Takeoffs and Test Scores: The Impact of Noise Pollution on Standardized Test Performance

Joshua Aarons*

6/16/26

Abstract

Noise pollution has become so ubiquitous in urban life that it is generally considered innocuous. However, research suggests that noise pollution can be more than a minor distraction and is associated with diminished health and productivity. I leverage exogenous shifts in noise pollution from changes in aircraft noise and use a value-added model that accounts for student ability and school quality to estimate the causal impact of airplane noise pollution on standardized test performance. I find a large negative impact on math test performance, but no impact on English Language Art test performance. Moving from a school at the 25th percentile of exposure to one at the 75th percentile is associated with a math test score reduction of 0.10 standard deviations, twice as large as the impact of missing 10 school days during the year. Additionally, I find evidence that these negative impacts may be driven by the loudest days. These findings suggest a potentially large value to investing in noise insulation for schools exposed to high noise levels.

*Department of Economics, University of California San Diego; jaarons@ucsd.edu. I thank Judson Boomhower for excellent guidance and support during this project and Julian Betts and Joshua Graff Zivin for helpful advice and discussions. I thank seminar participants at UCSD and the 2024 AERE Summer Conference for their feedback. The research presented here uses confidential data from the Education Research and Data Center (ERDC) located within the Washington Office of Financial Management (OFM). ERDC's data system is a statewide longitudinal data system that includes de-identified data about people's preschool, educational, and workforce experiences. The views expressed here are those of the author and do not necessarily represent those of OFM or other data contributors. Any errors are attributable to the author.

1 Introduction

Noise pollution, unwanted sound from transportation, construction, and other activity, has become a feature of modern urban life. High noise levels, even those well below the level required for hearing damage, are linked to increased stress, poor sleep, and adverse cardiovascular health outcomes (Münzel et al., 2014; Saucy et al., 2020; Vienneau et al., 2022; Evans et al., 2001), resulting in a significant public health burden (WHO, 2011). Even at relatively low levels, noise has negative non-health effects in the form of reduced productivity (Noweir, 1984; Dean, 2024). These negative impacts are especially concerning given how pervasive noise pollution is in everyday life. In the United States, more than 45 million people live within 300 feet of major transportation infrastructure such as a large road (EPA, 2014) and one in eleven public K-12 schools is within 500 feet (Hopkins, 2017).

Despite the ubiquity of noise in modern life, there is only a nascent literature on the causal impacts of noise on cognition and learning. Recent work has found that increased noise impairs memory and reduces productivity (Hygge, 2003; Jahncke et al., 2011; Dean, 2024). Noise insulation on building has the potential to reduce impacts, but the prevalence of noise may lead individuals to underestimate its effects. Prior research suggests workers are unaware that increased noise reduces their productivity and therefore do not make valuable defensive investments (Dean, 2024). Therefore, documenting the effects of noise pollution is important to ensuring schools invest in the efficient amount of noise insulation and other defensive technologies that can mitigate negative impacts of noise pollution.

Cognitive psychologists have documented cross-sectional associations between noise and student performance and comprehension (Evans, Hygge and Bullinger, 1995; Evans and Maxwell, 1997; Haines et al., 2001; Stansfeld et al., 2005; Matheson et al., 2010; Klatte, Bergström and Lachmann, 2013).[1] However, these cross-sectional studies do not account for residential sorting. There is a well documented property value penalty associated with noise pollution such that houses in louder neighborhoods tend to have lower property values (Crowley, 1973; Nelson, 1980, 2004). Since most students attend local schools, students attending schools exposed to more noise pollution may be from lower income households than those attending schools in quieter areas. Given these results, researchers should worry estimates relying on cross-sectional variation in noise likely capture both direct impacts from noise pollution on cognition as well as indirect impacts from residential sorting. In this paper, I leverage changes in aircraft noise pollution at schools around

[1]There are a handful or experimental studies that find impacts of noise on memory tasks (Hygge, 2003; Hygge, Boman and Enmarker, 2003; Jahncke et al., 2011). These studies expose older, usually college-aged students to relatively low-level noise while reading and attempting to remember what they read. These studies unfortunately do not measure any impacts on standard measures of learning and do not attempt to determine whether noise impacts non-reading recall or skills. Potentially due to their often small sample sizes or participant motivation, results in this field have contradicted each other (van Kempen et al., 2010).

Seattle-Tacoma International Airport (SeaTac) to estimates the causal impact of noise pollution on standardized test performance.

Utilizing data from noise monitors around SeaTac, I measure the number of loud noise events from overhead airplanes near schools from 2008 through 2019. I implement a value-added model of student performance at schools near monitors to adjust for student ability and isolate the direct effect of noise pollution on performance. My approach exploits both across- and within-school variation in noise pollution over time that comes from a change in flight paths into and out of SeaTac as well as year to year variation in flight noise. My results show that noise pollution had a negative effect on math test performance but no impact on English language arts test performance.

Importantly, the effect size of the impact on math test scores is large. Under my preferred specification, I estimate that an additional loud flight overhead per hour during diminishes math test scores by 0.0037 standard deviations. In terms of noise levels, a one decibel increase in aircraft noise is associated with a 0.0073 standard deviation decrease in math test scores. These are large effect sizes. Moving from a school in my dataset at the 25^{th} percentile of exposure (Leq of 46.2 dB) to the 75^{th} (60.2 dB) in 2019 is associated with a 0.1 standard deviation reduction in math test scores, larger than the impact of missing 20 days of school (Liu, Lee and Gershenson, 2021).

I also consider a nonlinear models of the share of school days at different levels of flight noise. My results suggest that the negative effects of noise pollution on test performance are driven by increases in the number of very loud days at the schools exposed to the most aircraft noise. This result suggests that defensive investments might have large positive impacts on student performance for schools in the loudest areas and potentially no impact for schools exposed to lower levels of noise pollution.

My work primarily contributes to the literature on identifying causal impacts of noise pollution in the labor context. Dean (2024) uses a randomized control trial at a textile training facility in Kenya to document a negative impact of noise on cognitive task productivity. The study also elicits willingness to pay for quiet working conditions and finds that workers who are paid based on output are no more willing to pay for quiet conditions, suggesting workers are unaware that increased noise reduces their productivity. Hener (2022) uses random variation in which direction planes approach Frankfurt airport on a given day that comes from wind direction to estimate that higher daily background noise increases the violent crime rate. I use a change in landing and takeoff paths to capture the effect of an unanticipated shift in exposure to aircraft noise pollution. I am not the first to use this policy change; Kuminoff and Mandia (2025) and Allroggen et al. (2025) leverage the same shift to estimate the effect of noise pollution on the Phoenix and Seattle, Chicago, and Boston housing markets, respectively. Beyond my results, I add to this literature by showing that there is a large amount of year to year variation in flight noise around Seattle Tacoma International Airport before and after the flight path change that can be exploited to measure noise pollution impacts.

This paper also contributes to the broader and more robust literature measuring the causal impacts of pollution on education and human capital accumulation. It is well documented that both contemporaneous (Ebenstein, Lavy and Roth, 2016; Graff Zivin et al., 2020; Zhang, Chen and Zhang, 2018) and cumulative (Heissel, Persico and Simon, 2022; Wen and Burke, 2022; Stafford, 2015) air pollution exposure reduces exam scores. Beyond diminishing test scores, air pollution is tied to increased absence rates and behavioral incidents (Heissel, Persico and Simon, 2022). Cumulative lead pollution also diminishes student achievement (Hollingsworth et al., 2025) and increases absences and suspensions (Gazze, Persico and Spirovska, 2024). This paper extends the work on the adverse impacts of pollution exposure on educational outcomes to include noise pollution.

The remainder of this paper is organized as follows. Section 2 provides information on noise levels and the implementation of new flight paths at SeaTac. Section 3 discusses noise data and the construction of an analytic dataset. Section 4 explains the value-added model and strategy for estimating the impact of loud flights and flight noise on test performance. Section 5 presents results. Finally, Section 6 concludes.

