School Bus Diesel Retrofits, Air Quality, and Academic Performance: National Evidence Using Satellite Data

Wes Austin
Georgia State University
gaustin4@gsu.edu

October 25, 2019

Abstract

Prior work shows that air pollution affects cognitive performance. School bus diesel emissions meaningfully contribute to this exposure for school-age children. I exploit variation in the timing and location of 17,901 school bus diesel engine retrofits or replacements across the US from 2008 to 2016 to test how these bus fleet investments affect air quality and student test scores. I use satellite-based fine particulate matter (PM 2.5) measurements from the Atmospheric Composition Analysis Group to provide the first evidence that these engine retrofits significantly improve surface-level ambient air quality, suggesting potentially large spillover benefits. Retrofitting school buses is also associated with a 0.05-0.06 standard deviation increase in standardized test scores. Moreover, each additional $\mu g/m^3$ of fine particulate matter is associated with a precisely-estimated decrease in English and math test scores of 0.0056 standard deviations. Finally, I calculate the benefit of these test score and air quality improvements, finding that \$170 million spent in grants by the EPA led to approximately \$4.75 billion in external benefits. Whether considered from a mortality and clinic cost or test score perspective, the retrofits pass a benefit-cost test.

Keywords: Satellite Air Pollution Measures; Test scores; School bus; Diesel retrofit.
JEL: *I18, I20, Q53*

1 Introduction

Diesel emissions are more harmful than gasoline emissions, containing higher levels of particulate matter, nitrogen dioxide, gaseous aldehydes, carbon monoxide, and polycyclic hydrocarbons.[1] Despite composing a small fraction of all vehicles, diesel automobiles contribute a third of nitrogen oxide emissions and a quarter of particulate matter emissions.[2] School buses play no small role in this discrepancy; older buses are sometimes dirtier emitters than tractor trailers.[3] Unlike tractor-trailers, school buses concentrate their routes in crowded school zones and residential neighborhoods, contributing to pollution exposure both for riders and bystanders.[4] Recent evidence suggests this pollution affects the health and cognitive performance of students.[5]

A simple, cheap, and effective way to reduce emissions from school buses is to install pollution abatement engine modifications, henceforth simply "retrofits."[6] One common tailpipe retrofit, a diesel particulate filter, may decrease overall emissions of fine particulate matter between 60% and 90%, while decreasing in-cabin pollution levels by 15-26%.[7] A combination of a tailpipe filter and crankcase modification reducing draft-tube self pollution may reduce in-cabin pollution to background levels.[8] These improvements persist with good engine care; reductions in particulate matter emissions may remain up to 95% by mass after four years of use.[9]

A growing literature examines how air quality affects cognitive performance.[10] The medical literature has demonstrated that ultrafine particles in air pollution deposit in the brain via the olfactory bulb, leading to inflammation, behavioral changes, and cognitive impairment.[11] The economics literature supports this evidence, showing that air quality affects the quality of political speeches and the likelihood of errors in chess tournaments.[12] In a causal study exploiting variation in wind direction, Heissel et al. (2019) demonstrate that air quality also affects student test scores.

[1] Commins et al. (1957); Muzyka et al. (1998); Waller et al. (1985).
[2] Alexander and Schwandt (2019); EPA (2003).
[3] Harder (2005); Monahan (2006).
[4] Marshall and Behrentz (2005); Ngo (2017); Xu et al. (2016); Zuurbier et al. (2010).
[5] Austin et al. (2019).
[6] Barone et al. (2010); Tate et al. (2017).
[7] Biswas et al. (2009); EPA (2003); Hammond et al. (2007).
[8] Borak and Sirianni (2007); Jiang et al. (2018); Li et al. (2015); Zhang and Zhu (2011).
[9] Barone et al. (2010).
[10] Currie and Neidell (2005); Lavy et al. (2014).
[11] Calderón-Garcidueñas et al. (2012); Freire et al. (2010); Guxens and Sunyer (2012); Sunyer et al. (2015).
[12] Heyes et al. (2019); Künn et al. (2019).

Such negative impacts stem from both contemporaneous and long-term exposure.[13] [14] I join the literature on air quality and cognitive performance to a more limited body of work on the relationship between school bus emissions, student health, and test scores.[15]

The purpose of this paper is threefold. First, I estimate how diesel retrofits improved ambient air quality. Next, I determine the average national relationship between improved school bus emissions and student cognitive performance. Finally, I provide benefit-cost estimates of federal school bus retrofit grants by quantifying the mortality, hospital visit, and cognitive benefits of the retrofits. To address potential endogeneity between diesel retrofits, air quality, and student cognitive performance, I exploit variation in the timing and location of 17,901 diesel bus retrofits over 219 grant cycles across 137 counties from 2008 to 2016. I adopt treatment measures related to the total number of buses retrofitted to estimate the effect on air quality. For cognitive performance, I estimate the relationship between the likely proportion of a county bus fleet retrofitted and English and math standardized test scores for all students in 3rd to 8th grade.

I find that retrofits improve ambient air quality and student performance on standardized tests. Each retrofitted bus improved ambient particulate matter by 0.0003-0.0005 $\mu g/m^3$ per month. Moreover, retrofitting all school buses in a county is associated with at least a 0.05 SD increase in English Scores and a 0.06 increase in math scores. I also show the baseline relationship between fine particulate matter and test scores in a county, finding that each additional $\mu g/m^3$ of fine particulate matter decreases test scores by precisely 0.0056 standard deviations in both language arts and math test scores. Results closely resemble previous findings on the link between diesel retrofits and cognitive performance. Results also support prior evidence that school bus emissions, and air quality more broadly, impact cognitive performance. Finally, the benefits of the retrofits likely far outweigh the amount spent to clean bus fleets.

