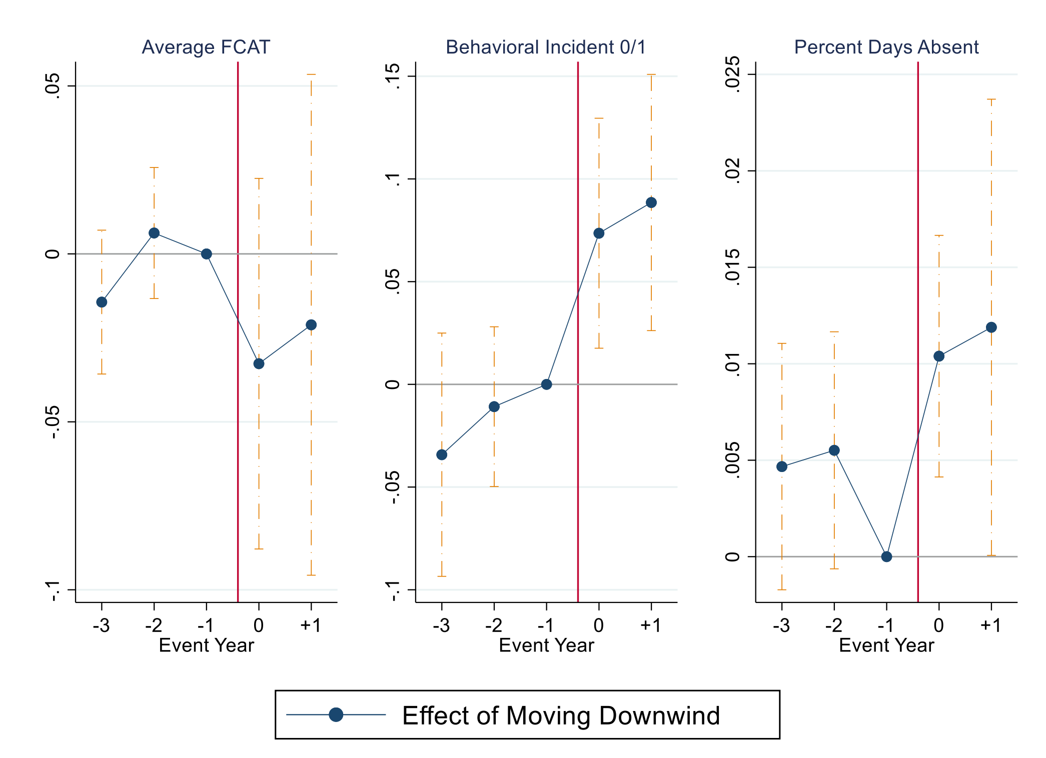
Notes: Each column shows results from a different regression where average FCAT is the dependent variable. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, a grade-moved indicator, the number of highways within a mile, distance from nearest highway dummies, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Standard errors clustered on zip and id are in parentheses. Column 1 runs the main specification on the subsample with wind data available during the testing week but defines downwind at the annual level. Column 2 defines downwind at the testing-week level. Column 3 includes both measures to test whether annual or testing week variation in pollution exposure more strongly predicts FCAT scores. All models are estimated using only the "policy-induced" moves to middle/high school by dropping "K through 12" schools and limiting the sample only to students who change schools from fifth to sixth or eighth to ninth grade.
$^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0$.

