Internal Consistency and Stability of Measurements of Community Reaction to Noise

R. F. S. Job

Correlations between transportation noise exposure and community reaction indicate that on average only 18 to 22 percent (depending on the type of transportation noise) of the variation in reaction is attributable to noise exposure. Given the importance of these studies in determining not only the acceptable levels of noise but also how the noise is measured (equal energy units or not; with or without time of day weightings, etc.), the impact of reliability of measurement of community reaction and how reliability might be improved are considered. It is often suggested that error in measurement of reaction is a major reason for the relatively low noise-reaction correlations observed. However, few data exist on the internal consistency of composite scales of reaction. Data are presented on the internal consistency of various reaction scales, indicating that the typical general reaction scale and annoyance scales are reliable (in terms of internal consistency) and superior to the typical disturbance and complaint disposition scales. Further, reliability is increased by the use of several questions in a scale, rather than a single question. Thus, the data suggest that the best measure of community reaction is a composite general reaction scale (based on questions such as, "How much are you personally affected . . . "; and "How disturbed are you by . . . ") with a number of contributing questions. Noise-reaction correlations would be increased moderately with the use of an extremely reliable measure of reaction. However, when the limited reliability of the measurement of reaction and noise exposure, and modifying variables (especially attitude and sensitivity) are taken into account, the proportion of variation in reaction left unaccounted for is only around 20 percent.

Since the 1950s, community reaction to noise has been studied through the collection of noise exposure data and self-reported reaction data from residents around the noise source of interest. Such studies serve important purposes by attempting to identify the factors influencing community reaction to noise, most obviously including the noise itself. Such information has a number of consequences: it is relevant to basic theory of psychophysics, of annoyance in general, and of annoyance caused by noise in particular (*1,2*). Because the ultimate aim of these studies and subsequent countermeasures is not reduction of noise but reduction of reaction (annoyance, disturbance, etc.), the precise noise-reaction relationship is central to issues of acceptable noise level, compensation for residents exposed to noise, land-use planning, noise insulation, and adjustments to noise at the source.

As a simple, perhaps extreme, example, one of the theoretical and practical issues of concern is the value of the decibel equivalent number (k), which still receives considerable research attention (*3,4*). Studies have suggested values of k that best predict reaction, ranging from below 0 to 25 or 35 (*4*). High values of k would suggest that adding low-decibel noises would increase community reaction, whereas low values of k would suggest that such additions of low-decibel noise would decrease reaction. If this were true, one way of reducing reaction would be to add noise, not decrease it. If the number factor is irrelevant or trivial, many quieter noises are better, but if the number factor is large, fewer louder noises may have less impact on reaction, and if equal energy units provide the best prediction of reaction [again, a point of contention (*4–6*)], then overall energy reduction is the most likely beneficial option. Clearly then, the practical and theoretical implications of k are profound.

Similar arguments may be made for many features of noise (frequency range, time of day, duration, etc.). Consequently, considerable care with noise-reaction studies is justified. However, such studies have indicated correlations between the noise exposure and the reaction of individuals to the noise, ranging from 0.19 (*7*) to 0.64 (*8*), with a recently reviewed mean of 0.42 (*9*). This suggests that only 4 to 41 percent of the variation in individual reaction is accounted for by noise exposure. Although the average correlations are slightly higher for transportation noise than for impulsive noise [mean r = 0.42, 0.46, 0.46, 0.28 for road, aircraft, rail, and impulsive noise, respectively (*9*)], the percentages of variance in community reaction accounted for by noise still only average 17.6, 21.2, and 21.2 percent for road, rail, and aircraft noise exposure, respectively. Such low correlations in these critical studies deserve explanation. There are three principal explanations for these results:

1. The relationship between noise exposure and reaction is, in reality, not strong;
2. The noise measurement techniques or indices used have been inadequate; and
3. The reaction measurement techniques or indices have been inadequate.

