Relation between aircraft noise reduction in schools and standardized test scores

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

1 Harris Miller Miller & Hanson Inc., 77 South Bedford St., Burlington, MA

2 Harvard University Graduate School of Education, Appian Way, Cambridge, MA 02138

* corresponding author: e-mail: meagan@hmmh.com

INTRODUCTION

Research on the effects of aircraft noise on children's learning suggests that aircraft noise can interfere with learning in the following school areas: reading, motivation, language and speech acquisition, and memory (Evans et al. 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 LAeq is 65 dB or higher (Stansfeld et al. 2000). It is also possible that, for the same outdoor LAeq, the effects of aircraft noise on classroom learning may be greater than the effects of road and railroad noise (Hygge et al. 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.

METHODS

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; student group; or test type?

Airports and schools

Aircraft sound within classrooms can change for many reasons. For adequate analysis in this study, 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 commercial airports, and (3) school sound insulation.

The following three airports/states met these constraints and were therefore chosen for this study: Airport 1: One airport in Texas (airport closing); Airport 2: Another airport in Texas (school sound insulation); and Airport 3: One airport in Illinois (school sound insulation). Only public schools were chosen for this study, because state-wide testing in the U.S. is mandatory only for students in public schools. Near these three airports, a total of 35 public schools have experienced reduction in aircraft noise during the last ten years, due either to commercial-airport closure or to school sound insulation. In particular:

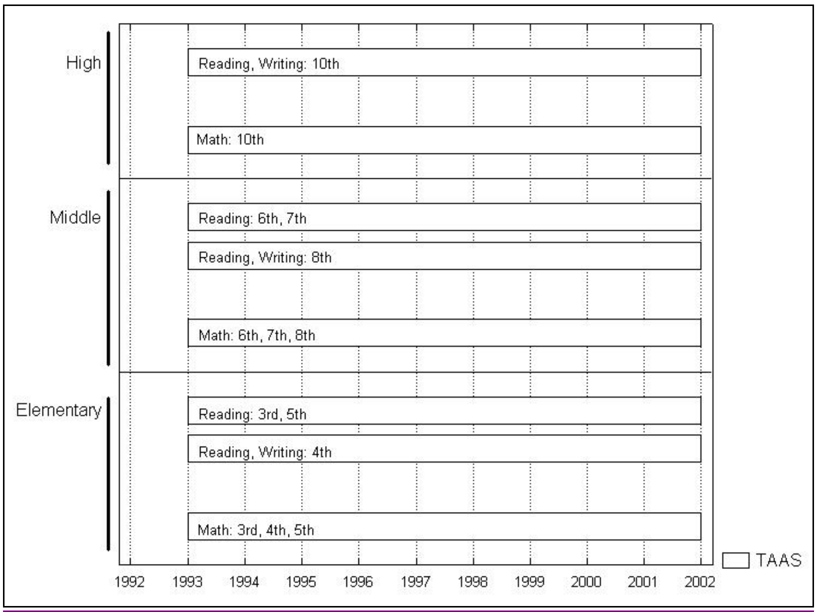
- 20 public schools near Airport 1—all those within the DNL 65 dB contour, plus a 2,000-foot buffer outside this contour
- 4 public schools near Airport 2—all those that were sound-insulated the summer of 1994 or later
- 11 public schools near Airport 3—all those that were insulated the summer of 1995 or later.

