FICAN

Federal Interagency Committee on Aviation Noise

Mr. Alan F. Zusman, Chairman
Department of Navy
Washington Navy Yard
1322 Patterson Ave. SE
Washington, DC 20374-5065
Alan.Zusman@navy.mil

FINDINGS OF THE FICAN PILOT STUDY ON THE RELATIONSHIP BETWEEN AIRCRAFT NOISE REDUCTION AND CHANGES IN STANDARDIZED TEST SCORES

July 2007

Research on the effects of aircraft noise on children's learning suggests that aircraft noise can interfere with learning in the following areas: reading, motivation, language and speech acquisition, and memory. The strongest findings to date are in the area of reading, where more than 20 studies have shown that children in noise impact zones are negatively affected by aircraft. In September 2000, FICAN undertook a pilot study to evaluate the effectiveness of school sound insulation programs. This finding reports on the results of that study.

The study was designed to answer the following: Is abrupt aircraft noise reduction within classrooms related to mandatory, standardized test-score improvement, after controlling for demographics? Does this relationship vary by age group, by student group, and/or by test type? The study included 35 public schools nearby three airports in the U.S. Abrupt noise reduction at these schools was caused by either airport closure or newly implemented sound insulation. In the analysis, the noise-reduction group (each school, before-to-after the summer of noise reduction) was compared to the control group (same school, but for years prior to noise reduction). Analysis consisted of multi-level regression with "change in test scores" regressed against a range of variables such as "change in cumulative noise exposure".

After controlling for demographics, the study found (1) a substantial association between noise reduction and decreased failure (worst-score) rates for high-school students, and (2) significant association between noise reduction and increased average test scores for student/test subgroups. In general, the study found little dependence upon student group and upon test type.

FICAN recommends that additional studies be conducted that expand the scope of this work in several ways: incorporating a larger number of airports and schools; following individual students from year to year; determining which tests were actually given in "teaching" classrooms and which were given elsewhere; obtaining airport data directly from airports; and incorporating actual outdoor-to-indoor measurements at each school. In general, wherever these recommendations increase the amount of data, compared to this current study, they will increase the levels of confidence for all results.

BACKGROUND

Research on the effects of aircraft noise on children's learning suggests that aircraft noise can interfere with learning in the following school subjects: reading, motivation, language and speech acquisition, and memory (Evans, 1998). The strongest findings to date are for the school subject of reading, for which the majority of studies have shown that children in noise-impact zones are negatively affected by aircraft. Recent research, which confirms conclusions from the 1970s, shows learning decreases in reading when outdoor-noise L_{Aeq} is 65 dB or higher (Stansfeld, 2000). It is also possible that, for the same outdoor L_{Aeq}, the effects of aircraft noise on classroom learning may be greater than the effects of road and railroad noise (Hygge, 2003).

In February 2000, the Federal Interagency Committee on Aviation Noise (FICAN) held a public forum to address the issue of the effects of aircraft noise on children. As a result of that forum, FICAN decided to sponsor this current study, which is based upon existing publicly available data. In brief, this study is designed to investigate the relation between (1) reduction in indoor classroom noise levels through airport closure or school sound insulation and (2) student academic performance, as measured by scores on state-standardized tests.

OVERVIEW

Research questions

This study concerns the relation between aircraft sound in classrooms and concurrent student test scores. More specifically, this study attempts to answer the following: Is aircraft noise reduction within classrooms related to test-score improvement, after controlling for demographics? Moreover, does this relationship vary by: age group (high, middle, and elementary school), student group (Individualized Education Program, IEP and non-IEP), or test type (verbal and math/science)?

Airports and schools

Aircraft sound within classrooms can change for many reasons. For adequate analysis, aircraft-sound changes needed to be relatively large in magnitude and not highly disruptive of the socio-economic environment. Three types of changes met these constraints: (1) the opening or closing of individual airport runways, (2) the opening or closing of entire airports, and (3) school sound insulation.

Three airports/states were chosen; the schools were in Illinois and Texas. Only public schools were chosen, as state-wide testing is mandatory only for students in public schools. Near these three airports, a total of 35 public schools experienced reduction in aircraft noise during the last ten years, due either to commercial-airport closure (20) or to school sound insulation (15). Of these 35 schools, three are high schools (grade 9 and higher), 13 are middle schools (grades 7 and 8), and 19 are elementary schools (grade 6 and lower).