2 Context: Performance-Based Navigation

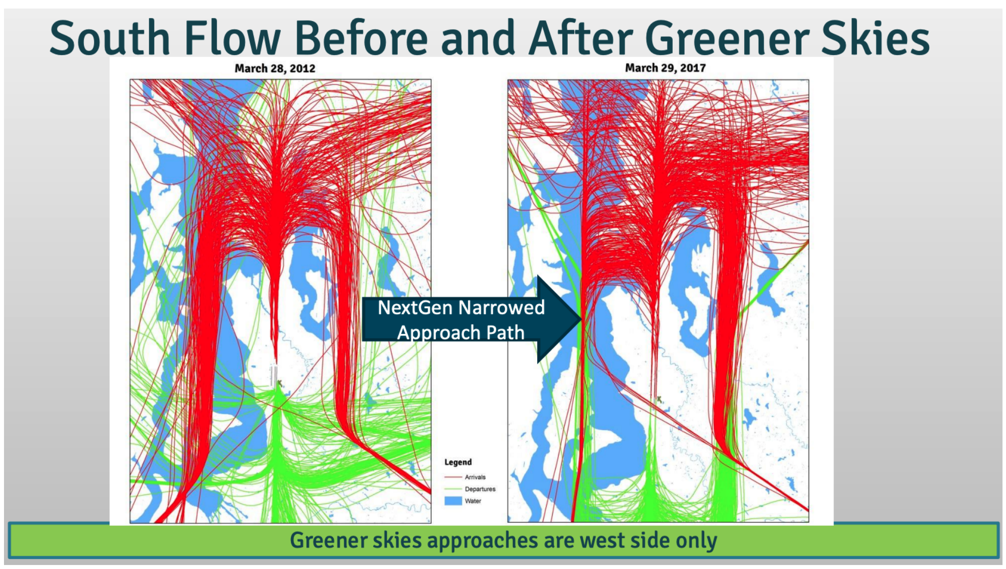
Over the past two decades, the Federal Aviation Administration (FAA) has made a series of changes to commercial flight paths and procedures. The backbone of this suite of programs commonly known as NextGen is an advancement in navigation technology. Starting in the early 2010s, the FAA began using Automatic Dependent Surveillance-Broadcast (ADS-B) technology to track planes and mandated all planes be equipped with ADS-B by 2020. ADS-B determines an airplane's position using GPS (satellites) rather than radar (radio waves), which allows for more precise location tracking. This advancement allowed the FAA to shift more planes to the most efficient landing and takeoff routes to reduce approach and takeoff flight distance.

Using ADS-B, the FAA implemented Performance-Based Navigation (PBN), a program that altered the routes pilots used for takeoff and landing, at the nation's busiest airports. The program sets limits on flight paths, giving air traffic controllers and pilots less autonomy in determining routes. With more precise location information from ADS-B, air traffic controllers can safely have planes fly in closer proximity rather than encouraging pilots to spread out over a wider area. This has concentrated more planes on the most efficient routes for takeoff and landing.

PBN has caused noise corridors to tighten and become louder as planes fly over the same optimal paths in quick succession, leading to pushback from members of the public. Airports and the FAA have seen dramatic increases in noise complaints.[2,3] When flight paths shift, there are both winners and losers in terms of noise pollution – some areas experience declines and others increases. This project does not attempt to measure the impact on total social surplus, but to instead estimate the effect on a subset of those who were affected by the path changes.

The FAA piloted new PBN flight paths on 1% of flights at SeaTac beginning in March 2013 (United States Department of Transportation, 2014). SeaTac expanded the program to all planes equipped with ADS-B technology in April 2015 (Alaska Airlines, 2015). As a result, flight paths used for landings and departures concentrated along more narrow corridors, which is particularly noticeable for southern departures depicted in green in Figure 1. PBN induced both an immediate shock and gradual change in flight paths because ADS-B technology was not required until 2020. This effect is visible in the number of loud noise events from aircraft each school year shown in Appendix Figure 1.

Figure 1: Change in Air Traffic Flow for Southern Departures and Northern Arrivals



3 Data and Treatment Measurement

3.1 Noise Data

The Port of Seattle operates a network of noise monitors around SeaTac, and 24 of their monitors covered 2008 through 2019. These monitors record sound pressure continuously and docu-

[2] *Advances in airport technology mean sleepless nights for some* and *New flight paths lead to airplane noise complaints across US*

[3] These complaints do not impact the way that air traffic controllers route flights day to day, though a lawsuit in Phoenix did lead to temporary reversion to the previous flight paths.

ment all time spans that correspond to an airplane flying overhead. The Port of Seattle quantifies the sound pressure from flights with a commonly used measure, sound exposure level (SEL). My data capture each instance a flight overhead results in the SEL exceeding 65 decibels (dB), approximately the same volume as using a vacuum cleaner or standing next to city traffic. The vast majority of the events captured by the monitors are not loud enough to cause hearing damage, but they are loud enough to startle, distract, or drown out talking (Appendix Figure 2 provides the distribution of loud noise events).

SEL events are helpful for counting the number of loud noise events, but not the amount of noise students hear. There are two common measures of aggregate noise: equivalent continuous sound level (L_{eq} or, more commonly, Leq) and day-night sound level (L_{dn} or, more commonly, DNL). Both measure the total noise pressure from loud noise events during a period of time rather than the frequency of loud noise. The decibel scale is logarithmic, so a 90 dB noise generates 100 times as much sound pressure as a 70 dB noise (even though the human ear only perceives the louder event as four times as loud as the quieter). This means that the loudest events largely determine the Leq and DNL. For example, consider one hour at two different noise monitors: at one there was one flight overhead with a SEL of 90 decibels (dB), at the other there were 100 flights overhead each with SELs of 70 dB. The Leq of these two monitor-hours is the same even though one indicates potentially 100 times as many interruptions. The difference between DNL and Leq is that DNL weights overnight hours ten times as heavily as daytime hours whereas Leq weights all hours equally. I use Leq rather than DNL since school is during the day and I am less interested in the effect of overnight flights than those during school hours.[4]

In this paper, I use aircraft-specific Leq and the number of SEL events as treatments. The Port of Seattle also collects Leq data and calculates a level for both community and aircraft noise, which allows me to isolate noise from aircrafts from community noise, which may be endogenous to economic conditions. The school year runs from the start of September to the start of June in Washington, and schools are allowed to administer standardized tests any time during the final 12 weeks of the school year. I do not have data on the date of exams and therefore cannot determine the amount of exposure during exams. I calculate the average number of SEL events per hour and aircraft Leq from September through the end of February to focus on the effect of cumulative exposure during learning. However, noise pollution during March, April, and May is highly correlated with noise pollution during the rest of the school year, so I cannot separate the impact of noise pollution during exams on recall from the impact of cumulative noise pollution on learning.

One difficulty with the SEL data is determining when data is missing. One quarter of school hours in my sample have no SEL events. This could reflect that there were no loud flights overhead

[4]There are relatively few loud noise events captured by these monitors overnight, so the DNL and Leq measure are actually very similar.

during that hour or that the monitor was not collecting data. The probability of not recording any SEL events is non-random both across monitors and days. For example, the share of monitor-days with no SEL events is lowest in December when there are more flights. To address occasions where data is missing across all stations, I drop all days where at least three quarters of stations do not record a single SEL event.[5] These days are concentrated in five school year months (November and December 2010 and January, February, and November 2015) which are missing data. Of the 115 September - February school days I classify as missing data, 108 are during these five months. Unfortunately, most monitors are missing all data for the first four months of 2015. Without data for the start of the year and April, when flight paths changed, I cannot tell whether flight volumes changed notably between the learning period and exam period in 2015. I choose to drop the 2014/15 school year from my data because if there is a large increase during the exam period in 2015 but I record a school year with no large change in cumulative exposure, then including the school year would increase measurement error. I conduct sensitivity analyses including the 2015 tests and show that this choice does not change my results.

To address cases where specific monitors are not collecting data, I count consecutive days with no SEL events. I drop all hours during spells of at least two weeks without an SEL event. This process removes 3.6% remaining monitor-hours. The two-week threshold trades off underestimating the number of SEL events by including hours when monitors aren't functioning with overestimating by excluding hours where there were no flights overhead for many days in a row. Changing the threshold by a week in either direction has a very small impact on the share of hours dropped and no impact on the results.

3.2 Student Data

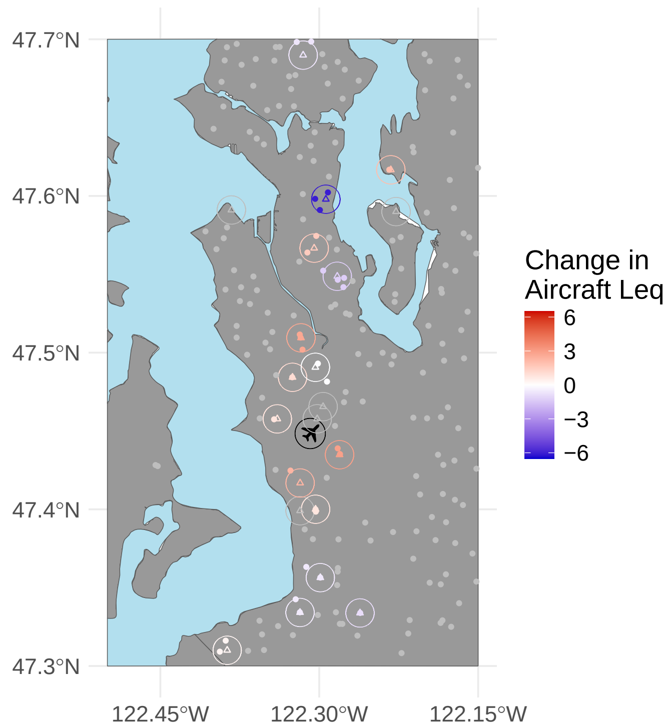
Washington state's Education Research and Data Center provided data on student enrollment, demographics, and test scores for this project (Education Research and Data Center , ERDC). My outcomes of interest are annual performance on mathematics and English Language Arts (ELA) standardized tests taken by all third through eighth grade students. I measure a student's performance on a test as number of standard deviations from the statewide mean that year. This accounts for year to year changes to the testing instrument that impact difficulty.