Unlike two previous studies on school bus diesel retrofits, this paper expands our understanding of the causal mechanism linking retrofits, health, and cognitive performance by estimating the relationship between retrofits and ambient air quality. Beatty and Shimshack (2011), for example, are unable to test whether 4,000 school bus retrofits improved ambient air quality in Washington

[13] Chen et al. (2017); Ebenstein et al. (2016); Ham et al. (2014).

[14] The effects on test scores may arise, in part, from increased absences due to pollution-related illness (see Currie et al. (2009)).

[15] Adar et al. (2015); Austin et al. (2019); Beatty and Shimshack (2011).

state due to the sparsity of air quality monitors. Austin et al. (2019) and Adar et al. (2015), meanwhile, do not consider changes to ambient atmospheric air quality in their analyses, which focus instead on in-cabin changes in air quality. I also provide external validity to a previous study on the relationship between school bus emissions and student test scores. A third contribution of this paper is an expanded benefit-cost analysis of diesel retrofits relying both on test scores and mortality changes.

2 Data

2.1 School Bus Retrofits

I obtained extensive information on all EPA-funded diesel retrofits through a Freedom of Information Act request. 26,000 unique diesel engine retrofit episodes are covered in the EPA report. These retrofits were disbursed through 178 separate projects from 2008 to 2016. Many of these projects were funded by congress through the Diesel Emissions Reduction Act and the American Investment and Recovery Act. Relevant information in the EPA report includes the type of vehicle affected, the type of technology installed, the grant amount, the application and recipient locations, and estimates of various pollutants reduced.[16] Sample restrictions eliminate non-bus engine retrofits and retrofits not clearly linked to a (single) county location.[17] 18,914 school buses remain after sample restrictions. The bus projects entailed disbursements of $170 million in grants. Figure 2 maps counties receiving grants. Table 1 shows that a typical retrofitting district improved 73 buses, or close to 21% of its bus fleet, in each retrofit cycle.

2.2 Particulate Matter 2.5

I use satellite-based monthly PM 2.5 concentration estimates to determine the relationship between school bus retrofits and surface-level ambient air quality. The Atmospheric Composition Analysis Group at Dalhousie University publishes satellite-based monthly PM 2.5 concentration estimates

[16] Estimates of pollutants reduced are mostly based on the Diesel Emissions Quantifier. They include nitrogen oxides, particulate matter, hydrocarbons, carbon monoxide, and carbon dioxide.

[17] Grants to freight trucks, agriculture equipment, and construction vehicles are common. Many grants went to state environmental protection divisions, who then broke up the funds among many municipalities. I exclude any grants going to multiple regions and any grants not clearly linked to one single county.

over a long time horizon.[18] These data are created by applying a machine-learning algorithm to repeated daily satellite images of aerosol optical depth, a measure of cloudiness, across small pixels of coordinate grid size 0.01*0.01 on the earth's surface.[19] Using GIS software, I convert these raster-pixel data to county-month variables for the average, minimum, and maximum PM 2.5 for each month from 2000-2017.[20] Figure 1 displays these pixel average particulate matter data points in North America in December of 2016. One advantage to satellite-based data is a wider coverage region than would be possible using air quality monitors, so more retrofits can be included in the sample. Coverage across counties is also even, lowering measurement error associated with monitor location and allowing school bus retrofits occurring in any portion of a county to be reflected in the county-month means. Prediction errors may render satellite-based estimates less accurate for tiny regions or high pollution levels.[21] A recent study nevertheless demonstrated very similar fetal health outcomes from diesel-related air pollution when using either satellite-based or monitor-based air quality measurements at the county level.[22] I take this as evidence that the county is a suitable unit of aggregation for this study, which analyzes trends over many years.

2.3 Academic Achievement

The Stanford Education Data Archive provides standardized English language arts and math test score estimates at the school, school district, and county level for 2009 to 2016.[23] I use the long-form county files, which provide standardized test scores for each subject-grade-year-county combination.[24] Achievement information is populated based on raw data files in the EDFacts data system housed by the U.S. Department of Education (USEd). EDFacts is composed of test score information aggregated by student subgroups; each cell is a school-subject-grade-year-subgroup proficiency

[18] Monthly estimates for North America were downloaded via secure ftp here. Files are named GWR-wSPEare_(yearmonth)_(yearmonth)_RH35-NoNegs.asc.zip.

[19] van Donkelaar et al. (2019).

[20] In ArcGIS, I use the extract by polygon tool to assign average pixel values to the entire region within a county. I therefore assign the average concentration of particulate matter in a county region as equal to the average of the concentrations in all pixel regions within that county.

[21] Fowlie et al. (2019). Indeed, the creators of the Dalhousie Atmospheric Composition data caution that "Users are reminded that these datasets are intended for long-term, large-scale studies. Increased uncertainties are expected when used at finer spatial/temporal resolution."

[22] Alexander and Schwandt (2019). The authors were provided precisely the same re-constructed air quality panel as this study, in part to test whether the use of a county-month panel is reasonably unbiased.

[23] Reardon et al. (2017b)

[24] Note this data file is titled "SEDA_county_long_cs_v30.dta" and can be accessed here.

count indicator.[25] The Stanford Education Data Archive perform many steps to convert these school-level proficiency counts into county-level means that are comparable across states (with associated standard errors).[26] I use the CS mean scale (rather than the GCS scale) because it is intended to be interpretable as an effect size relative to the grade-specific standard deviation of scores in a common cohort. I collapse these mean z-scores across grades within a county, weighting by cohort size, to derive a mean English language arts and math z-score for each county and year. The Archive also creates covariate information describing county socioeconomic, demographic, and segregation information of families with children enrolled in the schools represented in the test-score data.[27] I summarize these covariate values in retrofitting and non-retrofitting counties in Table 1. Counties receiving retrofit grants are notably larger, more urban, more educated, and better-performing on standardized tests.