These airports and schools are not guaranteed to be representative. For that reason, results of this study should not be used nationally without subsequent studies of additional airports and schools.

Standardized tests

This study used mandatory state-standardized tests, exclusively, as the measure of student performance. This was decided because standardized test results have become increasingly important in the U.S. in recent years; among other things, such tests help determine student class credit, student grade advancement, student graduation, school funding, and official school accreditation. In addition, detailed test results are all available publicly, either on the internet or from research divisions of the two state departments of education. Non-mandatory tests were excluded even for public schools, because self-selection by school or by individual students might bias the school's test results.

The study's database included 1-year and 2-year "lags" after noise reduction occurred. However, only the lag-1 values were analyzed – that is, noise reduction was only assessed after one year of noise-reduced schooling. For the tests in the analysis, three types of test scores were available and used in this study: percentage of students with the "worst" test grade, average numerical score, and percentage of students with the "best" test grade.

For most tests in most years, these scores were available separately for the two student groups of interest: IEP (learning disabled) and non-IEP. Average numerical score was available for fewer than half the tests.

Aircraft noise exposures

This study departs from most prior studies in the details of its major predictor variables – that is, its noise exposures. First, this study used *computed* noise exposures, rather than measured ones. Computation resulted in noise exposures that:

- Included each *entire* school year, rather than just sampled measurement periods during that year
- Included just the school months of each year, rather than the full year
- Included just school hours, rather than 24 hours
- Converted all computed noise exposures to *indoor* values, to account for outdoor-indoor noise reduction of school/window structures.

As a result, this study's noise exposures are potentially more closely linked with actual student noise exposure than in most prior studies.

ANALYSIS

Noise exposures

The major predictors of interest in this study concern before-after changes in cumulative noise exposure. Although contours of day-night sound levels (DNL) were available for each airport, such contours are influenced by early morning, evening and nighttime aircraft activity, and were not used. Instead, a series of noise exposures were developed – all for the 9-hour school day (7am to 4pm), and all inside the school classrooms.

In brief, outdoor school-hour metrics were computed – separately for each of the three airports in the study, and separately for each study. Outdoor SEL and LAmax for each aircraft flyover, at each school, were computed with the Federal Aviation Administration's Integrated Noise Model (INM) Version 6.1. As stated above, input was restricted to school hours (7am through 4pm) during the school year. Next, these outdoor sound levels were converted to indoor values and different noise exposures, using school-specific construction details. In brief, this process involved:

- Computation of outdoor-to-indoor level reduction (OILR), in octave bands, using construction details of individual schools
- Conversion of outdoor aircraft spectra (from INM) to indoor spectra, based upon the computed values of OILR
- Computation of the specific indoor cumulative noise exposure for the study.

For the relevant years and time periods, the following indoor cumulative noise exposures were computed:

- A-weighted noise exposures: Equivalent sound level, ALeq, the indoor equivalent sound level, averaged over the 9-hour school day (7am to 4pm).
- Speech Intelligibility Index (SII): Number of events disrupting indoor speech – for students in the back of the classroom, when the teacher (either gender) uses "raised voice" – ANEv<0.98SII (disrupts one percent of words).
- Speech Interference Level (SIL):
  - Number of events disrupting indoor speech – Articulation Index (AI) equals 0.50 for students in the back of the classroom, when the teacher (either gender) uses "raised voice", above 40 dBA, ANEv>40SIL
  - Fraction of indoor time speech is disrupted, above 40 dBA, AFnTm>40SIL

Multi-variate multilevel regression

To conclude that aircraft-sound change is associated with test-score change, the analysis must determine confidence intervals for any computed association. Such confidence intervals depend upon the various sources of variability in the data, and this variability exists at several levels. To account for these sources of data variability, multilevel regression was used for all analyses. Such regression is commonly used in educational studies, because the underlying data are commonly "nested" – schools sampled first, then test years, then classroom test scores. As a result of nested sampling, classroom test scores are not all statistically independent. Instead, they might tend to "cluster" by school and/or by test year. Multi-level analysis increases the statistical uncertainty to properly account for the reduced number of truly independent samples.