The Washington Office of Superintendent of Public Instruction (OSPI) provides data on school locations. I match monitors to all schools within one kilometer.[6] Figure 2 provides a visual representation of this process. Each monitor is depicted as a triangle at the center of a circle with a

[5]This is to remove the 5% of days with a high missing rate, which are usually days where all monitors or around 80% of monitors are missing data. Only 1.2% of days have missing rates between 50% and 75%, so the decision to set the threshold at 75% is not impactful.

[6]In Section 5.2, I show that my results are not sensitive to this choice.

Figure 2: Monitor and School Matching



one kilometer radius. The color of each monitor shows the difference in aircraft Leq between the post- and pre-PBN periods (the 2016-2019 and 2010-2014 school years, respectively). Monitors that experienced the largest increases in aircraft noise are the most red, while white monitors saw little change and blue monitors experienced less noise. Schools inside the 1 km radius around a monitor are matched and depicted using the monitor's color rather than the gray of the unmatched schools. My final dataset includes school years for all third through eighth grade public school students who were enrolled at a single school during the school year.

4 Empirical Design

4.1 Variation in Noise Pollution Exposure

In my setting, treatment is at the school level and there is no variation in treatment exposure within schools during a school year.[7] Variation in noise pollution comes from assignment to schools, year to year variation in flight path use, gradual increases in the number of flights in and out of SeaTac, and the shift in flight paths from the PBN implementation. If noise pollution

[7]This is a limitation of the data. It is possible that some students take classes in rooms that are more exposed than their peers, however I have no information on student classrooms or school acoustics.

were randomly assigned to schools in all years, one could simply regress test scores on noise exposure to recover a causal estimate of the effect of noise pollution. However, noise pollution is consistently higher at some schools than others, which could lead to residential sorting resulting in different types of students at more and less exposed schools. Therefore, the key challenge to identifying a causal impact is ensuring that the noise variation I isolate is not correlated with student characteristics that explain test scores.

The most common tool for assessing the impact of a one-time policy change is the difference in differences model. In this context, a difference in differences model with continuous treatment can compare the outcomes for students before and after the flight path shift and would attribute differences in these differences to the relative change in aircraft noise. This approach captures variation in noise from flight path changes both from an immediate shock and the average of gradual increases or decreases following the changes. Since planes were not required to be equipped to use the new flight paths until 2020 and the PBN implementation provided the opportunity for SeaTac to increase the number of annual flights, this gradual change is an important potential avenue for variation in exposure. The difference in differences model however does not leverage year to year variation in noise pollution or student assignment to schools.

To understand the relevant sources of variation I observe in my data, I plot year to year change in noise pollution against prior year annual noise pollution in Figure 3 (I show annual flight noise Leq by monitor in Appendix Figure 3). In the figure, each point represents a monitor-year, the color denotes the monitor, and the shape denotes the year relative to the PBN implementation. If the PBN implementation caused an immediate, permanent shift in flight noise at each monitor, then all points denoting other years (circles and squares) would be clustered tightly around the X-axis since there would be very little year to year change, and all points for the year after the shift (triangles) would show the total variation. If the implementation induced both an immediate shock and gradual shift, then pre-implementation years (circles) would be clustered around the X-axis with post-implementation years (triangles and squares) all on one side of the X-axis for each monitor. The figure instead shows a mix of shapes on each side of the X-axis for most monitors, suggesting that year to year variation is a key component of the noise pollution students experience.

Therefore, I utilize student data to take advantage of year to year variation in noise pollution while still accounting for potential sorting. For completeness, I also estimate a continuous treatment difference in differences model to show that a) my results are not driven by this choice, and b) the PBN implementation does not appear to have changed sorting patterns (at least in terms of observable student demographics). These results are available in Appendix Table 1 and Appendix Table 2, respectively.

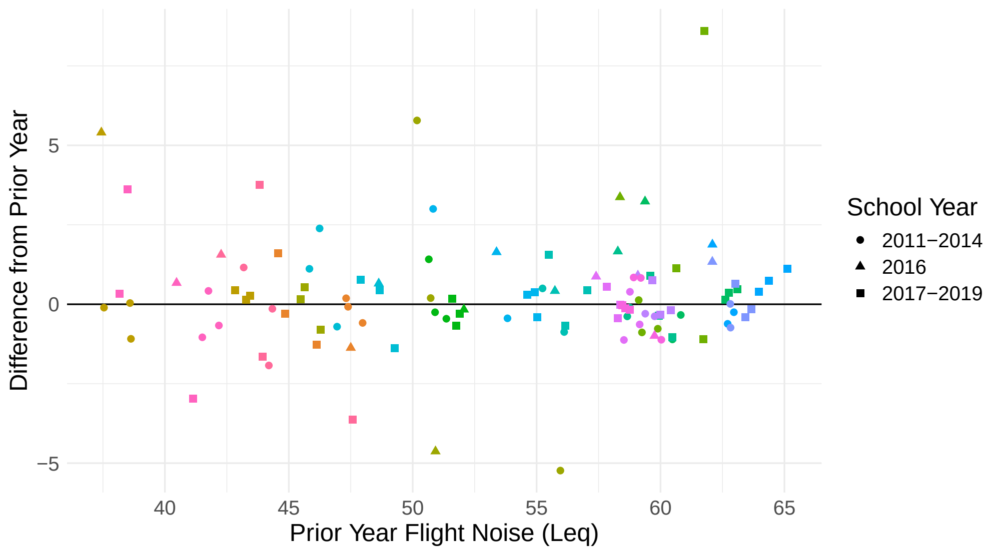


Figure 3: Annual Change in Flight Noise by Prior Year Flight Noise

4.2 Value-Added Model

I identify the impact of noise pollution on student performance by implementing a value-added model. The logic of the value-added model is that students perform similarly across years, so past test scores are a good estimator of future test scores. I model student i's standardized test performance as a function of the number of loud noise events at the student's school s during school year t, a vector of student characteristics, X_{it}, that includes student demographics and their test performance from the prior year, and idiosyncratic student-level variation, ε_{it}. To account for variation in testing instrument, I measure performance, y_{it}, as the number of standard deviations from the mean score for all Washington students in grade g during year t. Since I do not have teacher or classroom identifiers, I include school by grade fixed effects. I estimate the following equation using ordinary least-squares (OLS) regression:

$$y_{it} = \beta X_{it} + \gamma FlightNoise_{st} + \rho PM_{st} + \lambda_{sg} + \varepsilon_{it} \tag{1}$$

The parameter of interest is γ, which is the estimate of the causal impact of increased noise, either a 1 dB increase in aircraft Leq or an additional loud noise event per hour during a school year. The vector X_{it} includes the student's lagged test performance, gender, race/ethnicity, a flag for whether English is not their primary language spoken at home, and whether they receive free or reduced-price lunch during the school year. PM_{st} is the average daily $PM_{2.5}$ the school was exposed to during the school year.[8] λ_{sg} are time-invariant school-grade fixed effects that account for any

[8] After adjusting for the school and year averages, both measures of flight noise are uncorrelated with the average

differences in school composition or quality not captured by demographics. These include teacher quality and any unobserved differences in average students for example.

I use the value-added model to estimate the effect of year to year variation in noise pollution on student test performance. The identifying assumption is that noise pollution exposure is uncorrelated with student ability after controlling for the student's prior year exam performance, school and grade at that school, and student demographics. By controlling for student ability using prior year exam performance, this model and student-level data can mitigate concerns about a negative estimate being the result of sorting rather than noise impairing student learning or test taking.