A wide variety of concerns may be raised with respect to use of the Stanford Education Data Archive. Test score means are estimated, geographic aggregation may be problematic, (non-virtual) charter schools are included, data is suppressed in cells with fewer than 20 students, and, among other issues, additional noise is even intentional randomly injected into the data. To deal with some of these potential idiosyncrasies, all models include state-year fixed effects. In addition, covariate information, which may be imprecisely measured or imputed and used in estimation of test scores, is not included as a control in any model. Rather, it is used to assess differences across counties that do or do not receive retrofit grants. I also test whether the test scores may credibly capture changes in air quality in Table 3.

3 Empirical Strategy

To discern the relationship between diesel school bus retrofits and fine particulate matter concentration, I exploit variation in the county location of a retrofit, the year of the retrofit, and months

[25]For example, an observation may represent the number of students of a given ethnicity in a given grade in a given school who do not meet basic standards on a state-standardized test of the given subject.

[26]A full description of these steps is beyond the scope of this paper. For more information on the procedure, see Reardon et al. (2017a).

[27]Covariate information is derived from the Education Demographic and Geographic Estimates (EDGE) database, which is based on the American Community Survey, and the Common Core of Data (CCD), which is an annual survey of all schools and school districts in the nation. For more information on how covariate information is populated, imputed, or estimated, see Reardon et al. (2017a). Covariate files used for summary statistics in this paper can be found in the data file "SEDA_cov_county_long_v30.dta."

of the year in which school buses are most likely to be used. Consider the following regression specification:

$$PM_{imst} = \beta Buses_{imt} + \tau_{im} + \tau_{st} + \eta_i + \epsilon_{imst}. \tag{1}$$

PM_{imst} is the average concentration of atmospheric PM 2.5 in $\mu g/m^3$ in county i, month m, state s, and year t. τ_{im} is a county-month fixed effect to control for average pollution levels at different times of the year in a county. τ_{st}, a state-year fixed effect, controls for secular changes in pollution levels in a given state. I use a state-year fixed effect instead of a year fixed effect because within-state secular changes in air pollution are a better control than national changes in annual levels of air pollution. η_i is a county fixed effect to control for any time-constant factors affecting pollution in a county. I consider two formulations of the treatment variable, $Buses_{imt}$, which represents the total number of vehicles that receive engine modifications. In the first formulation, I assign the number of buses modified to the county-year of the retrofit from the months of September to the following September.[28] This formulation of $Buses_{imt}$ asks how particulate matter changes only in the year following the retrofit, where the retrofit air quality improvements are assumed to begin in the month of September. In the second formulation of $Buses_{imt}$, I assign the number of buses retrofitted to the county-month of September in the year of the retrofit as before, and then I accumulate this variable with additional retrofits in all future county months.[29] This formulation of $Buses_{imt}$ is designed to test for persistent improvements in air quality from all bus retrofits combined.

Intuitively, Equation 1 estimates the per-vehicle change in a county's monthly average PM 2.5 concentration, comparing average monthly concentration in the county before the retrofit occurred to the concentration after the retrofit. Identification requires that no county-level policy or institutional changes are associated both with the number of buses retrofitted and the average change in air quality, holding annual state and monthly county fluctuations constant. A violation of the

[28]For example, I assign a value of 0 to the month of August in 2010 in a county with a retrofit of 20 buses occurring in 2010, while the month of September of the same year would receive a value of 20. September of the following year, 2011, would also receive a value of 20, but September of the year after thatwould receive a value of 0 again.

[29]For example, a retrofit of 20 buses in 2010 followed by a retrofit of 10 buses in 2012 would receive a value of 0 until September of 2010, 20 until September of 2012, and then 30 in all future periods.

identifying assumption may occur if a county that retrofits many buses also performs other actions that improve air quality in the same year of the retrofit.

To test for changes in test scores associated with school bus retrofits, I adopt both a first-differences estimation strategy with state-year fixed effects and a fixed effects panel model. Following Austin et al. (2019), the first differences approach controls for unobservable county-level characteristics that might be correlated with the timing of the retrofit. The state-year fixed effects control for any state changes to testing practices, educational institutions, or SEDA estimation procedures common to all counties in a state-year pair. Consider the following estimating equation:

$$\Delta y_{it} = \beta Retrofit_{it} + \tau_{st} + \Delta\epsilon_{it}. \tag{2}$$

The dependent variable y_{it} is math and English langage arts and math test z-scores. Δ represents a change in a variable from year $t-1$ to year t. τ_{st} is a state-year fixed effect that controls for any institutional or schooling changes in a state in a given year. The first-differences model in Equation 2 captures year-on-year changes in test scores resulting from retrofitting a share of a county's bus fleet in year t, $Retrofit_{it}$.[30] For example, if 10% of a county's likely bus fleet is retrofitted in year t, $Retrofit_{it}$ is 0.1. In all non-retrofit periods, $Retrofit_{it}$ is set equal to zero. I use the share, rather than perhaps a simple count variable, because the same number of retrofitted buses is likely to have a different effect when spread over many vs. fewer students. In the presence of serial correlation, the Equation 2 has advantages over a fixed effects model. However, the first differences model requires knowing the exact timing of a retrofit; every retrofit assigned to the wrong year will bias estimates to zero. For this reason, I prefer a fixed effects model in which $Retrofit_{it}$ accumulates over time, allowing poorly-timed treatments to be partially captured in any post-retrofit test-score changes. Consider the following regression:

$$y_{it} = \beta Retrofit_{it} + \tau_{st} + \eta_i + \epsilon_{it}. \tag{3}$$

[30] I do not observe the size of a county bus fleet. I therefore assume that all counties have one bus per 55 students. This corresponds to the bus-to-student ratio observed in Austin et al. (2019). The proportion of a fleet retrofitted is therefore calculated as the quotient of the number of vehicles retrofitted and the annual estimated bus fleet size (i.e. $\frac{TotalStudents}{55}$).

y_{it} and τ_{st} are as before; η_i controls for time-constant county-level factors. β is the relationship between the proportion of a bus fleet retrofitted and English language arts and math test scores in county t. Now, I consider two formulations of $Retrofit_{it}$. In the first, $Retrofit_{it}$ is the proportion of a fleet retrofitted in year t, and zero in all years without retrofits. In the second, $Retrofit_{it}$ is the cumulative share of buses retrofitted in a county.