In all, regressions were performed for three score types: failure rate (percent of students with worst test score); average test score (scaled from 0 to 100); and top-score rate (percent of students with best test score), as well as for each of these four combinations of: age group (high, middle and elementary school); student group (IEP and non-IEP); and test type (verbal and math/science).

Aircraft noise-exposure change was mathematically compared to concurrent test-score change, to look for a potential relationship between the two. All changes were quantified over a one-year period.

Between one year and the next, a change in classroom noise exposure may influence standardized test scores. But demographic changes over the same time period may also influence these test scores. It is necessary to "control" for these demographic changes during the analysis. In that manner, only the proper portion of test-score change will be associated with noise-exposure change, and the remaining portion with these demographic variables. The relative portions will be determined mathematically in the analysis and will depend upon how strongly each variable relates to test-score change in the data.

As the primary method of demographic control, comparisons were made while holding "school" constant, as follows: (1) first the resulting regression equation was evaluated for all tests given in that school on the year after noise reduction, (2) then the same regression equation was evaluated for all tests given in prior school years (prior to noise reduction), and (3) finally, these two results were subtracted, to obtain the "effect" of noise reduction, controlled for results on non-noise-reduced (prior) years.

This method of demographic control works well because school demographics are not likely to change much from year to year. Their relative constancy is a great benefit to before-after studies of this type. This constancy means, to a first approximation, that these variables are automatically controlled in the analysis – by holding "school" constant from "before" to "after." With this demographic control, the study asks, "How much *different* is test-score change, before-to-after noise reduction, from test-score change at these same schools but when they were not concurrently experiencing noise reduction?"

As a result of the study's primary demographic control, "noise-reduction" and "control" groups automatically have the same demographics, at least over a ten-year average. Even so, possible year-to-year changes in demographics remain.

To explicitly control for year-to-year demographic changes (and also for each school's long-term average demographics), publicly available demographic data were collected from individual school records, state boards of education, and from the year-2000 census.

Table 1 contains the 24 demographic variables that were available in both Texas and Illinois. The table contains each variable's abbreviation in this study, its more complete definition, and whether it describes an entire school district or a specific school.

Table 1. Available demographic variables

Abbreviation in the study	Definition	Type: School district	Type: Specific school
DStTchExp	Teacher experience (years), average	x	
DStStntTchRat	Student-teacher ratio	x	
DStTchSal	Teacher salary ($), average	x	
DSt$PrStnt	School expenditure per student ($), average	x	
DSt%OwnOcc	% owner-occupied housing	x	
DSt%Pvty	% poverty (households)	x	
DSt%ChldPvty	% child poverty (under 18 years of age)	x	
DSt%NoSch	% adults with no schooling	x	
DSt%8orLess	% adults who finished 8th grade or less	x	
DSt%9to12	% adults with some high school education (9th through 12th grade)	x	
DSt%SmCollg	% adults with some college education	x	
DSt%GradDeg	% adults with graduate degrees	x	
DStHsVal	House value ($), representative	x	
DStHsInc	Household income ($), representative	x	
DStEnrl	Enrollment in the school		x
DSt%Attnd	% student attendance		x
DSt%LwInc	% low-income students		x
DSt%RcWht	% race, white		x
DSt%RcBlk	% race, black		x
DSt%RcHsp	% race, Hispanic		x
DSt%RcAsn	% race, Asian		x
DSt%RcNAm	% race, native American		x
DSt%LmtEng	% with limited English proficiency		x
DSt%Drpout	% drop out		x

The last variable in this table (percentage drop out) had many missing values in the database and was therefore dropped from the study, thereby leaving 23 demographic variables in the analysis. None of the other variables had any missing values, whatsoever.

Because of the large number of demographic variables, Principal Components Analysis was used to (1) condense the number of variables in the analysis from 23 to four principal components, and (2) guarantee that these four components are mutually independent. Each principal component is a linear combination of all 23 original variables, each with its own "factor coordinate" between plus 1 and minus 1. Where a demographic variable's factor coordinate is small (nearly zero), that variable is unimportant to that principal component. In all, the following principal components were identified and named:

- D1: Overall wealth and level of parental education
- D2: Spanish language
- D3: Socio-economic status
- D4: School-district size.