4.3 Nonlinear Impacts

In the previous model, I use full school year Leq. An issue with aircraft-specific Leq is determining how to handle hours with no noise. The decibel scale is logarithmic and therefore accentuates differences at the low end. The difference in decibels between true silence, 0 dB, and a whisper, 30 dB, is nearly as large as the difference in flight noise at the most (70 dB) and least (37 dB) exposed monitors (in sound pressure – an exponential transformation of decibels – the latter difference is orders of magnitude larger). Taking the arithmetic mean of a day with no aircraft noise and a day that constantly had flights overhead constantly would average to an average aircraft Leq below the lowest total sound level the schools observed. Calculating the Leq for the full school year addresses the issue of zeros having a large impact on the average daily Leq, but also assumes a functional form for the impact of noise levels on student performance since relatively quiet days have little impact on the Leq. Therefore, I consider a model of nonlinear impacts of daily aircraft noise as well.

Specifically, I estimate a nonlinear model of the impact of the share of days at different levels of daily aircraft Leq. This takes the form:

$$y_{it} = \beta X_{it} + \sum_{j \in J} \gamma_j FlightLeq_{mt} + \rho PM_{st} + \lambda_{sg} + \varepsilon_{it} \tag{2}$$

where $FlightLeq_{mt}$ is the share of days in each decibel range bin. Ideally, I would create bins with relatively small Leq ranges to count the number of days at different aircraft noise levels and estimate the effect of having an additional school day with a specific flight noise Leq (relative to a day at the reference level of flight noise). One complication with this approach is that the support of the aircraft Leq distribution varies across noise monitors (see Appendix Figure 4). Therefore, it is not possible to estimate the causal effect of switching from a 45 dB aircraft noise day to a 70 dB

daily $PM_{2.5}$.

day.

I handle the lack of overlap of support by setting a relatively large reference bin to compare other days against so that all monitors have days in the reference bin. My specification has eight decile bins: <42 dB, 42-46, 46-50, 50-60, 60-62, 62-64 64-66, $\geq$ 66. Estimates of impacts of additional exposure at the upper end of the distribution come from comparing greater exposure years at louder schools to lower exposure years at those schools while estimates of the impacts of additional exposure at the lower end of the distribution come from comparing lower exposure years at lower exposure schools to higher exposure years at those schools.

4.4 Smarter Balanced Assessments Implementation

One complication with my context is that Washington switched from testing students using the Measurements of Students Progress (MSP) to the Smarter Balanced Assessment system (SBA) during the 2014/15 school year. This means that a large shift in noise pollution coincided with a change in test instrument. The left panel of Figure 4 shows that students were less likely to meet the exam standard after the switch to the SBA than they were before.

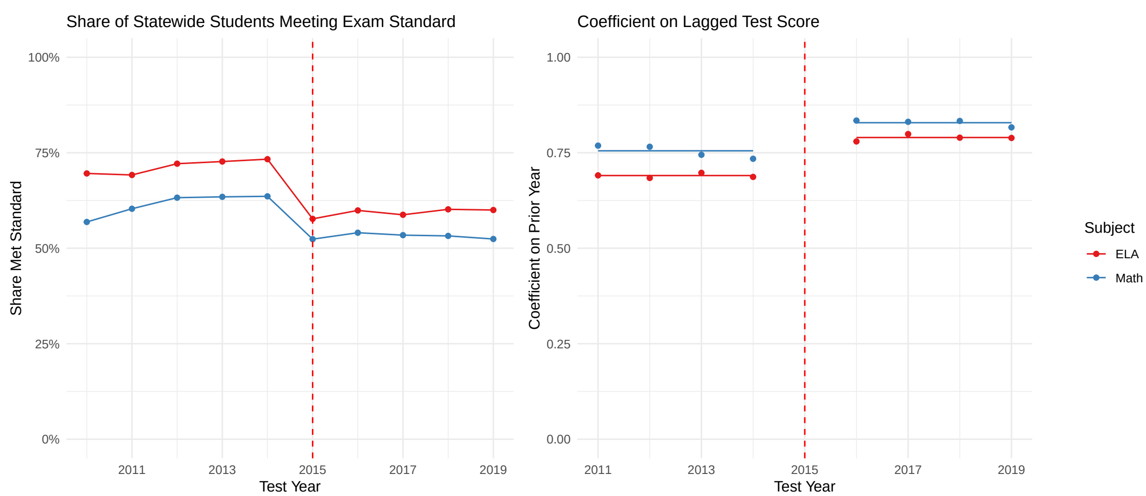


Figure 4: Impact of Test Regime Change

The regime change creates four problems for this analysis. First, students are less likely to meet the exam standard on the new exam and the tests have different scoring scales. I use standard deviations from the state mean score to scale everyone's performance so that this impact does not affect the results.[9] Second, prior to 2014/15, all third through eighth grade students took a reading exam and both fourth and seventh grade students also took a writing exam. The SBA assessments

[9]One exception would be if the change in test instrument had differential impacts across students of different demographics. I show in Appendix Figure 5 that there do not appear to be notable differences in impacts across demographic groups.

used from 2015 onwards are composed of just a single ELA exam rather than multiple tests. To minimize the impact of this change, I use the reading exam only prior to 2015 and the ELA exam from 2015 onward. Third, during the final year of the MSP regime, schools could choose whether or not to administer the old exam, resulting in a smaller sample that year.[10] The value-added model I use addresses concerns about selection into test taking by conditioning on prior year performance, however I also conduct sensitivity analysis around the inclusion of this year and find that it does not change my results.

The final concern pertains to the prior year test score variable in the value-added model. Lagged test scores serve as a measure of student ability, but if their predictive value varies from year to year, then residuals will be correlated with year. This is particularly concerning since the testing regime and flight paths change during the same year. The right panel of Figure 4 shows that the SBA exams better predict subsequent exam performance than the MSP exams do. In this figure, each point is the coefficient on the previous year's exam performance from a regression of performance on prior year performance interacted with test year. The horizontal lines are the coefficients from a separate regression where prior year performance is interacted with assessment regime instead of year. The annual and assessment regime coefficients show that each within regime, each year's testing instrument is similarly predictive of future performance.

In my model, I interact prior year performance with the assessment regime to account for the difference in predictiveness of testing instruments. This solves the problem of using a single variable for lagged performance, which would overestimate the prediction value of the prior year performance variable under the MSP regime and underestimate under the SBA regime, potentially leading to more residual variation to explain during the years that noise pollution levels grow and diverge. One might have additional concerns that the more predictive previous year performance is, the less residual variation there is for noise pollution to explain. Since the SBA instrument used when noise pollution is higher across my sample is more predictive than the MSP instrument, this would bias my results towards zero leading me to underestimate the true effect of noise pollution.

5 Results

5.1 OLS Results

I present my primary results from estimating the value-added model using OLS in Table 1. The table provides estimates for the impact of an additional loud flight per hour and of increasing the amount of flight noise during the school year by 1 dB on standardized test performance measured

[10]Schools did not have the option to implement the new exam, just the choice on whether or not to implement any exam.

in standard deviations (SDs) above or below the annual mean test score. My estimates suggest a negative impact of additional loud noise events on math test scores, but not on ELA test scores. I estimate that an additional loud noise event per hour reduces standardized math test scores by 0.0037 SDs. Using the same model, I estimate that increasing school year flight noise by 1 dB is reduces math test scores 0.0073 SDs. These estimates are smaller in magnitude than the difference in differences estimates available in Appendix Table 1, though the difference in differences estimate of the impact of SEL events is not significant at the $p < 0.05$ level.

Table 1: Impact of One Additional Loud Noise Event per School Hour on Exam Performance

	Aircraft Leq		SEL events	
	Math	ELA	Math	ELA
Estimate	-0.0073**	-0.0001	-0.0037***	0.0002
	(0.0035)	(0.0025)	(0.0009)	(0.0007)
Observations	63,379	63,357	63,379	63,357
R^2	0.74	0.68	0.74	0.68

*/**/*** denote significance at the $p < 0.1/0.05/0.01$ level. Test performance is measured in standard deviations from the mean score.
Standard errors are clustered at the school level.