The timing and magnitude of the retrofits, holding county and state-year conditions constant, provides identifying variation. The identifying assumption is that there are no factors correlated with the proportion of a bus fleet retrofitted and changes in test scores within a county. This assumption might be violated if counties more likely to install diesel engine retrofits are more likely to make other changes that improve academic conditions and hence test scores.

4 Results

4.1 Air Pollution

I present findings on the relationship between diesel school bus retrofits and surface-level concentrations of fine particulate matter in Table 2. Column (1) suggests that, for each bus retrofitted in county i, the average monthly concentration of particulate matter in that county in the following year fell by 0.0004 $\mu g/m^3$. The coefficient in column (2) suggests that the average change in particulate matter concentration from year $t-1$ to retrofitting year t is -0.0005 across all likely affected months m. Since column (2) reflects year to year changes from, for example, February of 2015 to February of 2016, county-month fixed effects are not necessary. Columns (3) and (4) perform the same analysis, except instead of testing for changes only in the year following a retrofit, I test for changes in all months after a retrofit likely took place. The coefficient in column (3) therefore suggests that each retrofitted bus decreased monthly average particulate matter concentrations in the same county by 0.00036 $\mu g/m^3$ in all future periods. Although these effect sizes are small in relation to the county average of 9.122 $\mu g/m^3$, they represents a per-bus decrease, where counties typically retrofitted 73 buses per retrofit cycle or 172 over the entire sample. The row labeled Δ PM Concentration therefore scales these effect sizes by the average relevant number of buses driving the effect. It should be noted that these effects represent changes in average concentration across all pixels within a county; they therefore reflect much larger air quality improvements in highly-affected environments such residential neighborhoods and schools.

4.2 Academic Achievement

Before assessing the relationship between school bus retrofits and cognitive performance, I first test the average association between fine particulate matter concentration and performance on standardized English language arts and math test scores in Table 3. While this relationship is interesting by itself, it also provides some credibility to the use of SEDA test scores, which are based on estimations of means from school-level proficiency counts. I regress two measures of air quality on two variations of standardized test score outcome. The measures of county air quality are year-to-year changes in particulate matter concentration in a county and also simply the average levels of particulate matter. I split the academic outcomes similarly by year-on-year changes and average levels. My preferred specifications are columns (1) and (2), which correspond roughly to the first-differences specification laid out in Equation 2, and those in columns (7) and (8), which correspond to Equation 3. Although the first-differences estimates are noisy, the fixed effects specification in columns (7) and (8) are precisely estimated. They suggest that each additional $\mu g/m^3$ of fine particulate matter decreases test scores by 0.0056 standard deviations across both language arts and math test scores. I take this as evidence that the SEDA test scores are likely credible measures of standardized test performance.

I present findings on the relationship between diesel school bus retrofits and cognitive performance in Table 4. Outcomes are county-level English language arts (ELA) z-scores and math z-scores averaged across all test-takers in grades 3 through 8. Unlike the air quality regressions in Table 2, these regressions only cover the years for which SEDA test scores are available, 2009-2016. Columns (1) through (4) show the relationship between retrofits and year-on-year changes in test scores, where columns (1) and (2) directly replicate Austin et al. (2019). The coefficient in column 1 implies that retrofitting an entire bus fleet would raise ELA scores in the following year by 0.01 standard deviations, while that of column (2) suggests retrofitting an entire fleet would improve math scores by 0.05 z-scores. The coefficients in columns (3) and (4) imply that retrofitting an entire bus fleet in year t would affect average improvements in test scores in all future periods by 0.00065 and -0.00322. Since this relationship is not easily interpretable or meaningful, especially when counties have multiple retrofit cycles, it is not surprising that both coefficients in these columns are precisely estimated zeros. When replicating the identification strategy of Austin et al. (2019) in columns (1) and (2), the coefficient on ELA is much smaller and insignificiant than in the previous study, while the results for math are very similar. This discrepancy may relate to the relatively noisier retrofit timing employed in this paper or issues in the underlying SEDA test-score estimation.

Columns (4) through (8) do not difference the outcomes across years, and therefore they include county fixed effects to control for time-constant differences across counties. The coefficient in column (5) suggests that retrofitting an entire bus fleet would improve ELA z-scores in the year of the retrofit by 0.0678. The coefficient in column (6) implies that retrofitting an entire bus fleet would improve math z-scores in the year of the retrofit by 0.086. When regressing the cumulative proportion of a bus fleet retrofitted on test z-scores, these improvements are 0.0549 and 0.06 in all post-retrofit periods. Since columns (7) and (8) implicitly test for improvements that last for the remainder of the sample, it is not surprising that they are slightly smaller than the coefficients in columns (5) and (6) due to likely depreciation of diesel parts. The results in columns (5) through (8) closely resemble the results of Austin et al. (2019), which found z-score improvements of 0.08 and 0.05 for ELA and math, respectively, in a similar study in Georgia. Improvements in test scores observed in columns (5) through (8) are comparable to the difference in scores observed between students of a rookie teacher and those of a teacher with five years of experience.[31] Since the average retrofitting county in the national sample retrofitted a cumulative 20% of its fleet, the average observed county improvement in test scores is likely between 0.01 and 0.02 z-scores.

5 Benefit-Cost Analysis

I estimate the benefits of school bus engine grants with respect to two outcomes. First, I incorporate valuations of particulate matter improvements to estimate the benefit of air quality improvements associated with the retrofits. Next, I quantify the long-term wage benefits of higher test scores. I then compare these benefits to the overall grant dollars distributed.

