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

RELATION BETWEEN AIRCRAFT NOISE REDUCTION IN SCHOOLS AND STANDARDIZED TEST SCORES:

Mary Ellen Eagan[1], Grant Anderson[1] Bradley Nicholas[1],
Richard Horonjeff, Terry Tivnan[2]

BACKGROUND

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 (Evans, 1998). The strongest findings to date are in the area of reading, where the majority of studies have shown that children in noise impact zones are negatively affected by aircraft.

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.

STUDY OVERVIEW

Research questions. 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 (IEP[3] and non-IEP)?
- Test type (verbal and math/science)?

Standardized tests. This study uses state-standardized tests, exclusively, as the measure of student performance. This was decided because standardized test results are increasingly important in the United States in recent years. Among other things, such tests help determine student class credit, student grade advancement, student graduation, school funding, and school accreditation.

---

[1] Harris Miller Miller & Hanson Inc.

[2] Harvard University Graduate School of Education

[3] IEP means Individual Education Program.

Airports and schools. Three airports were chosen in states with publicly available test-score results in electronic format (only available over the last 10 years). Near these three airports, a total of thirty-two public schools have experienced reduction in aircraft noise during the last ten years, due either to airport closure or to school sound insulation. Non-public schools were excluded from the study, because they are not required to give state-standardized tests to all their students.

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 many additional airports and schools.

Time periods for computed noise exposure. For each of these schools, separately by year, aircraft noise was computed inside a typical classroom. Novel to this study are the time periods used to quantify aircraft noise. Compared to studies using pre-computed noise contours, this study:

- Used just school months, rather than the full year
- Used just school hours, rather than 24 hours
- Converted to indoors, to account for school/window structure.

As a result, this study's noise metrics are closely linked with actual student noise exposure. In addition, this study used the full school year to determine noise exposure, rather than just sampled measurement periods.

ANALYSIS METHOD

Computation of noise within classrooms. Year-by-year air traffic was obtained from a combination of information in Part 150 studies, the Official Airline Guide (OAG), and air-carrier aircraft inventories. From these input, outdoor SEL and L_{Amax} were computed with Version 6.1 of the FAA's Integrated Noise Model (INM). As stated above, input was restricted to typical school hours (7 A.M. through 4 P.M.) during the school year.

Next, these outdoor sound levels were converted to indoor values and different noise metrics, using INM aircraft spectra and school-construction details—average classroom dimensions, window types (before and after sound insulation), wall structure, and roof structure—for main school buildings and portable classrooms.

These computations resulted in the following noise metrics (school year, school hours, inside classrooms):

- School-day L_{Aeq}
- Percent of time $L_A > 40$ dB
- Number of events with $L_{Amax} > 40$ dB
- Number of events disrupting speech—that is, Speech Intelligibility Index (SII) < 0.98.

Demographic "control." To answer the research questions, aircraft-noise 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. These comparisons were made separately for:

- The "noise-reduction" group: Each school, before-to-after the year of noise reduction
- The "control" group: Same schools, but for all years prior to noise reduction.

By this arrangement, "noise-reduction" and "control" groups automatically have the same demographics (same schools mean same demographics) over a 10-year average. As a result, the mathematical question can be phrased as follows: How much <u>different</u> was test-score change, before-to-after noise reduction, than for schools not concurrently experiencing noise reduction?

Possible year-to-year changes in demographics remain, however. To explicitly control for these, demographic variables were collected for each school, from individual school data and year-2000 census data. Use of demographic variables in the analysis regression avoids associating test-score improvement with noise reduction, if test-score improvement is more strongly associated with demographics.

Analysis mathematics. Multi-level regression was used for all analyses. Such regression is commonly needed for educational studies, because the underlying data are commonly "nested"—schools sampled first, then test years, then test scores. As a result of nested sampling, student test scores are not all statistically independent. Instead, they might tend to "cluster" by school and/or by year. Multi-level analysis increases the computed statistical uncertainty, to properly account for fewer truly independent samples.