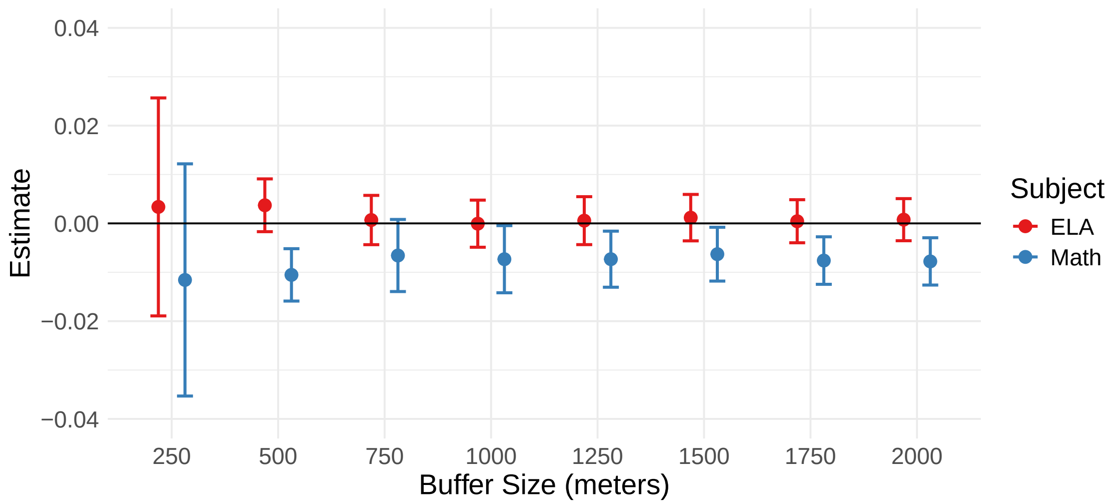
The magnitude of my math estimates is fairly large. Heissel, Persico and Simon (2022) estimate that attending a school just downwind of a highway instead of one just upwind of a highway reduces standardized test scores by 0.04 SDs due to the large increase in exposure to air pollution.[11] Moving from a school in my dataset at the 25$^{\text{th}}$ percentile of exposure during the 2019 school year (46.2 dB) to the 75$^{\text{th}}$ (60.2 dB) is associated with a 0.10 SD reduction in score, about 20% larger than the impact of missing 20 school days during the year (Liu, Lee and Gershenson, 2021).[12]

The difference in my estimates of the impacts of noise on math and ELA is consistent with studies on the importance of working memory for cognitive task completion. Working memory is a key input for mathematical problem solving (Lin and Powell, 2022) and distractions impede math performance (Vallée-Tourangeau, Sirota and Vallée-Tourangeau, 2016). Further, the effects of working memory on math and language tasks are not equal. Problem solving draws on domain-specific working memory as well as general working memory. Visual-spatial memory loads draw more additional effort from general working memory, and visual-spatial tasks have greater atten-

[11] Based on their first stage, the effect of an additional microgram per cubic meter of $PM_{2.5}$ is 0.02 SDs, which is similar to the 0.016 that Gilraine and Zheng (2024) find.

[12] Goodman (2014) uses moderate snowfall as an instrument for absences and estimates a much larger impact of absences on math test scores (0.05 SDs per absence) than do Liu, Lee and Gershenson (2021). However, he notes that if teacher absences increase during moderate snowfall days, then this estimate captures the effect of teacher absences as well.

Figure 5: Robustness to Alternate Noise Monitor Radius Choices



Bars indicate the 95% confidence interval around each estimate.

tion demands than verbal ones (Zehra Emine Ünal and Cowan, 2022). Visual-spatial working memory has a larger impact on visual-spatial task performance than verbal working memory does on verbal task performance (Giofrè, Donolato and Mammarella, 2018). The relatively larger taxation of math tasks on working memory and importance of math working memory suggest that distractions during exams have the potential to tax math performance more than verbal performance.

5.2 Sensitivity Analysis

I conduct several sensitivity analyses to determine whether my findings are robust to alternate specifications. First, I consider whether my choices of matching radius around noise monitors or minimum decibel level are driving my results. Expanding the radius around noise monitors increases the sample size, but reduces the accuracy of the noise pollution variable, potentially inducing more measurement error. Figure 5 shows that for almost all buffer distances considered from 250 to 2000 meters the estimate of the impact of noise pollution on math test scores is significant, and the magnitude is consistent.

I also consider whether my choice to exclude the 2015 test year and include the 2014 test year drives my results. I excluded the 2015 test year from my preferred specification as the flight path change happens in April of 2015 (and therefore may impact the exam period) and I am missing data for January through April 2015. I run a regression including the 2015 tests where the measure of noise pollution is the average number of SEL events per school day during September through December 2014 for the observations that year to account for the missing January to April data. I also consider excluding the 2014 test year as schools had flexibility in choosing whether or not to administer the exam that year. I compare my primary specification to these two specifications in Table 2. The results are similar across all specifications.

Table 2: Robustness to Alternate Year Specifications

	Math			ELA		
	(1)	(2)	(3)	(4)	(5)	(6)
Aircraft Leq	-0.0073**	-0.0066*	-0.0087**	-0.0001	-0.0018	0.0003
	(0.0035)	(0.0034)	(0.0034)	(0.0025)	(0.0019)	(0.0030)
Includes 2014	X	X		X	X	
Includes 2015		X			X	
Observations	63,379	68,227	58,672	63,357	68,176	58,649
R^2	0.74	0.74	0.74	0.68	0.68	0.69

*/**/*** denote significance at the $p < 0.1/0.05/0.01$ level.
Preferred model excludes 2015 test year due to change in flight paths in April 2015.

5.3 Nonlinear Results

The results from my preferred nonlinear model are shown in Figure 6. The effect of additional loud days is consistently negative and generally significant for math. Additionally, the estimate for an additional day in the loudest flight noise bin is twice as large as any other estimate, meaning the negative OLS result is driven by the loudest days. This suggests the annual Leq measure accurately represents the effect of cumulative flight noise on performance.

The lack of coverage of exposure to days with different amounts of flight noise limits both the precision and interpretation of my model. There are more than 180 schools within 20 kilometers (12 miles) of SeaTac, but I can only match one in six to a noise monitor, meaning many treatments and responses remain unmeasured, inhibiting my ability to estimate a dose response curve. Estimates of noise that map noise dispersion from flight path data, such as those used by Allroggen et al. (2025), could allow for more coverage of student treatment levels and prove more fruitful than monitor data for nonparametric estimates and enable researchers to estimate a dose response curve.

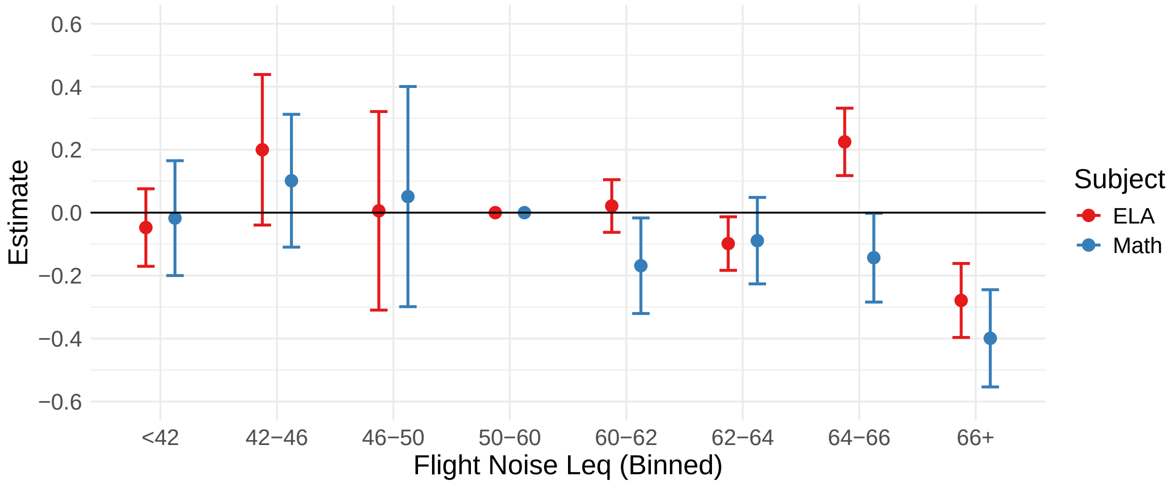


Figure 6: Estimate of Increased Share of Days in Flight Noise Leq Bin on Annual Standardized Test Performance

Bars indicate the 95% confidence interval around each estimate.

6 Conclusion

This is the first study, to my knowledge, to provide causal estimates of the impact of noise pollution on standardized test performance. I use exogenous changes in aircraft noise to assess the impact of noise pollution on student performance. I find a meaningfully large impact of loud noises during the school day on math test scores but not on ELA test scores. Further, my estimates suggest that the magnitude of these results is large. Moving from a school in my dataset at the 25$^{\text{th}}$ percentile of exposure during the 2019 school year (46.2 dB) to the 75$^{\text{th}}$ (60.2 dB) is associated with a 0.10 SD reduction in score, larger than the impact of missing 20 days of school (Liu, Lee and Gershenson, 2021).