To determine the benefits of air quality improvements, I borrow estimates from a recent study exploiting changes in wind direction to estimate the mortality and hospital visit cost of fine partic-

[31] Rice (2010).

ulate matter pollution. Deryugina et al. (2016) find that a 1 $\mu g/m^3$ increase in PM 2.5 causes the loss of 2.99 life-years per million beneficiaries over three days, implying a mortality cost of $299,000 per million affected individuals; hospital visits from the same change in air quality cost $19,000. I combine these measures, assuming external damages per unit increase of $\mu g/m^3$ of $320,000 over three days. For simplicity, I assume that the improvements in air quality from retrofits last only one year. Then, the external benefit of the diesel school bus retrofits can be conservatively estimated as $245 million.[32] Since the total amount awarded by the EPA in my sample is $170 million, these conservative external health improvement estimates pass a benefit-cost test.[33]

I calculate the benefit of test score improvements based on the monetizations in Chetty et al. (2011). According to their analysis of the Tennessee STAR experiment, a one-percentile increase in test scores in Kindergarten is associated with an increase in lifetime earnings of $1,041.[34] The results presented in Table 3, columns (7) and (8) indicate that retrofitting 100% of a district's fleet will increase the z-score of the ELA tests by 0.05 and of the math tests by 0.06. Scale these to the average proportion of a fleet retrofitted (i.e. 20%), and then convert z-scores to percentile increases of 0.39 and 0.47, respectively. Using the average of these two values (0.43), and multiplying by the valuation implied by the Chetty et al. (2011) estimates, the benefit of retrofitting 20% of a district's fleet may be valued at $4.5 billion.[35] This is roughly 26 times the amount spent on school bus retrofit grants in my sample, which was $170 million. Interestingly, this is nearly the exact same benefit ratio determined in Austin et al. (2019), which found test-score benefit-cost ratios of 25 times the cost of the retrofits.

[32] Loss in life years from a unit increase in PM concentration over 365 days is 2.99*121.66= 363.78 life years per million inhabitants in a retrofit year. Retrofitting counties had an average population of 199,427 in 2014. Since a typical retrofit led to a fall in particulate matter concentration of 0.0481 in the following year, this means an average retrofit can be valued at $(320,000 * 200,000 * .0481 * 363.78/1,000,000 =)$ $1.12 million in mortality and clinic costs. Multiplying this by the total number of retrofit cycles results in a total benefit of $245.2 million.

[33] It is worth noting that these estimations are sensitive to average population size in retrofitting counties; that counties with larger populations are more likely to receive retrofits inflates the external health benefit. Similarly, allowing pollution abatements to last more than one year would dramatically increase the external health valuation.

[34] Assume improvement of $94 in wage earnings at age 27 lasts from ages 25-54 and discount at an annual rate of 3%.

[35] 0.43 percentile points $*\$1,041$ per percentile point per student $*46,000$ students per retrofitting district $*219$ total retrofit cycles.

6 Conclusion

I estimate the effect of retrofitting diesel school bus engines on ambient air quality and academic achievement. Retrofit cycles reduce one of the most harmful diesel emissions, fine particulate matter 2.5, by between 0.4% and 0.7%. They are also associated with positive and significant improvements in English language arts and math tests scores. Back-of-the-envelope calculations reinforce the findings of previous studies, demonstrating that the benefits of the retrofits were likely magnitudes greater than the costs. This study could be extended by testing for differences in test score improvements across demographic groups and type of city. Health impacts from school bus retrofits could also be analyzed directly by, for example, estimating changes in mortality or fetal health in retrofitting counties. The same analysis could be extended to other pollutants influenced by school bus diesel emissions, such as nitrogen oxides. Results have policy relevance, as they extend prior work on school bus retrofits in showing that external spillover benefits of the air quality improvements may have been underestimated.

References

Adar, S. D., D'Souza, J., Sheppard, L., Kaufman, J. D., Hallstrand, T. S., Davey, M. E., Sullivan, J. R., Jahnke, J., Koenig, J., Larson, T. V., et al. (2015). Adopting clean fuels and technologies on school buses. pollution and health impacts in children. *American journal of respiratory and critical care medicine*, 191(12):1413–1421.

Alexander, D. and Schwandt, H. (2019). The impact of car pollution on infant and child health: Evidence from emissions cheating. IZA Discussion Papers 12427, Institute of Labor Economics (IZA).

Austin, W., Heutel, G., and Kreisman, D. (2019). School Bus Emissions, Student Health, and Academic Performance. NBER Working Papers 25641, National Bureau of Economic Research, Inc.

Barone, T. L., Storey, J. M., and Domingo, N. (2010). An analysis of field-aged diesel particulate filter performance: Particle emissions before, during, and after regeneration. *Journal of the Air & Waste Management Association*, 60(8):968–976.

Beatty, T. and Shimshack, J. (2011). School buses, diesel emissions, and respiratory health. *Journal of Health Economics*, 30(5):987–999.

Biswas, S., Verma, V., Schauer, J. J., and Sioutas, C. (2009). Chemical speciation of pm emissions from heavy-duty diesel vehicles equipped with diesel particulate filter (dpf) and selective catalytic reduction (scr) retrofits. *Atmospheric Environment*, 43(11):1917 – 1925.

Borak, J. and Sirianni, G. (2007). Studies of self-pollution in diesel school buses: methodological issues. *Journal of occupational and environmental hygiene*, 4(9):660–668.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Styner, M., Gómez-garza, G., Zhu, H., Torres-Jardón, R., Carlos, E., Solorio-López, E., Medina-Cortina, H., Kavanaugh, M., and D'Angiulli, A. (2012). White matter hyperintensities, systemic inflammation, brain growth, and cognitive functions in children exposed to air pollution. *Journal of Alzheimer's Disease*, 31(1):183–191.