Most socioacoustic surveys have emphasized the search for noise indices to more accurately predict human reaction, rather than examining many possible measures of reaction itself. For example, Fields and Walker (*10*) examined 44 indices, Bradley and Johan (*11*) examined 25 indices, and Bullen et al. (*12*) examined 88 indices. The examination of such large numbers of noise indices is justified by the common aim of testing the indices against one another, and the need for such indices for

Department of Psychology, University of Sydney, Sydney, New South Wales 2006, Australia.

legislation and land-use planning. However, with some exceptions (*13–15*), this concentration has resulted in a lack of research directed at obtaining more accurate measurement of reaction. In most surveys, it is likely that reaction measurements contribute a significant amount of error—more than is contributed by noise measurements as identified by computations of reliability. Therefore, the issue of reliability of measurement of community reaction to noise is addressed in the following actions.

TEST-RETEST RELIABILITY

Following the classic literature, Hall and Taylor (*15*) have distinguished two meanings subsumed under the general label of reliability: internal consistency and stability. Stability refers to the change in the true magnitude of the characteristic being measured, whereas internal consistency refers to the consistency of the items used to construct a composite scale. Of course, stability in the sense of change in true magnitude is not directly measurable. Rather, differences in repeated measurements at different times are contributed to by changes in the true score and changes in the error components of the observed score. Ultimately, in terms of measurement, the difference between stability and internal consistency becomes this: stability is measured by repeating exactly the same questions to the same respondents with a significant time interval—usually several months to 1 year in studies of noise reaction stability. Internal consistency is measured by asking nonidentical questions, designed to measure the same characteristic of the same respondent over a very short interval—within the one interview of typically 30 min to 1 hr.

However, measures of stability may be confounded by error components. For example, the supposedly identical questions may be put slightly differently by the interviewer on the two occasions: voice intonation, eye contact, and nonverbal cues from the interviewer are known to influence survey outcomes (*16–19*). If these effects are conceived of as constituting slight changes in the question, then part of the variation observed as stability is actually a matter of internal consistency. Furthermore, it is possible that some of the variation observed in measurements of internal consistency really reflects changes of stability over the brief interval of the interview. In fact, the conduct of the interview itself may point out many possible effects of the noise not previously identified all at once. For example, for questions on disturbances to sleep, watching television, conversation, telephone calls, symptoms such as restlessness, nerves, and headaches may occasion a change in the true component of reaction. Accordingly, the theoretical distinction between stability and internal consistency may be somewhat blurred at the point of measurement. Nonetheless, the distinction is recognized as being of theoretical and practical value.

Four studies have been reported on the stability of social survey measures of reaction to noise across time (*11,15,20,21*). Despite substantial differences in sample sizes and the scales of reaction examined, the results are similar. Scales based on single items achieved the following reliabilities: (a) seven-point, semantic differential scale of dissatisfaction with traffic noise, 0.642 (*22*), 0.61 (*20*); (b) four-point, fully labeled unipolar scale of the extent to which the person is bothered by traffic noise, 0.63 (*21*); (c) nine-point, bipolar fully labeled scale from extremely agreeable to extremely disturbing, 0.53 (aircraft noise), 0.58 (traffic noise), 0.30 (overall noise) (*15*); and (d) eleven-point (0 to 10), unipolar scale with endpoints labeled "Not at all Disturbed" and "Unbearably Disturbed", 0.50 (aircraft noise), 0.58 (traffic noise), and 0.26 (overall noise) (*15*). Hall and Taylor (*15*) noted that the lower correlations for aircraft and overall noise compared with road noise may be because of changes in the noise environment. Therefore, the results for traffic may be taken as the least ambiguous measure of test-retest reliability in this study. Thus, over the different measurement scales, reliability varied over a very narrow range (0.58 to 0.64).

Bradley and Johan (*11*) reported data on the test-retest reliability of composite scales of reaction to traffic noise. The annoyance scale achieved a time-separated test-retest reliability (stability) of 0.75, while for the interference scale the stability was 0.81. Consistent with Hall and Taylor's (*15*) claim that little data existed on the internal consistency of reaction scales, Bradley and Johan's study is the only one that bears on internal consistency, albeit with a comparatively small sample size. The reaction measurement was based on analysis of 80 response items. The alpha coefficient, a measure of internal consistency of a multi-item scale (*22*), was 0.95 for the annoyance scale and 0.94 for the interference scale.