A limitation of my work is that I cannot disentangle what share of the effect is from cumulative impacts on reduced learning and what share is from contemporaneous impacts on exam days. Washington allows schools to set their own testing dates and times (within a range of a couple months), and the school districts I contacted have not retained this information for the period of my study. As a result, I cannot determine the amount of pollution during test times and would need to use the average amount during the testing period instead. Average noise during the testing period is highly correlated with noise during the rest of the school year. Therefore, I cannot separate the effects on learning and encoding information from the effects on test taking and information recall. Future research should address this issue by leveraging a setting where there is variation between cumulative noise pollution exposure and exposure during the test. One possibility would be to take the approach of Ebenstein, Lavy and Roth (2016) and measure day-of-test exposure controlling for prior exposure in a setting where there is sufficient daily variation. Many airports change flight paths depending on the direction of the wind (Hener, 2022) or when runways are under construction or repair, so this should be possible in an airplane noise pollution context.

My finding of a negative impact of noise pollution on academic performance suggests two immediate steps for reducing the impacts of noise pollution. First, in the school context, sound insulation may mitigate damages at schools exposed to considerable amounts of flight noise. SeaTac and the Port of Seattle have been proactive about reducing noise pollution – their noise pollution monitoring system is more thorough and has been around longer than similar networks at other major airports, and they've focused on methods to reduce noise pollution.

Additionally, the Port of Seattle and FAA have spent hundreds of millions of dollars funding sound insulation projects for houses to mitigate pollution, and recently allocated \$70 million to insulate nine schools (with \$30 million set aside for six more).[13] This funding was focused on the Highline School District, which is closer to the airport than many of the schools treated with an increase in loud noise events following the PBN implementation. This suggests a role for expanding the funding to schools further from the airport, but underneath flight paths. Future research should investigate noise insulation as a mitigating factor on the impact of noise pollution on cognition and and learning. However, it is worth noting that my estimates are local average treatment effects that are identified off noise pollution variation at schools exposed to a considerable amount of flight noise. Therefore, I provide no evidence that noise insulation could provide meaningful benefits at low levels of flight noise.[14]

In line with the previous recommendation, most airports do not currently have a robust noise monitoring system. The areas identified as affected by airport noise are based on 65 dB DNL measures and are therefore generally very close to airports. My findings suggest large impacts to schools further away (a cluster of schools that were impacted most negatively by the flight path changes was 8-10 miles – 13-16 kilometers – away from the airport). This suggests sound proofing and other assistance is focused too narrowly and should be expanded to loud areas further from the airport. Ensuring adequate data collection to measure impacts to communities further from the airport would allow for monitoring of the impacts of noise pollution for schools and individuals (and better understanding of the dose response curve for researchers). Moreover, if airports or the FAA see benefits from adjusting flight paths, they should consider offsetting or mitigating damages to negatively affected communities through noise insulation funding or another form of transfer payment.

[13] The FAA funds insulation for houses exposed to excessive noise pollution from air travel, and the Port of Seattle has used airport revenues to fund additional sound insulation as well: https://www.westsideseattle.com/highline-times/2020/02/26/port-commission-approves-acceleration-noise-insulation-efforts-benefit.

[14] For example, moving from 10 to 11 dB of flight noise per year is unlikely to have any detectable impact on learning given the low amount of cumulative exposure.

References

Alaska Airlines. 2015. ”Report Shows Benefits of Greener Skies Approaches at SeaTac Airport.” https://news.alaskaair.com/newsroom/report-shows-benefits-of-greener-skies-approaches-at-sea-tac-airport/.

Allroggen, Florian, R. John Hansman, Christopher R Knittel, Jing Li, Xibo Wan, and Juju Wang. 2025. “Planes Overhead: How Airplane Noise Impacts Home Values.” National Bureau of Economic Research Working Paper 34431.

Crowley, Ronald W. 1973. “A Case Study of the Effects of an Airport on Land Values.” *Journal of Transport Economics and Policy*, 7(2): 144–152.

Dean, Joshua T. 2024. “Noise, Cognitive Function, and Worker Productivity.” *American Economic Journal: Applied Economics*, 16(4): 322–60.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. “The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution.” *American Economic Journal: Applied Economics*, 8(4): 36–65.

Education Research and Data Center (ERDC). 2025. P20W Data Warehouse Comprehensive Education and Data Research System, Core Student Record System, and Annual Statewide Assessment. Washington Office of Financial Management, Forecasting and Research Division. Data obtained on 02/11/2025 via ERDC’s data request process, https://erdc.wa.gov/dataresources/data-request-submission.

EPA. 2014. “Near Roadway Air Pollution and Health: Frequently Asked Questions.” Environmental Protection Agency.

Evans, Gary W., and Lorraine Maxwell. 1997. “Chronic Noise Exposure and Reading Deficits: The Mediating Effects of Language Acquisition.” *Environment and Behavior*, 29(5): 638–656.

Evans, Gary W., Peter Lercher, Markus Meis, Hartmut Ising, and Walter W. Kofler. 2001. “Community noise exposure and stress in children.” *The Journal of the Acoustical Society of America*, 109(3): 1023–1027.

Evans, Gary W., Staffan Hygge, and Monika Bullinger. 1995. “Chronic Noise and Psychological Stress.” *Psychological Science*, 6(6): 333–338.

Gazze, Ludovica, Claudia Persico, and Sandra Spirovska. 2024. “The Long-Run Spillover Effects of Pollution: How Exposure to Lead Affects Everyone in the Classroom.” *Journal of Labor Economics*, 42(2): 357–394.

Gilraine, Michael, and Angela Zheng. 2024. “JUE insight: Air pollution and student performance in the U.S.” *Journal of Urban Economics*, 143: 103686.

Giofrè, David, Enrica Donolato, and Irene C. Mammarella. 2018. “The differential role of verbal and visuospatial working memory in mathematics and reading.” *Trends in Neuroscience and Education*, 12: 1–6.

Goodman, Joshua. 2014. “Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time.” National Bureau of Economic Research Working Paper 20221.

Graff Zivin, Joshua, Tong Liu, Yingquan Song, Qu Tang, and Peng Zhang. 2020. “The unintended impacts of agricultural fires: Human capital in China.” *Journal of Development Economics*, 147(C): S0304387820301358.

Haines, MM, SA Stansfeld, RF Job, B Berglund, and J Head. 2001. “Chronic aircraft noise exposure, stress responses, mental health and cognitive performance in school children.” *Psychological Medicine*, 31: 265–277.

Heissel, Jennifer A., Claudia Persico, and David Simon. 2022. “Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance.” *Journal of Human Resources*, 57(3): 747–776.

Hener, Timo. 2022. “Noise pollution and violent crime.” *Journal of Public Economics*, 215: 104748.

Hollingsworth, Alex, Jiafang Mike Huang, Ivan Rudik, and Nicholas J. Sanders. 2025. “A Thousand Cuts.” *Journal of Human Resources*, 60(3): 950–976.

Hopkins, Jamie Smith. 2017. “The invisible hazard afflicting thousands of schools.” the Center for Public Integrity.

Hygge, Staffan. 2003. “Classroom experiments on the effects of different noise sources and sound levels on long-term recall and recognition in children.” *Applied Cognitive Psychology*, 17(8): 895–914.

Hygge, Staffan, Eva Boman, and Ingela Enmarker. 2003. “The effects of road traffic noise and meaningful irrelevant speech on different memory systems.” *Scandinavian Journal of Psychology*, 44(1): 13–21.

Jahncke, Helena, Staffan Hygge, Niklas Halin, Anne Marie Green, and Kenth Dimberg. 2011. “Open-plan office noise: Cognitive performance and restoration.” *Journal of Environmental Psychology*, 31(4): 373–382.

Klatte, Maria, Kirstin Bergström, and Thomas Lachmann. 2013. "Does noise affect learning? A short review on noise effects on cognitive performance in children." *Frontiers in Psychology*, 4: 578.

Kuminoff, Nicolai V., and Jasdeep Mandia. 2025. "Valuing Noise Pollution: Evidence from Flight Path Changes."

Lin, Xin, and Sarah R. Powell. 2022. "The Roles of Initial Mathematics, Reading, and Cognitive Skills in Subsequent Mathematics Performance: A Meta-Analytic Structural Equation Modeling Approach." *Review of Educational Research*, 92(2): 288–325.

Liu, Jing, Monica Lee, and Seth Gershenson. 2021. "The short- and long-run impacts of secondary school absences." *Journal of Public Economics*, 199: 104441.

Matheson, Mark, Charlotte Clark, Rocio Martin, Elice van Kempen, Mary Haines, Isabel Lopez Barrio, Staffan Hygge, and Stephen Stansfeld. 2010. "The effects of road traffic and aircraft noise exposure on children's episodic memory: the RANCH project." *Noise Health*, 12: 244–254.

Münzel, Thomas, Tommaso Gori, Wolfgang Babisch, and Mathias Basner. 2014. "Cardiovascular effects of environmental noise exposure." *European Heart Journal*, 35(13): 829–836.

Nelson, Jon P. 1980. "Airports and Property Values: A Survey of Recent Evidence." *Journal of Transport Economics and Policy*, 14(1): 37–52.