Chen, X., Zhang, X., and Zhang, X. (2017). Smog in Our Brains: Gender Differences in the Impact of Exposure to Air Pollution on Cognitive Performance.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Commins, B. T., Waller, R. E., and Lawther, P. J. (1957). Air pollution in diesel bus garages. *British Journal of Industrial Medicine*, 14(4):232–239.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does Pollution Increase School Absences? *The Review of Economics and Statistics*, 91(4):682–694.

Currie, J. and Neidell, M. (2005). Air pollution and infant health: What can we learn from california's recent experience?*. *The Quarterly Journal of Economics*, 120(3):1003–1030.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2016). The mortality and medical costs of air pollution: Evidence from changes in wind direction. Working Paper 22796, National Bureau of Economic Research.

Ebenstein, A., Lavy, V., and Roth, S. (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.

EPA (2003). Technical highlights: Questions and answers on using a diesel particulate matter filter in heavy-duty trucks and buses. Technical report, Environmental Protection Agency Office of Transportation and Air Quality.

Fowlie, M., Rubin, E., and Walker, R. (2019). Bringing Satellite-Based Air Quality Estimates Down to Earth. *AEA Papers and Proceedings*, 109:283–288.

Freire, C., Ramos, R., Puertas, R., Lopez-Espinosa, M.-J., Julvez, J., Aguilera, I., Cruz, F., Fernandez, M.-F., Sunyer, J., and Olea, N. (2010). Association of traffic-related air pollution with cognitive development in children. *Journal of Epidemiology & Community Health*, 64(3):223–228.

Guxens, M. and Sunyer, J. (2012). A review of epidemiological studies on neuropsychological effects of air pollution. *The European Journal of Medical Sciences*, 141(3):1–7.

Ham, J. C., Zweig, J. S., and Avol, E. (2014).

Hammond, D., M. Lalor, M., and Jones, S. (2007). In-vehicle measurement of particle number concentrations on school buses equipped with diesel retrofits. *Water, Air, and Soil Pollution*, 179:217–225.

Harder, B. (2005). School buses spew pollution into young lungs. *Science News*, 167(21):334–334.

Heissel, J., Persico, C., and Simon, D. (2019). Does pollution drive achievement? the effect of traffic pollution on academic performance. Working Paper 25489, National Bureau of Economic Research.

Heyes, A., Rivers, N., and Schaufele, B. (2019). Pollution and politician productivity: The effect of pm on mps. *Land Economics*, 95:157–173.

Jiang, Y., Yang, J., Cocker, D., Karavalakis, G., Johnson, K. C., and Durbin, T. D. (2018). Characterizing emission rates of regulated pollutants from model year 2012+ heavy-duty diesel vehicles equipped with dpf and scr systems. *Science of The Total Environment*, 619-620:765 – 771.

Künn, S., Palacios, J., and Pestel, N. (2019). Indoor Air Quality and Cognitive Performance. IZA Discussion Papers 12632, Institute of Labor Economics (IZA).

Lavy, V., Ebenstein, A., and Roth, S. (2014). The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation. Working Paper 20648, National Bureau of Economic Research.

Li, F., Lee, E. S., Liu, J., and Zhu, Y. (2015). Predicting self-pollution inside school buses using a cfd and multi-zone coupled model. *Atmospheric Environment*, 107:16–23.

Marshall, J. D. and Behrentz, E. (2005). Vehicle self-pollution intake fraction: children's exposure to school bus emissions. *Environmental Science & Technology*, 39(8):2559–2563. PMID: 15884349.

Monahan, P. (2006). School Bus Pollution Report Card 2006 : Grading the States. Report, Union of Concerned Scientists.

Muzyka, V., Veimer, S., and Shmidt, N. (1998). Particle-bound benzene from diesel engine exhaust. *Scandinavian Journal of Work, Environment Health*, 24(6):481–485.

Ngo, N. S. (2017). Emission standards, public transit, and infant health. *Journal of policy analysis and management : [the journal of the Association for Public Policy Analysis and Management]*, 36 4:773–89.

Reardon, S. F., Kalogrides, D., Ho, A., Shear, B., Shores, K., and Fahle, E. (2017a). Stanford education data archive: Technical documentation (version 2.0). http://purl.stanford.edu/db586ns4974.

Reardon, S. F., Kalogrides, D., Ho, A., Shear, B., Shores, K., and Fahle, E. (2017b). Stanford education data archive version 2.0. http://purl.stanford.edu/db586ns4974.

Rice, J. K. (2010). The impact of teacher experience examining the evidence and policy implications. *CALDER*, Brief 11.

Sunyer, J., Esnaola, M., Alvarez-Pedrerol, M., Forns, J., Rivas, I., López-Vicente, M., Suades-González, E., Foraster, M., Garcia-Esteban, R., Basagaña, X., Viana, M., Cirach, M., Moreno, T., Alastuey, A., Sebastian-Galles, N., Nieuwenhuijsen, M., and Querol, X. (2015). Association between traffic-related air pollution in schools and cognitive development in primary school children: A prospective cohort study. *PLOS Medicine*, 12(3):1–24.

Tate, J., Mason, R., and Schmitt, L. (2017). 2050 - the air quality emissions and health benefits of cleaner buses: A city of york (uk) case study using micro-scale models and a health impact toolkit. *Journal of Transport Health*, 5:S41.

van Donkelaar, A., Martin, R., Li, C., and T Burnett, R. (2019). Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environmental Science Technology*, 53.

Waller, R. E., Hampton, L., and Lawther, P. J. (1985). A further study of air pollution in diesel bus garages. *British Journal of Industrial Medicine*, 42(12):824–830.

Xu, W., Mai, G., Zhu, Q., Yu, Z., and Liu, Y. (2016). Pollution exposure at bus commuter stations in guangzhou, china. *International Journal of Environmental Technology and Management*, 19(2):103–119.