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

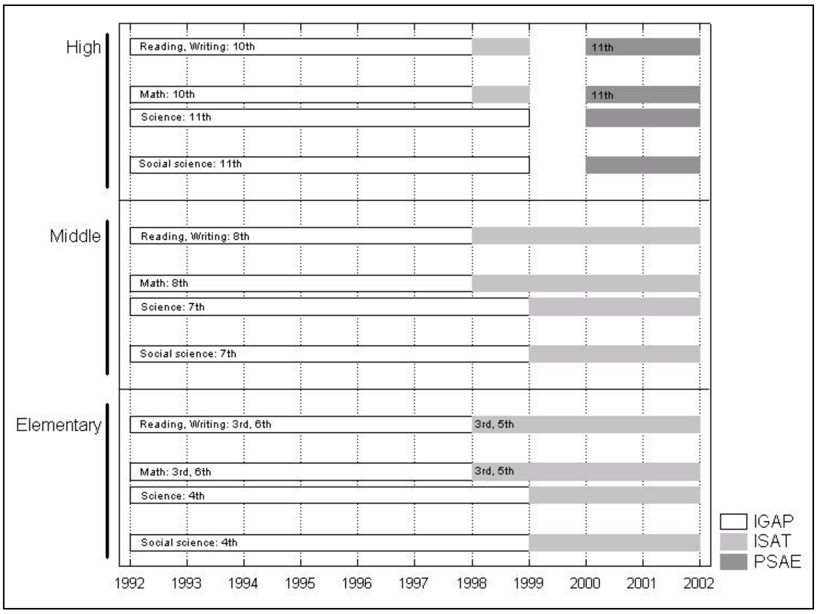
This study used mandatory state-standardized tests, exclusively, as the measure of student performance. This was decided because standardized test results have become increasingly important in the U.S. in recent years. Among other things, such tests help determine student class credit, student grade advancement, student graduation, school funding, and official school accreditation.

Figure 1 and Figure 2 show the available standardized tests for this study. Separately for high, middle and elementary schools, these two figures show the tested subjects, grade levels and school years of testing. The abbreviations denote different test regimes within each state, as shown below each figure. All these tests are mandatory in their state—all public schools, all students. In addition, their detailed test results are all available publicly, either on the internet or from research divisions of the two state departments of education.



TAAS: Texas Assessment of Academic Skills

Figure 1: Texas Standardized Tests



IGAP: Illinois Goal Assessment Program
ISAT: Illinois Standards Achievement Tests
PSAE: Prairie State Achievement Examination

Figure 2: Illinois Standardized Tests

For the horizontal bars in these two figures, each school year extends between its "start year" and its "end year." In this study, school years are numbered by the "year test given," the same way graduating classes are named; for example, the class of 2,000 graduates in June 2000 (after taking the year-2000 tests. The study's database included 1-year and 2-year "lags" after noise reduction occurred. However, only the lag-1 values were evaluated—that is, noise reduction was only assessed after one year of noise-reduced schooling.

For the tests in these two figures, 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; and converted all computed noise exposures to indoor values, to account for school/window structure. As a result, this study's noise exposures are potentially more closely linked with actual student noise exposure than in most prior studies.

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 too influenced by early morning, evening and nighttime aircraft activity to be of use in this study. Instead, a series of noise exposures were desired—all for the 9-hour school day (7am to 4pm), and all inside the school classrooms. This section describes the noise exposures of this study.

At each school location, the FAA's Integrated Noise Model (INM) was used to compute the following noise metrics that are relevant to this study: SEL (Sound Exposure Level) for each aircraft flyover; and LAmax (maximum A-weighted sound level) for aircraft flyover. Since the INM computes only *outdoor* aircraft noise, computation proceeded in two steps: (1) INM computation of outdoor metrics; and (2) conversion from these outdoor metrics to the desired indoor cumulative noise exposures.

First, outdoor school-hour metrics were computed, separately for each of the three airports in the study, and for each study year, using the INM (version 6.1) and standard airport noise modeling techniques. Annual traffic levels were adjusted based on airline schedules and other sources.

Next, these outdoor sound levels were converted to indoor values and different noise exposures, using school-specific construction details and proprietary software. 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.
- Resulting indoor cumulative noise exposures

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 (7 am to 4 pm)
  - Number of aircraft events with indoor LAmax greater than three candidate thresholds: ANEv>35 dBA; ANEv>40 dBA; and ANEv>45 dBA
  - Fraction of time with indoor LA greater than three candidate thresholds: AFnTm>35 dBA; AFnTm>40 dBA; and AFnTm>45 dBA.
- Speech Intelligibility Index (SII):
  - Number of events disrupting indoor speech—for students in the back of the classroom, when the teacher uses "raised voice"—per three candidate thresholds:
  - ANEv<0.80SII (disrupts five percent of words)
  - ANEv<0.90SII (disrupts three percent of words)
  - 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"—per three candidate thresholds: ANEv>35SIL; ANEv>40SIL; and ANEv>45SIL
  - Fraction of indoor time speech is disrupted, per three candidate thresholds:AFnTm>35SIL; AFnTm>40SIL; and AFnTm>45SIL.