Nelson, Jon P. 2004. "Meta-Analysis of Airport Noise and Hedonic Property Values: Problems and Prospects." *Journal of Transport Economics and Policy*, 38(1): 1–27.

Noweir, Madbuli H. 1984. "Noise exposure as related to productivity, disciplinary actions, absenteeism, and accidents among textile workers." *Journal of Safety Research*, 15(4): 163–174.

Saucy, Apolline, Beat Schäffer, Louise Tangermann, Danielle Vienneau, Jean-Marc Wunderli, and Martin Röösli. 2020. "Does night-time aircraft noise trigger mortality? A case-crossover study on 24,886 cardiovascular deaths." *European Heart Journal*, 42(8): 835–843.

Stafford, Tess M. 2015. "Indoor air quality and academic performance." *Journal of Environmental Economics and Management*, 70: 34–50.

Stansfeld, SA, B Berglund, C Clark, I Lopez-Barrio, P Fischer, E Öhrström, MM Haines, J Head, S Hygge, I van Kamp, and BF Berry. 2005. "Aircraft and road traffic noise and children's cognition and health: a cross-national study." *The Lancet*, 365(9475): 1942–1949.

United States Department of Transportation. 2014. ”FAA Faces Significant Obstacles in Advancing the Implementation and Use of Performance-Based Navigation Procedures Federal Aviation Administration.” Report No. AV-2014-057.

Vallée-Tourangeau, Frédéric, Miroslav Sirota, and Gaëlle Vallée-Tourangeau. 2016. “Interactivity mitigates the impact of working memory depletion on mental arithmetic performance.” *Cognitive Research: Principles and Implications*, 1(1): 26.

van Kempen, Elise, Irene van Kamp, Erik Lebret, Jan Lammers, Harry Emmen, and Stephen Stansfeld. 2010. “Neurobehavioral effects of transportation noise in primary schoolchildren: a cross-sectional study.” *Environmental Health*, 9: 25.

Vienneau, Danielle, Apolline Saucy, Beat Schäffer, Benjamin Flückiger, Louise Tangermann, Massimo Stafoggia, Jean Marc Wunderli, and Martin Röösli. 2022. “Transportation noise exposure and cardiovascular mortality: 15-years of follow-up in a nationwide prospective cohort in Switzerland.” *Environment International*, 158: 106974.

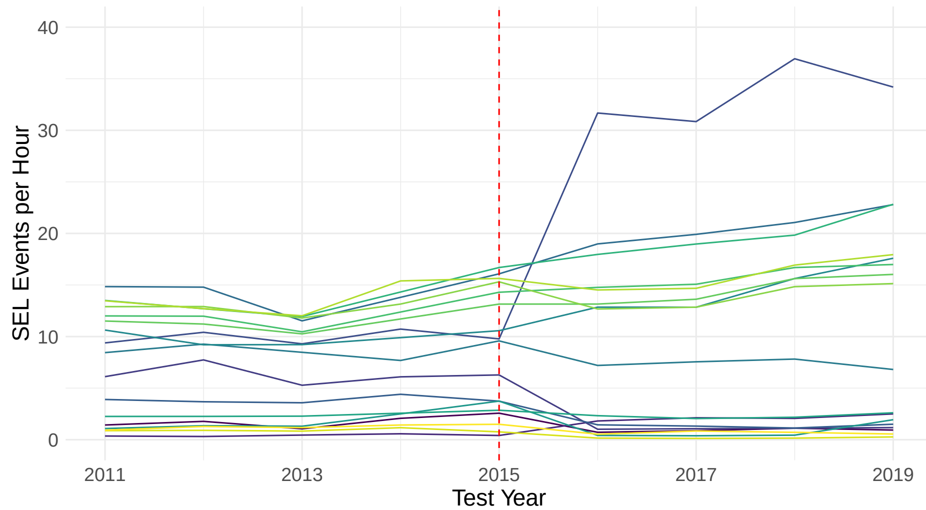
Wen, Jeff, and Marshall Burke. 2022. “Lower Test Scores from Wildfire Smoke Exposure.” *Nature Sustainability*, 66–95.

WHO. 2011. “Burden of disease from environmental noise: quantification of healthy life years lost in Europe.” World Health Organization. Regional Office for Europe.

Zehra Emine Ünal, Alicia Forsberg, David C. Geary, and Nelson Cowan. 2022. “The Role of Domain-General Attention and Domain-Specific Processing in Working Memory in Algebraic Performance: An Experimental Approach.” *American Psychological Association*, 46(3): 348–374.

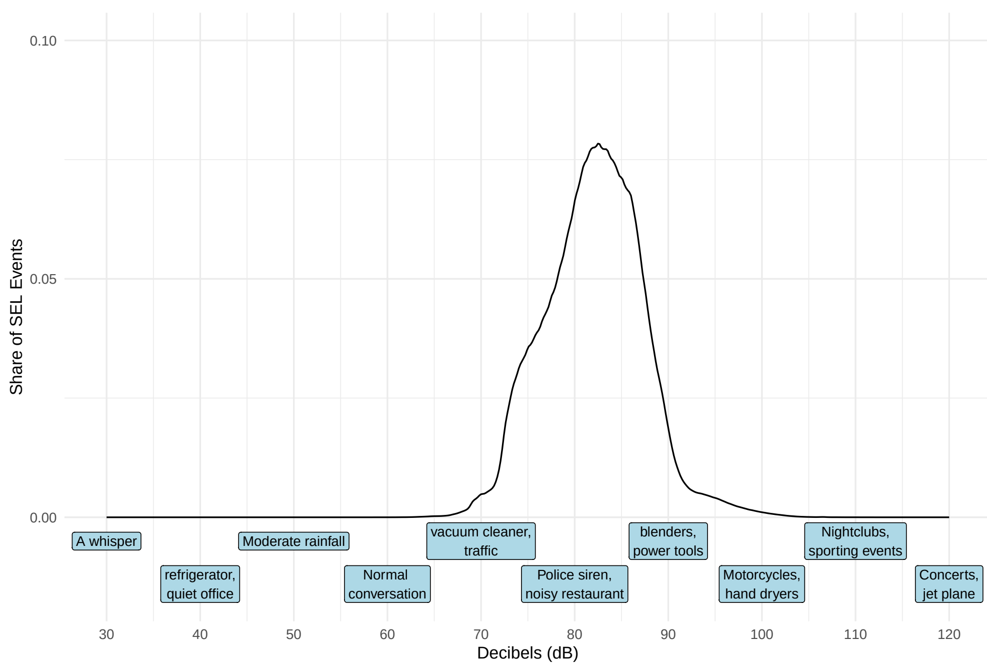
Zhang, Xin, Xi Chen, and Xiaobo Zhang. 2018. “The impact of exposure to air pollution on cognitive performance.” *Proceedings of the National Academy of Sciences*, 115(37): 9193–9197.

Appendix Figure 1: Average Number of Loud Noise Events per Hour by Noise Monitor and School Year



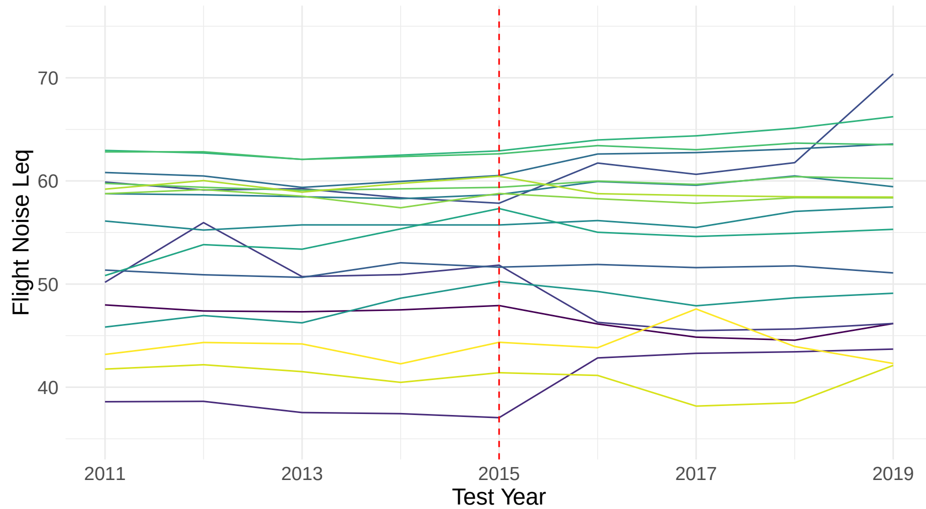
Notes: Each line depicts the annual average number of SEL events per hour for a noise monitor.