Several other predictor variables were included in the regression, to control for various factors:

- **Prior test score.** When a school class scores worse than average in a given year, it will most likely improve the following year. In essence, just by chance it will likely move towards its average performance the following year – which means upward. In statistical jargon, it will "regress towards its mean (average)." To control for this effect, each regression for a "change in test score" included as a predictor variable the prior year's actual test score, also. As a result, a portion of the change in test scores was ascribed to the prior year's test-score value. The regression coefficient for that prior year's test score thereby controlled for "regression towards the mean."
- **Prior noise exposure.** Each regression attempts to associate test-score change with noise-exposure change from "before" to "after" noise reduction. That association might be influenced by prior noise exposure, however. For example, whenever prior noise exposure is very low, then no test-score improvement can possibly be obtained from noise reduction. In that situation, the teaching/learning environment is not "sick," and therefore noise-reduction "medicine" cannot be expected to have any effect. To control for this potential effect, the prior year's noise exposure was added as a predictor in the regression.
- **Cause of an airport's noise reduction, combined with testing state (Illinois or Texas).** The three airports in this study involved two distinct causes of noise reduction (airport closing and school sound insulation) and tests within two different states (Texas and Illinois) – in all, three combinations of these two variables. To control for potential effects

of these distinctions, two additional dummy variables were added to each regression. The first of these applied to Texas schools that were sound insulated. The second applied to Texas schools that were near an airport closure. Then *neither* applied to schools near the Illinois airport. In all, the two dummy variables accounted for the three combinations of "cause" and "state."

- **Test-regime change within Illinois.** Illinois test regimes changed between 1998 and 2000. Some of the before/after test-score changes occurred simultaneously with these changes in test regime. For this reason, part of the test-score change might be more tightly associated with a change in the type of question or the method of scoring – and perhaps more tightly than with the change in noise exposure. To control for this possibility, a dummy variable tagged those particular before/after years in Illinois that involved test-regime change.

RESULTS

The key findings are the following:

- **Failure rate (all high-school students, both test types).** This study found substantial association between noise reduction and decrease in failure rate of high-school students. This improvement in test scores is essentially the same for all student/test subgroups. That substantial association was detected most "efficiently" when noise exposure was quantified as the percent time that the classroom L_A exceeded 40 dB. When that noise exposure decreased by 5 percentage points, the associated improvement was a substantial 20-percentage-point decrease in failure rate (with 99% certainty). This result was confirmed, though not as strongly, with the exposure called "any amount of change." In addition, it was confirmed for non-IEP students with the exposure called "number of events disrupting speech" reduced by 20. In fact for this subgroup, all tests show improvement in failure rate, and none show increased failure – further confirmation that improvement for failing high-school students is real.
- **Failure rate (all elementary and middle-school students, both test types).** This study found no substantial association between noise reduction and decrease in failure rate for elementary and middle-school students. All statistically significant tests show improvement (reduction in failure rate), but are very small in magnitude. Those "contrary" entries that show increased failure have extremely small confidences (44%, 39%, 4% and 0.1%) that the test-score change truly differs from zero.
- **Average test score (all subgroups).** This study also found significant association between noise reduction and average test scores, for all student/test subgroups. Measured by the percent time L_A was greater than 40 dB, all subgroups showed modest average-score improvement – between 7 and 9 percentage points, when this noise exposure decreased by 5 percentage points. In addition, when measured by the number of events with L_{Amax} greater than 40 dB, middle and elementary students showed modest average-score improvement – between 4 and 5 percentage points, when the number of such events decreased by 20. However, for high-school students reduction in the number of such events was associated with *poorer* average scores – between 17 and 19 percentage points.
- **Top-score rate (all subgroups).** This study found moderate association between noise reduction and change in top-score rates, mainly for IEP students on verbal tests. For those, a 5-point decrease in "percent time L_A was greater than 40 dB" was associated with *reduction* in the top-score rate by 5 percentage points.

Important caveats. The airports and schools in this study are not guaranteed to be representative. For that reason, results of this study should not be used nationally without subsequent studies of many additional airports and schools. In addition, this study's analysis is not yet fully reviewed.