The basic analysis equation is:

$$\begin{array}{c}\text{change in}\\\text{test score}\end{array} = C_1 + C_2\begin{pmatrix}\text{change in}\\\text{noise}\end{pmatrix} + C_3\begin{pmatrix}\text{prior}\\\text{test score}\end{pmatrix} + C_4\begin{pmatrix}\text{prior}\\\text{noise dose}\end{pmatrix}$$
$$+ \text{terms for demographics}$$
$$+ \text{terms denoting various subgroups}$$
$$+ \text{"interaction" terms with } \begin{pmatrix}\text{change in}\\\text{noise}\end{pmatrix}.$$

In this equation, "change in noise" was measured separately by each of the four indoor noise metrics bulleted above. In addition, it was measured by a variable not dependent upon the noise computations—simply whether the school had noise reduction that year, or not.

In the regression, if the net effect of all coefficients involving "change in noise" is statistically significant, then a relationship exists between change in test scores and change in noise. In addition, this relationship exists while simultaneously controlling for demographics—that is, after the effects of demographics are essentially subtracted out.

Analysis combinations. 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)
- "A" rate: Percent of students with best test score,

and each of these for all combinations of:

- Age group: High, middle and elementary school
- Student group: IEP and non-IEP
- Test type: Verbal and math/science.

INITIAL RESULTS

Change in failure rate associated with noise reduction. Statistically significant results were obtained for the percentage of students who failed. This paper reports only those failure-rate results. Results for average test scores and percentage of "A" students were smaller in magnitude and less statistically certain.

Table 1 shows partial results of this study. In that table:

- The first column contains each of the student age groups, in turn.

- The second column contains the change in failure rate associated with noise reduction, separately tabulated for high, medium, and low scores "before" noise reduction occurred.
- The third column contains the computed confidence that the change is "real"—that is, "greater than zero."

Table 1 Change in failure rate associated with noise reduction: Verbal tests

Age group	Change in failure rate associated with noise reduction	Confidence that change is real
High	High before: 60% before –12% = 48% after	99.9%
	Med. before: 40% before –10% = 30% after	99 %
	Low before: 15% before – 7% = 8% after	< 90 %
Middle	High before: 60% before – 1% = 59% after	< 50 %
	Med. before: 40% before + 1% = 41% after	< 50 %
	Low before: 15% before + 4% = 19% after	< 90 %
Elem	High before: 60% before – 0% = 60% after	< 50 %
	Med. before: 40% before + 2% = 42% after	< 50 %
	Low before: 15% before + 5% = 20% after	90 %

Figure 1 shows these same results graphically, and expands them to other analysis subgroups.