In consideration of this dearth of data, the first aim is to present analyses of data available from existing surveys to provide estimates of internal consistency of various reaction scales. The impact of the use of composite scales and single item scales of reaction on test-retest reliability is also examined. Finally, the effect of increases in reliability on the amount of variation in reaction that is accounted for (in terms of noise exposure and modifying variables) is examined.

The last purpose is relevant to consideration of the scientific and practical consequences of the small percentage of variation in reaction accounted for by noise exposure. From a practical point of view, the low correlations have led to criticism of the populations studied (*23*) and doubts about the environmental noise standards derived from these studies (*24*). Furthermore, the selection of the indices of noise most predictive of community reaction is hampered by the inaccuracy of reaction measures, which reduces the opportunity to distinguish between noise indices (*25*). From the viewpoint of basic science as well as practical application, research may be motivated by the need to account for the variation in reaction left unexplained by noise exposure. Thus, numerous other factors such as personality, various measures of socioeconomic status, sex, age, marital status, noise sensitivity, and attitude toward the noise source have all been examined. Langdon (*25*) has suggested that this may not be a problem. That is, the failure of noise exposure to account for more than a small percentage of the variation in reaction may reflect unreliability of the measures used. The present paper tests this possibility.

ANALYSIS OF EXISTING SURVEYS

Several studies by the socioacoustic research team of the Australian National Acoustic Laboratories have been conducted along similar lines, although a range of noise sources has been investigated: artillery noise, rifle range noise, and noise from civilian and military aviation. In all cases, respondents have been asked a large number of questions to assess their reactions. In all cases, a composite scale of overall reaction has been calculated using factor analyses and regression. The studies shared several scales of reaction. Thus, not only are data on internal consistency available for these scales, but also the internal consistencies can be compared across studies, noise sources, and populations (civilian or military). The following scales of reaction were used in the studies (*12,14,26–29*) and can be assessed for internal consistency.

1. General scale based on how much the person is affected by, and dissatisfied with, the noise;
2. Annoyance scale;
3. Complaint disposition scale;
4. Disturbance scale; and
5. Overall reaction based on all of the previous scales.

Table 1 presents the relevant results for each scale in terms of a measure of internal consistency, the maximum likelihood estimate of reliability known as the α coefficient (*22,30*). These data indicate that all the scales used have high internal consistencies regardless of the noise type or the population surveyed (civilian or military). The internal consistencies are themselves extremely consistent, with only a small range of values represented for each scale. The means of the α coefficients for each scale indicate that the scales do not differ greatly in terms of their internal consistency (the range of mean α values is 0.853 to 0.923). However, some of the scales required considerably more items to obtain the observed reliabilities than others. As a guide to the internal consistency of the scales—independent of the number of items used to compose the scale—the average interitem correlations have also been presented. These data suggest that the general scale and the annoyance scale are superior to the disturbance and complaint disposition scales. The general scale is also significantly superior to the annoyance scale ($p < 0.01$). However, in practical terms this statistically significant difference is not of great importance. As the last row of Table 1 indicates, both scales require three items contributing to them in order to achieve an α value of greater than 0.9. This row also indicates the practical consequence of the reduced internal consistency of the disturbance and complaint disposition scales that required large numbers of items to achieve an α value of 0.9.

TABLE 1 ALPHA COEFFICIENTS FOR THE VARIOUS SCALES OF REACTION

Study	Sample Size	Noise	General		Annoyance		Disturbance		Complaint Disposition		Overall Reaction	
Bullen, et al. 1986 (28)	3580	Aircraft	2	.90**	3	.92	8	.85	8	.88	6	.93*
Hede & Bullen 1982 (27)	201	Rifle	2	.92**	4	.94	6	.76	6	.90	4	.92 *
Bullen, et al. 1985 (12)	624	Aircraft	2	.88**	5	.94**	6	.90*	-	-	3	.90*
O'Loughlin, et al., 1986 [24]	318	Rifle Range	2	.89*	-	-	-	-	-	-	5	.92*
Bullen, et al. submitted (29)	1,626	Artillery Range	2	.88**	3	.89	9	.90	8	.84	5	.877*
Job & Bullen 1987 (14)	45 (interview sample)	Aircraft	-	-	-	-	-	-	-	-	5	.841
	41 (group sample)		-	-	-	-	-	-	-	-	5	.939
Mean No. Items			2.00		3.75		7.25		7.33		not computed, since only a small sample size study (14) has appropriate data	
Mean alpha for each scale			.894		.923		.853		.873			
Mean inter-item correlation			.808		.762†		.445†		.484†			
Number of items required for alpha to exceed .9			3		3		12		10			

* For overall reaction the figures reported are results of multiple regression of the various scales against the general scale.