Zhang, Q. and Zhu, Y. (2011). Performance of school bus retrofit systems: ultrafine particles and other vehicular pollutants. *Environmental science & technology*, 45(15):6475–6482.

Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., van den Hazel, P., and Brunekreef, B. (2010). Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, 118(6):783–789.

Tables

Table 1: County Characteristics

	(1) Non-Retrofitting Counties		(2) Retrofitting Counties	
Schooling Outcomes (2009-2016)				
English Language Arts Z-scores	-0.041	(0.236)	0.021	(0.193)
Math Z-scores	-0.040	(0.267)	0.046	(0.224)
Total Enrollment, Grades 3-8, thousands	5.890	(15.696)	46.001	(93.738)
% free lunch	0.460	(0.161)	0.411	(0.146)
% reduced-price lunch	0.093	(0.029)	0.078	(0.023)
% free or reduced lunch	0.553	(0.160)	0.489	(0.148)
% ELL Students	0.037	(0.059)	0.066	(0.074)
% Special Ed Students	0.137	(0.038)	0.131	(0.035)
% Urban Schools	0.067	(0.201)	0.299	(0.314)
% Suburban Schools	0.093	(0.235)	0.293	(0.298)
% Town Schools	0.307	(0.338)	0.188	(0.293)
% Rural Schools	0.533	(0.371)	0.220	(0.260)
County-Level Covariates (2009-2016)				
% White	0.730	(0.250)	0.648	(0.258)
% Hispanic	0.119	(0.175)	0.153	(0.174)
% Black	0.113	(0.195)	0.154	(0.184)
% Native american	0.026	(0.092)	0.014	(0.047)
% Asian	0.012	(0.022)	0.031	(0.029)
Log of median income	10.691	(0.223)	10.832	(0.196)
Bachelors and beyond rate	0.142	(0.064)	0.218	(0.080)
Poverty rate	0.163	(0.053)	0.149	(0.055)
Unemployment rate	0.083	(0.027)	0.080	(0.026)
Snap receipt rate	0.132	(0.056)	0.116	(0.046)
Single mother household rate	0.165	(0.053)	0.183	(0.062)
Bus Retrofits and Air Pollution (2000-2017)				
Particulate Matter 2.5 [$\mu g/m^3$]	8.825	(2.300)	9.122	(2.361)
Vehicles Retrofitted in year t	0.000	(0.000)	73.401	(141.99)
Proportion Fleet Retrofitted	0.000	(0.000)	0.2087	(0.2725)
Counties	2725		137	

Mean coefficients reported; standard deviations in parentheses. Observations are at the county level. Other demographic category includes Asian, American Indian, Pacific Islander, and Multiracial. Students represents the average student enrollment in thousands. Standardized math and ELA test scores are negative because the majority of Georgia school districts are rural, small, and under-achieving relative to larger urban districts. Aerobic capacity attempts / enrollment represents the number of attempts divided by K-12 enrollment, where certain grades in high school are never tested on the FitnessGram examination.

Table 2: The Relationship between School Bus Retrofits and County-Level Particulate Matter 2.5 Concentration [$\mu g/m^3$] from 2000-2017

	(1) Monthly Average	(2) Year-on-Year Change	(3) Monthly Average	(4) Year-on-Year Change
Buses Retrofitted in Year t	-0.00042**	-0.00054*		
	(0.0002)	(0.0003)		
Cumulative Buses Retrofitted			-0.00036***	-0.00027
			(0.0001)	(0.0003)
County Fixed Effects	✓	✓	✓	✓
State-year Fixed Effects	✓	✓	✓	✓
County-month Fixed Effects	✓		✓	
Δ PM Concentration	0.0356	0.0481	0.0619	0.046
% Change from Mean	0.39%	0.53%	0.68%	0.50%
R^2	0.739	0.142	0.739	0.142
County-Year-Months	636,072	598,656	636,072	598,656
Counties & County Equivalents	3,118	3,118	3,118	3,118

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the county level in parentheses. Implied % change is effect size times average buses retrofitted divided by baseline average particulate matter concentration. It is calculated assuming mean non-zero buses retrofitted in year t of 73 and mean non-zero cumulative retrofits of 172 school buses, where baseline all-time average PM 2.5 concentration is assumed to be 9.12 in retrofitting counties. County equivalents include parishes, boroughs, and the District of Columbia.

Table 3: The Relationship between County Fine Particulate Matter Concentration [$\mu g/m^3$] and Academic Performance (2009-2016)

	Year-on-Year Change in Test Scores				Annual Average Test Scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ELA	Math	ELA	Math	ELA	Math	ELA	Math
Year-on-Year Change in PM Concentration	-0.000910	-0.00253*	-0.000510	-0.00130				
	(0.0012)	(0.0015)	(0.0008)	(0.0009)				
Average Annual PM Concentration					-0.00299*	-0.00415**	-0.00588***	-0.00558***
					(0.0015)	(0.0019)	(0.0010)	(0.0013)
County Fixed Effects					✓	✓	✓	✓
State-year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.136	0.120	0.015	0.024	0.930	0.916	0.916	0.895
County-Year Observations	19,093	18,997	19,100	19,006	22,699	22,663	22,707	22,672
State-Year Observations	301	297	301	297	364	361	364	361

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the county level in parentheses. Outcomes in columns (1) through (4) are the year-on-year change in ELA or math test scores from year $t-1$ to year t, while columns (5) through (8) outcomes are simply the average standardized test score in year t.