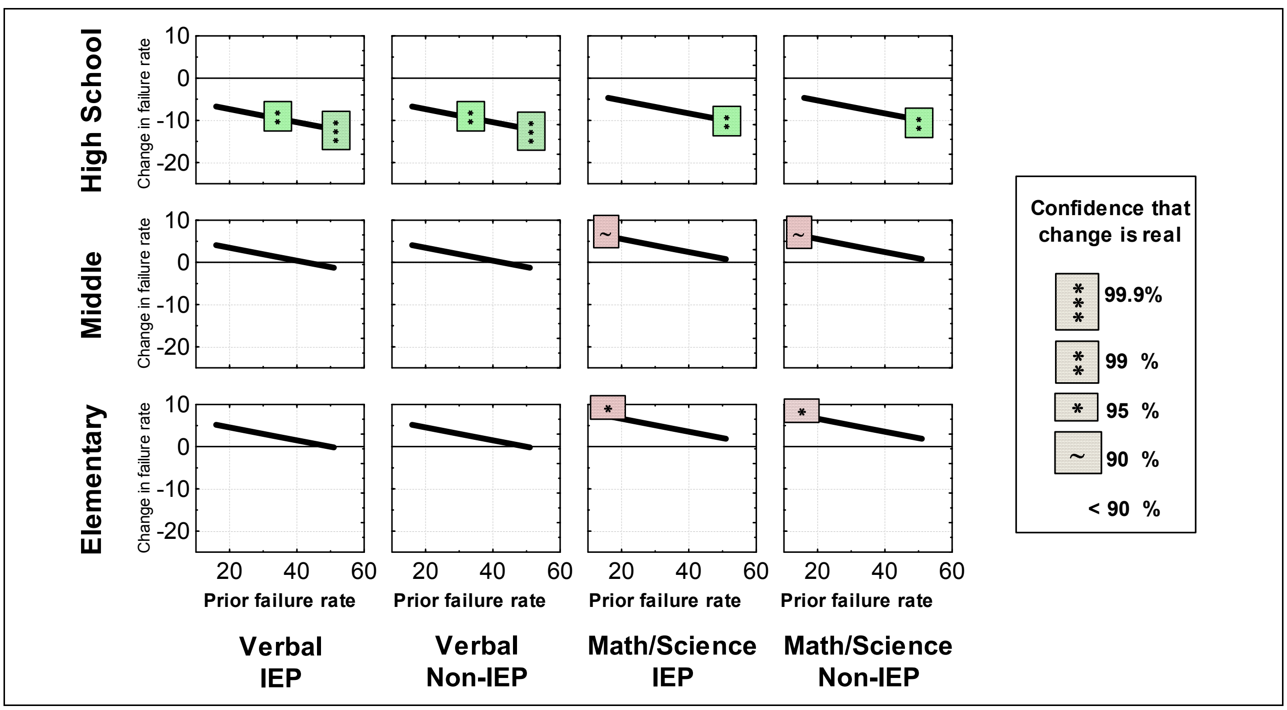

Figure 1 Change in failure rate associated with noise reduction

The figure contains three rows and four columns of small graphs. The rows show results for high, middle and elementary schools. The columns show these results separately for the two test types (verbal and math/science) and the two student groups (IEP and non-IEP). Plotted in each small graph is a line relating prior failure rate (horizontal axis) to the change in failure rate (vertical axis) after noise reduction. The plot therefore shows failure-rate change for all prior failure rates—not just high, medium and low prior failure rates, as in the table.

Confidence symbols appear within shaded boxes along each plotted line in the figure. The figure's key to the right shows the codes for these symbols, ranging from 99.9% confident to less than 90% confident. A box potentially appears at the left end of each line (low prior failure), the middle of that line, and the right end of that line (high prior failure). Missing boxes mean that the confidence is less than 90 percent.

Both Table 1 and the more-complete Figure 1 show substantial decreases in failure rates, for the "noise-reduction" group compared to the "control" group. This improvement depends upon the failure rate just before noise reduction. In the table's first row, for example, high-school failure rates of 60% (just before noise reduction) decreased to 48% after noise reduction—an improvement of 12. For lower prior failure rates (second row and third rows), improvement was less but still substantial. Note that for lower prior failure rates, there is less room for improvement.

For middle and elementary schools, the table and the figure both show some moderate increases in failure rate after noise reduction. These increases are smaller in magnitude than the high-school decreases. In addition, they are less confidently known.

The four graph columns in Figure 1 show no distinction between IEP and non-IEP students, and little distinction between verbal and math/science tests.

Change in failure rate when %Tm > 40dBA drops by 5 (like 7% to 2%). Figure 2 also shows substantial changes in failure rates—for the following specific change in noise dose: A drop of 5 in the percent time aircraft noise is greater than 40 dBA. In other words, the failure-rate changes in this figure are for a specific amount of noise reduction, in contrast to Figure 1.