Among these noise exposures, only the following were advanced into analysis: LAeq; AFnTm>40 dBA; ANEv>40 dBA; and ANEv<0.98SII.

Multi-variate multilevel regression

Multilevel regression was used for all analyses. The basic analysis equation for this study is:

$$\begin{aligned}\begin{matrix}\text{change in}\\ \text{test score}\end{matrix} = C_1 + C_2\begin{pmatrix}\text{change in}\\ \text{noise exposure}\end{pmatrix} + C_3\begin{pmatrix}\text{prior}\\ \text{test score}\end{pmatrix} + C_4\begin{pmatrix}\text{prior}\\ \text{noise exposure}\end{pmatrix}\\ + \text{four terms defining the various subgroups}\\ + \begin{pmatrix}\text{change in}\\ \text{four principal demographic components}\end{pmatrix}\\ + \begin{pmatrix}\text{prior values of}\\ \text{four principal demographic components}\end{pmatrix}\\ + \text{three terms defining "cause," "state" and "test-regime change"}\\ + \text{"interaction" terms of } \textit{every} \text{ variable with } \begin{pmatrix}\text{change in}\\ \text{noise exposure}\end{pmatrix}.\end{aligned} \tag{1}$$

In this equation, “change in noise exposure” was measured separately with each of the study’s four cumulative noise exposures. In addition, it was measured by a variable (*QQuiet*) 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 exposure. In addition, this relationship exists while simultaneously controlling for (1) demographics, (2) the cause of noise reduction, (3) the specific state, and (4) test-regime change (in Illinois). Wherever the regression associated with these control variables is stronger than with “change in noise exposure,” then the regression ascribes more association to them than to noise. In that way, the regression subtracts out their effect, when it predominates, rather than ascribing that effect to the change in noise exposure.

In all, regressions were performed for three score types: (1) Failure rate: Percent of students with worst test score, (2) Average test score (scaled from 0 to 100), and (3) Top-score rate: Percent of students with best test score; for all combinations of: age group (high, middle and elementary school), student group (IEP and non-IEP), and test type (verbal and math/science). For these conditions, an initial regression involved all possible predictor terms, while the final regression involved only those predictor terms that proved statistically significant in the initial regression. Numerically, terms were deemed statistically significant (retained) if their initial-regression standard uncertainties were smaller than their values. However, if an interaction term was retained per this test, then its “parent” term was also retained, no matter how large its own standard uncertainty.

Primary demographic “control”

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); (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 <u>different</u> 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?"

Supplemental demographic "control"

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. 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.

Table 1: Available demographic variables common to Texas and Illinois

Abbreviation in the study	Definition	Type	
		School district	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 8^{th} grade or less	x	
DSt%9to12	% adults with some high school education (9^{th} through 12^{th} grade)	x	

Abbreviation in the study	Definition	Type	
		School district	Specific school
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

This many demographic variables cause two problems in the analysis: first, their sheer number greatly increases the complexity of the analysis regressions; and second, their unavoidable correlation causes ambiguous regression results, due to "confounding" among the regression variables. To eliminate both difficulties, Principal Components Analysis was used to simultaneously (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.

These principal components enter Eq. (1) above, in two ways—as prior year's values and as the before/after change in value—with a separate coefficient for each.

Additional regression terms

Several other predictor variables were included in Eq. (1) above, to control for various nuisance factors:

- Prior test score. When a school class scores worse than average in a given year, it will most likely improve the following year, or "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.
- 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, when-

ever prior noise exposure is very low, then no test-score improvement can possibly be obtained from noise reduction. 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. As shown in Figure 2 above 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

Regression coefficients were combined, as appropriate, for various student subgroups—for example, IEP elementary-school students taking verbal tests. After these are summed, their respective uncertainties combine in the standard manner, which takes into account their individual standard uncertainties and the covariances among these standard uncertainties. Combining terms and their uncertainties in this manner yields Table 2. In the table, verbal tests are reported separately from math/science tests. Within each of these two categories, the three score types appear in the first column of each table. In generating this table, average values were used for each variable in the regression, where these averages were computed specifically for the relevant subgroup being computed.