Figure A1. Wind direction across Florida



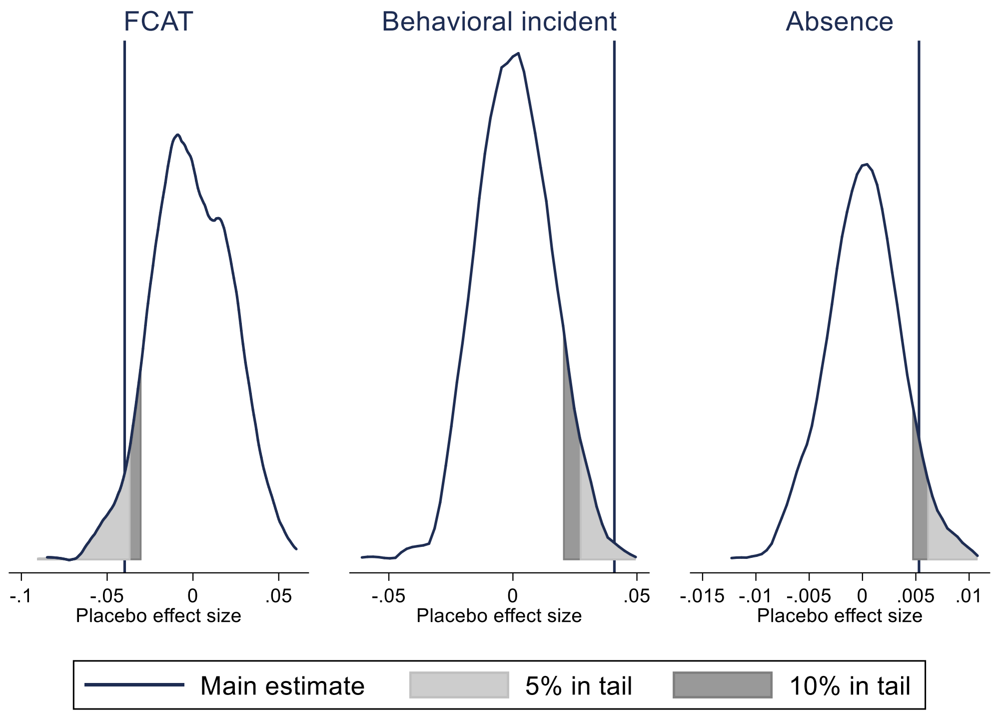
Notes: This figure depicts a series of box plots representing the distribution of wind directions throughout the day over a year for wind monitors in different geographic locations in Florida. The x-axes are hours in the day, and the y-axes are wind direction in degrees from north for the direction the wind is blowing at each hour. The gray dots represent outliers. This figure is constructed from NOAA (MADIS) wind data from 2010.

Figure A2: Event study for all students, relative to year of the move



Notes: This figure plots coefficients from an event study where the event is any move to a school that is downwind more than 60% of the time. The X-axis charts "event time" in grades relative to the move. The Y-axis plots coefficients of the effect of being in the given event year on the given outcome. All models include grade fixed effects, zip code fixed effects, student fixed effects, year fixed effects, distance from nearest highway dummies, a grade-moved indicator, the number of highways within a mile, school demographic characteristics, other school-level characteristics (percent of teachers with a master's degree, size, and stability rate), and whether the student was on FRL that year. Includes 95% confidence intervals based on standard errors clustered on zip and id. Sample includes all students in any grade as long as they attend school within 0.4 miles of a highway.

Figure A3: Placebo estimates



Notes: Distribution of results from 500 placebo tests per outcome. Our main estimates for our preferred specification are represented with a vertical line on the placebo effect size distribution. The lightly shaded gray region is the region of the graph where there is 5% in the tail of the distribution. The darker shaded gray region represents 10% in the tail of the distribution. For each placebo, schools were randomly assigned a percent of time downwind from the empirically observed distribution, which was then translated into an indicator variable equal to one if the placebo percent downwind was greater than 60%. The sample was limited to schools within 0.4 miles of a highway that were not actually treated. To ensure variation with zip and student fixed effects, placebos had to have at least 21 zips with variation and at least 46 treated schools.

Appendix 2: Online Data Appendix

FCAT scores are standardized by year and grade at the state level for each test, with a mean of 0 and a standard deviation of 1, and we average the math and reading scores by year to create one summary measure of academic performance. We decided to average reading and math to create an aggregate measure of student achievement that reduces noise and the likelihood of Type 1 error due to multiple hypothesis testing. That being said, we have also looked at these scores independently and find that they follow a similar pattern. We construct the rate of absences variable by dividing the number of days absent by the number of days in the school year.

The longitudinally-linked data follow students over time as long as they remain in the Florida public school system. The FDOE conducts the longitudinal matching process. About 90% of students are matched year-to-year by Social Security number, and any remaining students are matched by name and birthday. This matching process contains a small number of errors likely caused by multiple students with similar names or birthdays. To account for this, we ran a specification in which we exclude students who move backwards more than two grades, fail and then skip a grade, have a change in birthday, or change gender from year-to-year. In total, these deletions amount to about 5% of the original dataset. We lose few students in the longitudinal analysis; among students who took the third grade FCAT before 2009, we observe 90% taking an FCAT the following year and over 80% taking an FCAT five years later.

School Characteristics

School location data come from the 2010 National Center for Education Statistics data, which provides the latitude and longitude of every school in the United States. We use 2010 as a standard year due to slight variations in the reported locations of schools over time. Data on school characteristics come from the Florida School Indicators Report (FSIR), which is released annually by the Florida Department of Education. This data includes percent of teachers with a master's degree, school size, school stability rate, school racial demographics, and a variety of other school characteristics. Maternal education by school and percent of married mothers by school were calculated based on Florida vital statistics data that was aggregated at the school level.