Appendix Figure 2: Distribution of Sound Exposure Level Events Recorded by The Port of Seattle's Monitor System



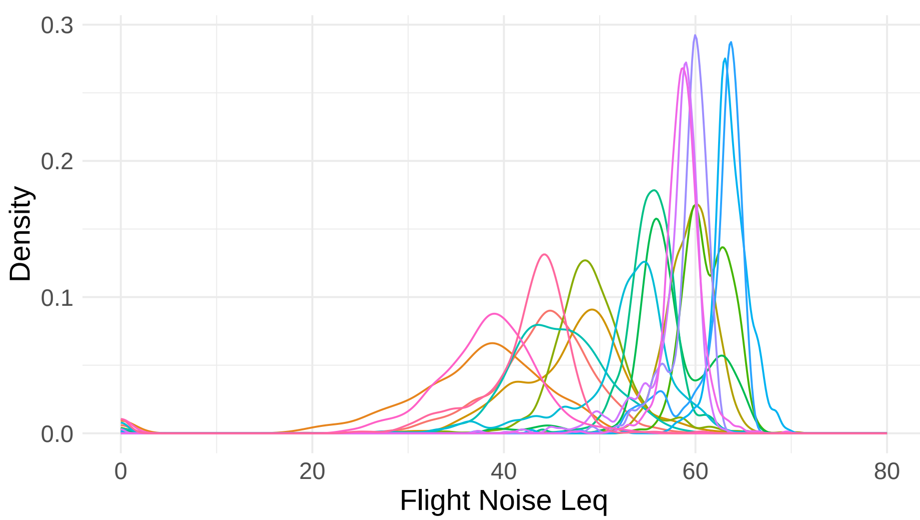
Notes: Monitor sample is limited to those that operated from 2009 through 2019. Density is for all sound exposure level events recorded by these monitors from 2009 through 2019.
Decibel level examples come from MDHearing.

Appendix Figure 3: Flight Noise Leq by Noise Monitor and School Year



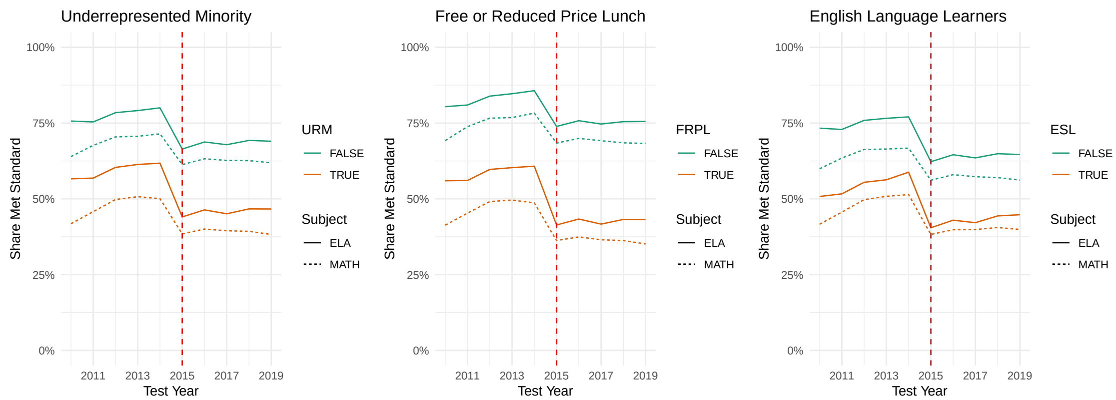
Notes: Each line depicts the annual flight noise-specific Leq during the school year for a noise monitor.

Appendix Figure 4: Density Plot of Daily Flight Noise Leq



Notes: Each line depicts the share of days at flight noise-specific Leqs for a noise monitor.

Appendix Figure 5: Impact of Test Regime Change by Demographic Group



Appendix Table 1: Difference in Differences Estimates of Impact on Test Performance

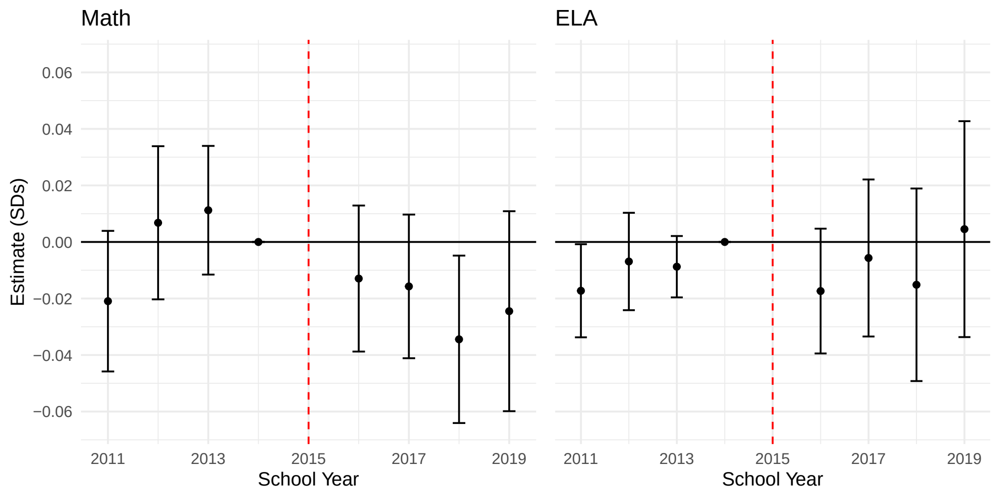
Independent Variable:	Post × Δ Flight Leq		Post × Δ SEL Events per Hour	
Dependent Variable:	Math	ELA	Math	ELA
Estimate	-0.0224**	-0.0004	-0.0071*	0.0008
	(0.0095)	(0.0120)	(0.0036)	(0.0033)
Observations	63,514	63,602	63,514	63,602

*/**/*** denote significance at the $p < 0.1/0.05/0.01$ level. Test performance is measured in standard deviations from the mean score.
Regressions include fixed effects for school-grade, gender, race and ethnicity, English as a second language, and receipt of free or reduced price lunch.
Standard errors are clustered at the school level.

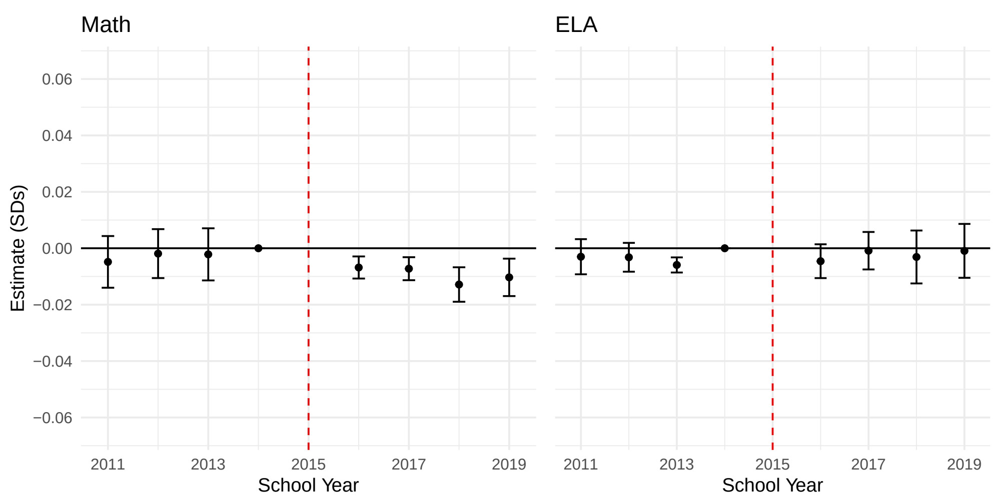
Appendix Table 2: Difference in Differences Estimates of Impact on Demographics

Independent Variable:	Post × Δ Flight Leq			Post × Δ SEL Events per Hour		
Dependent Variable:	FRPL	ESL	URM	FRPL	ESL	URM
Estimate	0.0020	0.0034	0.0063	-0.0015	-0.0047	0.0030
	(0.0061)	(0.0092)	(0.0077)	(0.0020)	(0.0028)	(0.0027)
Observations	63,706	63,706	59,117	63,706	63,706	59,117

*/**/*** denote significance at the $p < 0.1/0.05/0.01$ level. FRPL is an indicator variable for receipt of free or reduced price lunch. ESL is an indicator variable for whether English is the student's second language. URM is an indicator for whether the student is from an underrepresented minority racial/ethnic group.
Regressions include fixed effects for school-grade.
Standard errors are clustered at the school level.



Appendix Figure 6: Event Study Plot of Difference in Differences Model for Δ Flight Leq



Appendix Figure 7: Event Study Plot of Difference in Differences Model for Δ SEL Events