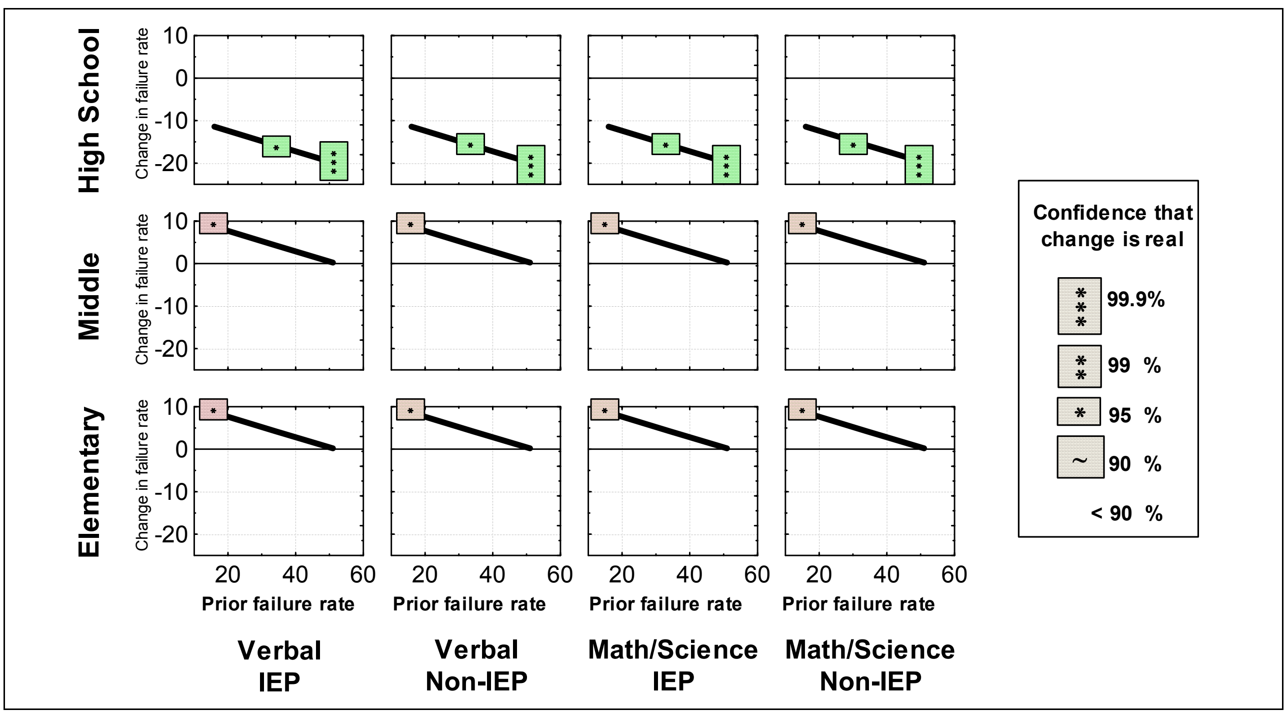

Figure 2 Change in failure rate when %Tm > 40dBA drops by 5 (like 7% to 2%)

Failure-rate change in Figure 2 again depends upon the failure rate just prior to noise reduction. In the figure's upper-left graph, for example, a prior failure rate of 50% decreases by 20%, for an "after" rate of 30%. When measured by this noise metric, changes in failure rate apply equally to verbal and math/science tests, and equally to IEP and non-IEP students.

Technical comments. When an analysis examines multiple subgroups of data, the criteria for confidence must be tightened. With subgroups, instead of desiring 95% confidence about the data as a whole, desired is 95% confidence about each and every one of the separate subgroups—a much stricter standard. With 12 subgroups, for example, that stricter standard requires 99.6% confidence about each subgroup. So when the computer reports 99.6% confidence, that value must be mathematically diluted to 95% confidence. Such confidence-level dilution has been done throughout this analysis.

Another matter: When a 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)." This study found regression towards the mean in all analyses, but then subtracted that effect out. As a result, the improvements cited here are additional improvements—"noise reduction" versus "control"—beyond regression towards the mean. Therefore, this additional improvement is truly associated with noise reduction.

SUMMARY AND DISCUSSION OF RESULTS

Summary of all results. In summary, this study:

- Found substantial association between noise reduction and decrease in failure rates, only for high-school students.
- Found some weaker association between noise reduction and increase in failure rates, for middle and elementary schools.
- Found little distinction between IEP and non-IEP students, and between verbal and math/science tests.
- Found little association between noise reduction and changes in "A" rate or average scores.

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 validated and reviewed.

Speculative thoughts about the results. 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, "A" 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 test type.

Potential limitation of the methodology. 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 cases. Nevertheless, both are potentially serious impacts on students. In particular, students who learn but cannot prove their knowledge during noisy tests, may suffer through lower grades, not advancing to the next grade level, or not graduating from school.

RECOMMENDATIONS FOR ANY FOLLOW-UP STUDIES

The authors make the following recommendations for any 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—perhaps in a quieter environment. Such knowledge would help distinguish between chronic and acute noise stress.
- ***Portable classrooms.*** In addition, identify classes taught in portable classrooms. Aircraft noise is louder in portable classrooms, because they do not insulate as well against aircraft noise. Knowing which classes—or better, which students—are taught in these portable classrooms will improve aircraft-noise estimates.
- ***Precision of noise computations.*** Obtain airport data directly from airports. Also incorporate 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 that has occurred in this 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.