Discussion

This study found substantial association between noise reduction and decrease in failure rates. Several mechanisms are possible for this association. Student failure may be due to impaired learning in the classroom, perhaps caused in part by noise stress. To the extent that noise stress contributes to student failure, then failing students are the ones most likely to benefit from noise reduction. In contrast, top-score students are less likely to benefit. Such a rationale is consistent with the results of this study.

In addition, this study found little distinction between test-score change and type of test: verbal or math/science. That finding is not consistent with past studies. However, to the extent that teacher-student communication is important to learning – for *all* academic subjects – then noise interruption of that communication would be detrimental to classroom learning, independent of the classroom subject (verbal or math/science).

The standardized tests used in this study are given to students in their classrooms – that is, they are tested in the "teaching" noise environment. As a result, a student's score might improve after noise reduction because either (1) the student learned more during the year (reduced chronic stress), or (2) the student was stressed less during the actual testing time (reduced acute stress). This study cannot distinguish between these two situations. Nevertheless, both are potentially serious impacts on students. Students who do not learn because classrooms are noisy will certainly suffer for lack of knowledge. In addition, students who *do* learn, but who *cannot prove their knowledge* during noisy tests, may suffer through lower grades, or not advancing to the next grade level, or not graduating from school.

Recommendations for future studies

The authors make the following recommendations for follow-up studies:

- ***Airports and schools.*** Include a larger number of airports and schools.
- ***Students.*** Follow individual students from year to year, rather than using only class-average results. Almost all of the statistical uncertainty in this study derived from test-to-test differences, where each test was a class average. This source of variability derives, in part, because different students take the test from year to year, at the same grade level. Instead, if scores of individual students were followed from grade to grade, such an analysis would intrinsically offer better precision.
- ***Testing location.*** Determine which tests were actually given in "teaching" classrooms and which were given elsewhere. Such knowledge would help distinguish between chronic and acute noise stress.
- ***Precision of noise computations.*** Obtain airport data directly from airports. Also incorporate actual outdoor-to-indoor measurements at each school.

In general, wherever these recommendations increase the amount of data, compared to this current study, they will increase the levels of confidence for all results. In addition, imprecise input always tends to partially reduce the numerical magnitude of (wash out) the associations found in regression analysis. It is likely this has occurred in the current study. Therefore, wherever these recommendations increase the precision of input data, they will tend to increase the numerical magnitude of all associations between noise reduction and test-score change.

REFERENCES

ANSI, 1977/86. American National Standards Institute. "Rating noise with respect to speech interference," ANSI S3.14-1977 (R-1986). Acoustical Society of America, New York.

ANSI, 1997. American National Standards Institute. "Methods of calculation of the Speech Intelligibility Index," ANSI S3.5-1997. Acoustical Society of America, New York.

Evans, 1998. Evans, G.W., Bullinger, M. and Hygge, S. "Chronic noise exposure and physiological response: A prospective study of children living under environmental stress." *Psychological Science*, 9, 75–77.

Goldstein, 2003. Goldstein, Harvey. *Multilevel Statistical Models, Third Edition*. Kendall's Library of Statistics, Arnold (a member of the Hodder Headline Group), London, 2003.

Hygge, 2003. Hygge, S., Boman, E. and Enmarker, I. "The effects of road traffic noise and meaningful irrelevant speech on different memory systems." *Scandinavian Journal of Psychology* **44**, 13–21.

MLwiN (software package for fitting multilevel models), Centre for Multilevel Modelling, University of Bristol, www.cmm.bristol.ac.uk/MLwiN/index.shtml.

Stansfield, 2000. Stansfeld, S.A., Haines, M.M, Head, J., Berry, B., Jiggins, M., Brentnall, S. and Roberts, R. "Aircraft noise at school and child performance and health initial results from the West London Schools Study." Proceedings of Internoise 2000.

Statistica (data analysis software system), Statsoft, Inc., www.statsoft.com

FICAN

Federal Interagency Committee on Aviation Noise

Mr. Alan F. Zusman, Chairman
Department of Navy
Washington Navy Yard
1322 Patterson Ave. SE
Washington, DC 20374-5065
Alan.Zusman@navy.mil