The various student subgroups (combinations of IEP and high/mid/elementary schools) are shown in the right set of six columns. For each subgroup, the table contains five numerical results—one result for each of the cumulative noise exposures in the second column. Rather than simple regression coefficients, the tabulated results consist of the expected test-score change for a particular noise-exposure change. For example, the third table entry for IEP high-school students is equal to –20. This is the change in failure rate (20 percentage points fewer failures) for a 5-point reduction in the percent time that indoor aircraft noise (L_A) is greater than 40 dB. In other words, a 5-percentage-point reduction in loud aircraft sound (those greater than 40 dB) is associated with a 20-percentage-point reduction in failure rate (an improvement in performance).

Table 2: Study results

Verbal Tests

Score type	For this amount of change in classroom noise (due to sound insulation or airport closure)	The study associates this amount of change in test score (percentage points)					
		IEP			Non-IEP		
		High N = 24	Mid N = 49	Elem N = 65	High N = 36	Mid N = 165	Elem N = 589
Failure rate	Any amount of change	**-12 *****	-1	0	**-12 *****	-1	0
	School-day L_{eq} down by 10 decibels	-1	-1	-1	-1	-1	-1
	Percent time L_A > 40 dB down by 5 points	**-20 ****	0	0	**-20 ****	0	0
	Number events L_{Amax} > 40 dB down by 20	-2 **	-2 **	-2 **	-2 **	-2 **	-2 **
	Number events disrupting speech down by 20	-7	4	3	**-14 ***	-3	-3
Average score	Any amount of change	Note 3					
	School-day L_{eq} down by 10 decibels	-6	13 ~	12	-6	13 ~	12
	Percent time L_A > 40 dB down by 5 points	**7 ***	**9 ****	**7 ***	**7 ***	**9 ****	**7 ***
	Number events L_{Amax} > 40 dB down by 20	**-18 *****	**4 *****	**4 *****	**-18 *****	**4 *****	**4 *****
	Number events disrupting speech down by 20	-4	-2 *	-3 ***	-4	-2 *	-3 ***
Top-score rate	Any amount of change	-3 ***	-3 ***	-3 ***	-2 ***	-2 ***	-2 ***
	School-day L_{eq} down by 10 decibels	**-4 *****	-1	-1	-3 ***	0	0
	Percent time L_A > 40 dB down by 5 points	**-5 ***	**-5 ***	**-5 ***	-2	-2	-2
	Number events L_{Amax} > 40 dB down by 20	-2	-2 *	-2 *	-2	-1	-1
	Number events disrupting speech down by 20	-5 ~	0	-2	-5 ~	0	-2

Math/Science Tests

Score type	For this amount of change in classroom noise (due to sound insulation or airport closure)	The study associates this amount of change in test score (percentage points)					
		IEP			Non-IEP		
		High N = 12	Mid N = 32	Elem N = 78	High N = 20	Mid N = 110	Elem N = 421
Failure rate	Any amount of change	**-10 ****	1	2	**-10 ****	1	2
	School-day L_{eq} down by 10 decibels	0	0	0	0	0	0
	Percent time L_A > 40 dB down by 5 points	**-20 ****	0	0	**-20 ****	0	0
	Number events L_{Amax} > 40 dB down by 20	-2 **	-2 **	-2 **	-2 **	-2 **	-2 **
	Number events disrupting speech down by 20	-7	4	3	**-14 ***	-3	-3
Average score	Any amount of change	Note 3					
	School-day L_{eq} down by 10 decibels	-7	12 ~	11	-7	12 ~	11
	Percent time L_A > 40 dB down by 5 points	**7 ***	**9 ****	**7 ***	**7 ***	**9 ****	**7 ***
	Number events L_{Amax} > 40 dB down by 20	**-17 *****	**5 *****	**4 *****	**-17 *****	**5 *****	**4 *****
	Number events disrupting speech down by 20	-4	-2 *	-3 ***	-4	-2 *	-3 ***
Top-score rate	Any amount of change	-3 ***	-3 ***	-3 ***	-2 ***	-2 ***	-2 ***
	School-day L_{eq} down by 10 decibels	**-4 *****	-1	-1	-3 ***	0	0
	Percent time L_A > 40 dB down by 5 points	-2	-2	-2	1	1	1
	Number events L_{Amax} > 40 dB down by 20	-2	-2 *	-2 *	-2	-1	-1
	Number events disrupting speech down by 20	-5	2	0	-5	2	0

Note 1. Cells are shaded when both (1) the test-score change is 4 or more, and (2) that change is more than 95% certain.
Note 2. Darker shading means test-score change is for the better. Lighter shading means test-score change is for the worse.
Note 3. These two regressions did not converge.