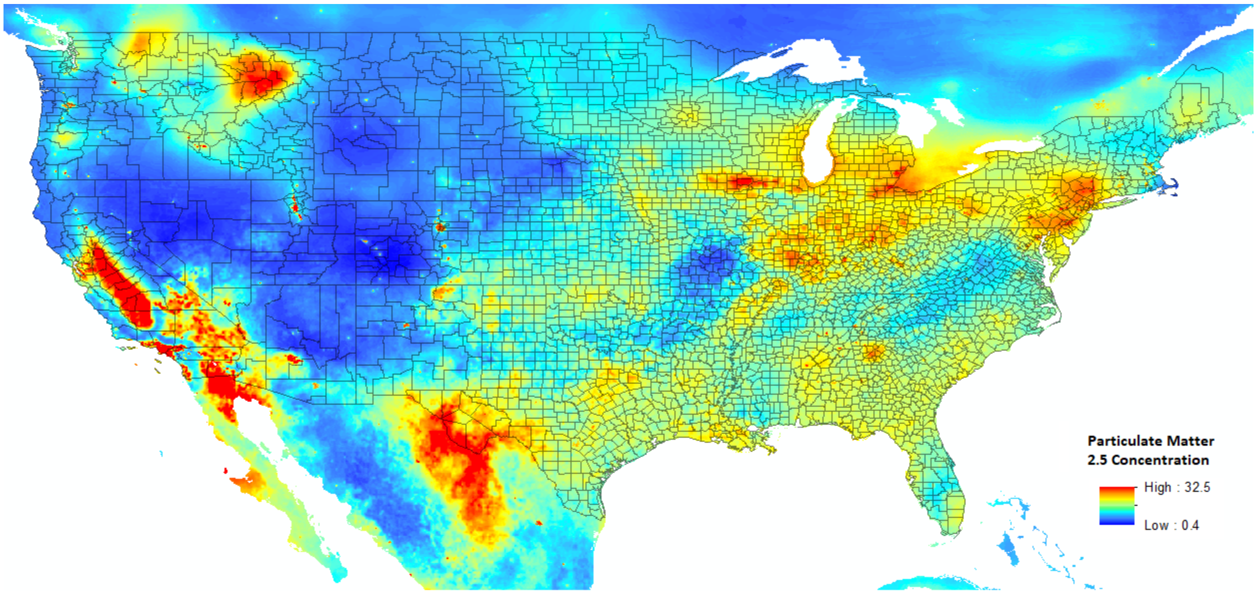
Table 4: The Relationship between School Bus Retrofits and Academic Performance (2009-2016)

	Year-on-Year Change in Test Scores				Annual Average Test Scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ELA	Math	ELA	Math	ELA	Math	ELA	Math
Proportion Retrofitted in Year t	0.0117	0.0513***			0.0678*	0.0860**		
	(0.0122)	(0.0154)			(0.0361)	(0.0408)		
Cumulative Proportion of Fleet			0.00065	-0.00322			0.0549**	0.0603**
Retrofitted			(0.0037)	(0.0053)			(0.0245)	(0.0246)
County Fixed Effects					✓	✓	✓	✓
State-year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.130	0.114	0.130	0.114	0.353	0.414	0.352	0.414
County-Year Observations	19,477	19,266	19,477	19,266	23,076	22,903	23,133	22,961
State-Year Observations	309	305	309	305	372	369	372	369

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the county in parentheses. Columns (1) through (4) regress the proportion of a bus fleet retrofitted in year t or the cumulative proportion of a fleet retrofitted on the change in test z-scores from year $t-1$ to year t. Columns (5) through (8) regress the retrofit treatment variables on the baseline z-scores in county i and year t.

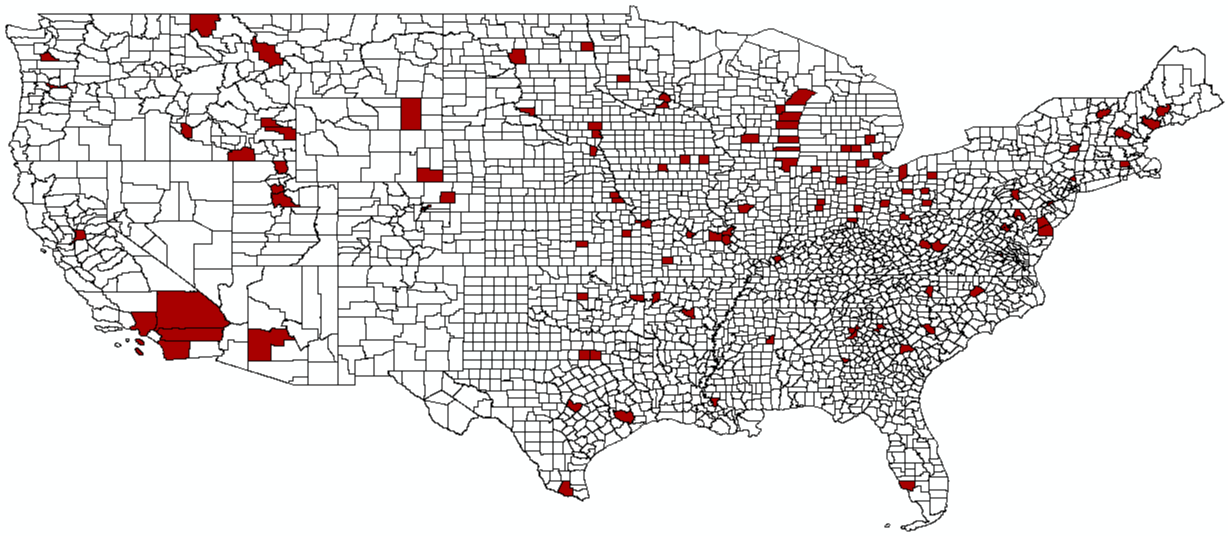
Figures

Figure 1: County-Level Particulate Matter 2.5 Concentration [$\mu g/m^3$] in December of 2016.



Notes: Bright red represents higher particulate matter concentrations, while cooler colors represent less air pollution. The hottest red represents particulate matter concentrations in excess of 30 $\mu g/m^3$.

Figure 2: Counties Receiving EPA Grant Funds to Retrofit School Buses (2008-2016)



Notes: Red counties have at least one retrofit from 2008-2016.