** Computed from mean inter-item correlations.

† Computed from the alpha, via equation 1 in text.

Effects of the Number of Items in a Scale

As pointed out earlier, with the brief interval test-retest procedure the result is interpreted as internal consistency. However, this measure may be affected by factors other than the differences in question wording aimed at measuring the same underlying feature. Reliability in social surveys may be reduced by unwanted influences such as voice intonation, nonverbal cues from the interviewer, and idiosyncratic interpretations of the wording of the question (*16–19*). Consistent with the possibility that the effects of the interviewer's voice intonation and nonverbal cues may add to the error, a recent study found superior internal consistency of a reaction scale when the questionnaire was self-administered rather than given by interview (*14*). The error variance attributable to such influences is reduced by user of several questions to measure the same variable. Thus, the use of several items not only allows examination of internal consistency, but is also likely to increase the observed stability. If a number of questions are used to form a scale rather than a single question, the maximum likelihood estimate of reliability is given by

$$\alpha = jr/[1 + (j - 1)r] \qquad (1)$$

where r is the average correlation between the questions and j is the number of questions (*31*).

As a guide to the increase in reliability which occurs with the use of more questions, this function is graphed for various values of r in Figure 1.

By the way of confirmation of this increase in reliability in relevant surveys, it is possible to calculate the brief interval test-retest reliability (or internal consistency) of various questions used to measure annoyance in surveys using more than one question. Bullen et al. (*25*) used four questions as measures of annoyance caused by aircraft noise. The average interitem correlation for the four items used by Bullen et al. was calculated, as was the average correlation between scales created by combining any two of the items. In the calculation of the latter correlation, only correlations between scales with no items in common were used; only three such comparisons exist: 1+2 versus 3+4; 2+3 versus 1+4; and 2+4 versus 1+3. The average correlation between the single items was 0.783, which yields a predicted correlation (or brief interval test-retest reliability) of 0.878 for the two-item scales. The average correlation between the two-item scales, calculated from the three comparisons above, was 0.879. This value is close to the predicted value and significantly different from the average single-item interitem correlation of 0.783 ($p < 0.05$). Similar analyses were conducted for other studies where the relevant data were available. The results are presented in Table 2.

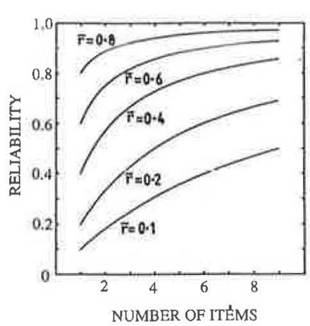

FIGURE 1 Composite scale reliability as a function of the number of items in the scale and the interitem reliability.

The examples in Table 2 indicate that Equation 1 yields a good prediction of the results of socioacoustic studies and that the reliability of self-reported noise reaction measurements may be significantly increased by the use of more than one item to measure reaction. Despite this increase in reliability, numerous surveys have used a single question as the basic measure of reaction even though many relevant questions existed in the questionnaires (*6,32–34*).

By taking into account the reliability of the reaction measure, it is possible to calculate the extent to which the true variance of reaction is predicted by the observed noise exposure. The equation is (*31*)

$$r_{nR} = r_{nr}/(r_{rr})^{1/2} \qquad (2)$$

where

- r_{nR} = correlation between noise exposure and the true component of reaction,
- r_{nr} = obtained correlation between noise exposure and reaction, and
- r_{rr} = reliability coefficient for the reaction measurement.