*** means more than 99.9% certain.
** means more than 99% certain.
* means more than 95% certain.
~ means more than 90% certain.
means less than 90% certain.

The asterisks in the table show the statistical confidence of individual results (see the table's footer). All entries with asterisks are statistically significant (95 % confidence or better). Shaded in the table are all values (1) greater than four and (2) statistically significant. The darker shadings mean that test-score change is for the better. In contrast, lighter shading means test-score change is for the worse—for example, a decrease in average test score after noise is reduced. The values of N in the column headings are the number of tested classes that contribute to each category.

Combined uncertainties

Note in Table 2 that several pairs of entries are identical (e.g., the two shaded entries of –12 in the table's first row). The numerical equality of these two entries means that these changes in high-school scores do not depend upon the IEP variable—that is, whether or not the student had an IEP. Notice also that several pairs of entries are identical between the upper and the lower table (e.g., the shaded entry of –20 in the third row of both the upper and lower table). Their numerical equality means that these changes in high-school scores do not depend upon the type of test: verbal or math/science.

In this study, many measures of test-score change were separately analyzed: different academic subjects, different student grade levels, and different percentiles for a given test. If each of these were to be analyzed with only 95 % certainty, it is quite likely that one or another of these analyses might appear statistically significant, just by chance alone. In brief, 95 % certainty allows a 5 % chance (1 out of 20) of apparent certainty, just by chance alone. So with 20 separate analyses, we would actually expect one to appear statistically certain. To guard against such mistaken certainty, this study analyzed to a tighter certainty than 95 %. The analysis was determined from the number of independent analyses. Whenever a regression examined multiple subgroups of data, the criteria for confidence was therefore tightened. With subgroups, instead of desiring 95 % confidence for the data as a whole, desired was 95 % confidence for each and every one of the separate subgroups—a much stricter standard. With twelve subgroups, for example, that stricter standard requires 99.6 % confidence for each subgroup. So when the regression mathematics reports 99.6 % confidence, that value must be mathematically diluted to 95 % confidence. Such confidence-level dilution has been done throughout this analysis. As a result, the confidence values in Table 2 incorporate this mathematical dilution, thereby becoming more stringent than without such dilution for multiple tests.

Summary of all results

The results of Table 2 above suggest:

- 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. The association was detected most "efficiently" when noise exposure was quantified as the percent time that the classroom LA 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 table entries 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 table entries do 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 LA 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 meas-

ured by the number of events with LAmax 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 LA was greater than 40 dB" was associated with reduction in the top-score rate by 5 percentage points.

CONCLUSIONS

This study found substantial association between noise reduction and decrease in failure rates on standardized tests for low-performing students. 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 then noise interruption of that communication would be detrimental to classroom learning, independent of the classroom subject (verbal or math/science).

Potential limitations of the methodology

The standardized tests used in this study are given to students in their classrooms, and as a result potentially measure both acute and chronic noise exposure. Thus, 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). Although this study cannot distinguish between these two situations, 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.
- 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.
- 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, American National Standards Institute (1977/86). Rating noise with respect to speech interference. ANSI S3.14-1977 (Rev. 1986). New York: Acoustical Society of America.

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

Evans GW, Bullinger M, Hygge S (1998). Chronic noise exposure and physiological response: a prospective study of children living under environmental stress. Psychol Sci 9: 75–77.

Goldstein H (2003). Multilevel statistical models. 3rd ed. London: Kendall's Library of Statistics, Arnold.

Hygge S, Boman E, Enmarker I (2003). The effects of road traffic noise and meaningful irrelevant speech on different memory systems. Scand J Psychol 44: 13–21.

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

Stansfeld SA, Haines MM, Head J, Berry B, Jiggins M, Brentnall S, Roberts R (2000). Aircraft noise at school and child performance and health initial results from the West London Schools Study. In: Proceedings of Inter-Noise.

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