Highway Data

We link schools to the nearest major highway using road data from the Florida Department of Transportation (FDOT), which maintains geo-coded files on the location and traffic density of large roads in Florida. We use the files from 2010, as there is little change in these roads from year to year. Each major highway is subdivided into smaller segments. The segment length varied from 74 feet to 29 miles, with a mean of 3.3 miles and a median of 2.0 miles. For matching roads to schools, we subdivide these road segments into a series of points that are a maximum of 0.1 miles apart. We match each school to its nearest five road segments using these points.

Of the schools with at least one major highway within 0.4 miles, the average distance to the nearest road is 0.21 miles. Fifty-three percent of those schools have only one major road within 0.4 miles, 24 percent have two, 10 percent have three, and the remaining 12 percent have four or more.

The FDOT calculates estimated average annual daily traffic (AADT) for each road segment. These counts are an annual measure that does not provide specific details on average traffic counts over the school year or school day. Likewise, there is not enough year-to-year variation in traffic to leverage within-road segment changes in annual average traffic volume over time. We instead use 2010 AADT to stratify our estimates by traffic volume, and we consider this to be a rough proxy of additional pollution exposure due to more heavily trafficked roads. For major highway segments, the AADT ranged from 1,150 cars per day to 306,000 cars per day. In our sample, the mean AADT of student's nearest road is 64,419 cars per day.

Wind Data

To determine whether a school is downwind of a major highway, we proxy prevailing school-day winds using 2010 data from the U.S. Meteorological Assimilation Data Ingest System (MADIS), a part of the National Oceanic and Atmospheric Administration (NOAA). The 2010 data is the most complete year available that is within the time frame of education data. Each MADIS station includes wind readings once per minute, and we take the first observation per hour.[1] There are 1,029 stations in the state of Florida, and we connect each school to its nearest station and assign it that station's hourly wind data. Across Florida, winds vary substantially both within a 24-hour period and across geographic location (see Appendix Figure A1).

For each hour of the school day, we define a school as downwind of a given highway if the wind direction blows within 45 degrees of a ray running from the nearest point on the highway to the school. That is, school j is downwind of the r^{th} nearest major highway segment in hour h, it would be considered downwind in that hour, as follows:

$$Downdwind_{jrh} = \begin{cases} |ray_{jr} - winddirection_{jrh}| < 45° = 1 \\ |ray_{jr} - winddirection_{jrh}| \geq 45° = 0 \end{cases}$$

Some schools are near multiple major highways. Taking r = 1–5 highway segments, we take the maximum of $Downdwind_{jrh}$ to obtain $Downdwind_{jh}$, a measure of whether a school is downwind of at least one of the nearest five major highway segments in a given hour. We then collapse the data over the year to obtain $Downdwind_{j}$, which provides the percent of time a school is downwind from any nearby major highway during the school day over the course of the year.

[1] We have also tried taking a random observation within the hour, and the results are the same.

We create a binary variable to delineate treated from non-treated schools, using a cutoff of 60% of time downwind.

In addition, we also show our results using a continuous measure of wind exposure in Table A4. We construct a variable that captures whether the average wind pattern blows directly at a school, directly away from a school, or in-between these extremes. Specifically, for each hour of observed wind data, we calculate the difference in degrees between bearing from the school to the highway and the direction the wind is blowing at a school: $Intensity_{j1h} = |ray_{j1h} - winddirection_{j1h}|$ so that it ranges from 0 to 180. We only calculate this from the nearest highway (r=1), as there is no obvious way to measure intensity across multiple nearby highways like we do with downwind status. Because we do not average across multiple highways when using wind intensity, we exclude parts of the road segment that go onto feeder routes in order to focus on only major U.S. highway and interstates, meaning our N is smaller in this analysis. After calculating intensity within an hour, we take the average over all hours ($Intensity_{j1h}$) and transform this to equal zero when the average degree is directly away from a school (completely untreated by the wind) and one when the wind is always blowing directly towards a school. This measure has the advantage of capturing more information about wind direction and variation in how directly the wind is blowing at the school as a measure of treatment intensity. We also create a "censored intensity" measure that assigns a zero to any schools that average being upwind of the nearest major highway, with the downwind schools ranging from near zero (for those where the wind blows close to parallel) to one (for schools where the wind blows directly from the road to the school). Across both measures, being "more" downwind is associated with worse outcomes for students.

Event Study Methodology

The event study regressions include all the same variables as in our primary specification (Equation 1). However, when estimating event studies, we need to take into account two specific issues that arise in our context. First, many students move multiple times such that there is not always one treatment occurrence per student. Second, treatment comes from both moving to upwind and moving to downwind; while these estimates are averaged together in our regression model, we need to estimate them separately to show distinct patterns in the event study. Students who never change treatment status are left in the reference group in all event study models.