This equation may be applied to the changes in reliability arising from the use of several items in measurement of reaction by Bullen et al. (*12,27*). Bullen, Job, and Burgess (*12*) found that the reliability of a single-item scale (i.e., the mean interitem correlation) was 0.740, whereas the measures of reaction constructed from several items achieved a reliability of r_{rr} = 0.902, with a noise-reaction correlation of r_{nr} = 0.580. Substituting the last two figures in Equation 2 results in correlation between noise exposure and the true component of reaction, of r_{nr} = 0.611. Substituting this figure in Equation 2 with the single-item reliability of 0.740 reveals that the observed correlation would have been r_{nr} = 0.526 with the use of a single-item reaction scale. Thus the drop in reliability of reaction measurement from r_{rr} = 0.902 to 0.740 would have resulted in a drop in the observed noise-reaction correlation from r_{nr} = 0.580 to 0.526. Similarly, the use of a single-item measure (r_{rr} = 0.783) rather than a composite measure (r_{rr} = 0.931) by Bullen, Hede, and Kyriacos (*27*) would have resulted in the observed noise-reaction correlation being reduced from r_{nr} = 0.361 to 0.331. Thus, in the two studies under consideration, the use of composite scales of reaction resulted in increases of 6.0 and 2.1 in the percentage of variation in reaction accounted for by noise exposure.

A more general case yields a slightly larger effect because the studies analyzed have unusually high reliabilities of reaction measurement. For example, assuming an r_{nr} value of 0.42 [the average correlation in a recent review (*6*)], and a single-item reaction scale with reliability of r_{rr} = 0.630 (*21*), the expected correlation with a reaction scale based on five items with reliability of 0.63 is r_{nr} = 0.500. This represents an increase of 7.4 percent in the variation in reaction accounted for by noise exposure. It would appear that although

TABLE 2 COMPARISON OF PREDICTED AND OBSERVED RELIABILITY FOR COMPOSITE SCALES OF REACTION, AND EFFECTS OF RELIABILITY ON THE NOISE-REACTION CORRELATION

Study	Mean inter-item Correlation ($\bar{r}$)	Number of Items (k)	Reliability of composite scale		noise/reaction correlation	
			Predicted	Observed	Predicted	Observed
Bullen *et al.*, 1985 (12)	.740	2	.851	.852		
	.740	4	.919	.902		
Griffith *et al.*, 1980 (22)	.642	2	.782	-	.397	.37
	.642	3	.843	-	.413	.43
	.642	4	.878	-	.421	.44
Hall & Taylor, 1982 (15)aircraft	.733	3	.890	.86		
Hall & Taylor, 1982 [9] road	.687	3	.868	.83		
Hall & Taylor, 1982 [9] overall noise	.653	3	.850	.84		
Bullen, et al. 1986 (28)	.783	2	.878	.879		
	.783	4	.941	.931		

increases in observed noise-reaction correlations will result from the use of multi-item scales of reaction, the increases will not be large.

Changes in the reliability of noise exposure measures will have similar effects on the observed r_{nr} value. However, given the already high reliability in noise measurement (*9*), little effect of improvements in noise measurement reliability can be expected. The correlation between the true component of reaction and the true component of noise exposure can be calculated (*31*) from the equation

$$r_{NR} = r_{nr}/(r_{nn} \cdot r_{rr})^{1/2} \qquad (3)$$

where

r_{NR} = correlation between the true component of reaction and the true component of noise exposure, and
r_{nn} = reliability coefficient of the noise measurement.

Thus, this correlation can also be calculated from the equation

$$r_{NR} = r_{nr}/\{jr_{rr}/[1 + (j - 1) \cdot r_{rr}] \cdot r_{nn}\}^{1/2} \qquad (4)$$

where

j = number of items contributing to the reaction scale, and
r_{rr} = average interitem correlation of the j items.