The “move to middle school” event study design (Figure 3) resolves the first issue by only focusing on moves from fifth to sixth grade. Further, we only include those students who we observe continuously between third and seventh grade, dropping students who attended K–12 schools and who moved between an upwind and downwind school before sixth grade. Because of missing data, sample sizes become much smaller if we try to extend the event study beyond seventh grade. We resolve the second issue of there being upwind and downwind moves by estimating upwind and downwind moves as separate events (see Equation 3 in the text).

We simplify the “All mover” event study (Appendix Figure A2) by just looking at downwind moves. Here we define a qualifying event as a student who we observe for five years, and who we observe for at least three years before and two years after a move. We include all students with a qualifying event in the sample as well as students who never change downwind status. We drop students who changed to downwind status without meeting the standard for a “qualifying event” or who had more than one qualifying event (in practice, very few students had more than one qualifying event).

Annual versus Testing Week Pollution Exposure

Prior research finds that testing-day pollution exposure affects test performance (e.g., Ebenstein, Lavy, and Roth 2016; Marcotte 2017; Roth 2016). In this section, we use our detailed wind data to assess this pattern, as well as test whether prevailing school year or testing week pollution better explains achievement outcomes. Though the designation of downwind status is static for each school for the annual variation, we can use the changing timing of the testing dates to examine whether testing-week-specific effects also drive results (Heissel and Norris 2018). We do not show the same test on behavioral incidents or absence rate, as the test-week fluctuations should not change these annual measures.[2]

As in the main analysis, we use the 2010 data as a proxy for all years; that is, the annual exposure for a given school is the same across all years and variation comes from students moving across schools. We proxy testing week exposure using the 2010 wind data by creating a downwind measure based only on the specific two-week period of testing for each year. Here, variation comes from both student moves and changes in the testing time. Not every school has wind data available to create this testing-time-specific downwind measure, so the N in this analysis is smaller than the main results.

Appendix Table A7 displays the results. Column 1 repeats the main specification on the subsample limited to schools with wind data available during the testing week. Column 2 instead only includes a measure of downwind status at the testing-week level. Finally, Column 3 includes both measures to test whether annual or testing week variation in pollution exposure more strongly predicts FCAT scores.

[2] We have examined these measures, and we find that test-week wind variation does not affect these outcomes after conditioning on annual variation, as would be expected.

In column 1 we come close to replicating our main findings with students in schools downwind 60% or more of the time over the course of the year having around a 0.042 standard deviations lower scores, relative to when they are in upwind schools. Moving to Column 2, we find that attending a school that is downwind 60% or more during the testing week specifically leads to 0.018 standard deviations lower test scores, relative to testing weeks in other years when the student was not downwind. When we include both of these measures together, the coefficients are similar, though we find that the impact of annual variation falls slightly to be roughly only 2.5 times the size of the weekly variation, though the standard errors on both coefficients are now larger due to collinearity. This suggests that continual exposure while students are learning during the school year has an impact that is two to four times greater than exposure on the day of the test. We note, however, that attenuation bias may play a role in these different outcomes, as measurement error is likely greater at the weekly rather than annual level when we use 2010 as a proxy. Thus, we take these estimates to be suggestive at best.

Wind Direction and Traffic Pollution Model

To estimate the relationship between pollution and wind direction, we use high frequency 2010 MADIS and EPA data at the hourly level. We limit our sample to all monitors within 0.4 miles of a major highway (or 0.1 miles in some specifications). As with measuring downwind status, we only use wind and pollution data during the school day. We then estimate the following regression:

$$P_{imh} = \gamma_0 + \gamma_1 Downwind_{imdh} + \varphi_i + \sigma_m + \varepsilon_{imdh}$$

where P_{imdh} is a measure of pollution for monitor i in month m during day d and in hour h. $Downwind_{imdh,}$ is measured at the day-hour level. Similar to the main analysis, we use an indicator for the wind blowing within 45 degrees of a ray running from the nearest point on the highway to the pollution monitor. φ_i is a vector of monitor specific fixed effects. We control for month fixed effects (σ_m) in order to account for seasonal differences in pollution and wind direction. Within 0.4 miles, there are five PM10 monitors, three NO_2 monitors, and seven CO monitors. Within 0.1 miles, there are two PM_{10} monitors, two CO monitors, and one NO_2 monitors. We cluster all models at the monitor site-day level (we do not have enough monitors to cluster at the monitor level). The majority of wind monitors do not measure the same pollutant. One of the 15 total wind monitors measures both PM_{10} and NO_2, and another measures both PM_{10} and CO.