The results of the application of Equations 3 and 4 to studies in which the relevant details were available are presented in Table 3. Three features of the data in Table 3 are noteworthy. First, the long and brief interval measures of reaction scale reliability differ (means = 0.724 and 0.916, respectively) indicating greater variability with more time between the measurement points. These data suggest that the internal consistency measure is an underestimate of the error component of the scales, or that true reaction changes significantly over time. Second, this difference allows a more powerful correction to the computed r_{NR} correlation in the case of the extended delay reliability test. Thus the originally very similar r_{nr} correlations for the studies using long and brief interval reliability tests (means = 0.382 and 0.386, respectively) diverge after correcting for reaction scale reliability [means = 0.450 and 0.402, respectively (see Table 3)]. Third, estimates of noise measurement reliability were only available in studies in which the brief delay reliability (internal consistency) was assessed. Accordingly, the values of r_{nR} computed would have been larger if those studies had assessed extended delay reliability. For example, substituting the average extended delay reliability (r_{rr} = 0.724) for the r_{rr} observed in each of these studies results in r_{NR} increasing to 0.696, 0.293, and 0.431 in the three studies (means = 0.473 compared with the mean of 0.423 in Table 3). Overall, these data indicate that estimates of the amount of variation in reaction accounted for by noise exposure increase slightly when the reliabilities of the measurements of noise and reaction are taken into account: the coefficient of determination ($r\_$) expressed as a percentage increased from an average of 15.0 to 22.4 percent for the three studies mentioned.

Computing the Influence of Modifying Variables

In an attempt to identify the amount of variation in reaction left unexplained, it is worthwhile to consider the role of modifying variables. (Modifying variables are variables in the respondent or environment that may modify reaction to noise.) Although many have been shown to be significantly related to reaction, only two are commonly assessed and consistently account for more than a small percentage of the variation in reaction: noise sensitivity and attitude to the noise source (*9*).

Three problems arise in further consideration of the impact of these variables. Firstly, a correlation between a modifying variable and reaction does not guarantee that the modifying variable influences reaction. Some third variable may influence both the supposed modifying variable and reaction, or the modifying variable may be influenced by reaction, not vice versa. Secondly, the calculation of the coefficient of determination for exposure and reaction on the basis of (the-

TABLE 3 COMPUTED EFFECTS OF RELIABILITY ON OBSERVED NOISE-REACTION CORRELATIONS

r_{nr} and reference source		r_{nn} and reference source (unless as for r_{nr})	r_{rr} and reference source (unless as for r_{nr})†		Computed r_{nR}	Computed r_{NR}
Long delay test-retest reliability						
.477	Bradley & Jonah Jonah, 1979 (11)	-	.75		.551	-
.36	Griffiths et al. 1980 (22)	-	.642		.449	-
.37		-	.782**		.418	-
.43		-	.834**		.471	-
.44		-	.878**		.470	-
.21	Langdon 1976 (41)	-	.626	(average of .6 (21) & .642 (22)	.265	-
.325*	Langdon 1978 (26)	-	.61		.416	-
.37	Langdon 1978 (37)	-	.76**		.425	-
.46	McKennell 1963, 1973 (38,39)	-	.63	(22)	.580	-
Means	(Long Delay) .382	-	.736	(excluding .626)	.450	-
Brief delay reliability (internal consistency)						
.477	Bradley & Jonah, 1979 (11)	-	.95		.489	-
.29	Bullen & Hede, 1982 (40)	-	.922	Hede & Bullen 1982 (27)	.302	-
.58	Bullen et al. 1985 (12)	.96	.902		.611	.623
.220	Bullen et al., submitted (29)	.78	.877		.235	.266
.361	Bullen et al. 1986 (28)	.969	.931		.374	.380
Means	(brief delay) .386	.903	.916		.402	.423

* Average of .33 and .32 reported.

** Computed by Equation 1 from information reported (k and $\bar{r}$)

† Where r_{rr} was not obtained from the same study as r_{nr}, the same reaction scale, same type of noise source and same country were involved.

oretically existing) completely reliable measures of both factors has already taken into account some variation in reaction possibly attributable to the modifying variables. This is confounded because the variation in reaction attributable to the modifying variables is treated as error variance that is eliminated in calculating for a completely reliable (zero error variance) reaction scale. Third, the modifying variables may themselves correlate.

The first problem is not easily handled. The data are conflicting as to whether attitude, for example, is a genuine modifying variable or not (*9,26,35*). In the ensuing analyses, it is assumed, with some justification (*35*), that both sensitivity to noise and attitude towards the noise source are genuine modifying variables. The second problem may be avoided by identifying the amount of variation in reaction that is reliable (true) and calculating how much of this component is accounted for by the other factors. The third problem presents particular difficulties when attempting to correct for the reliability of the scale used to measure the modifying variable. That is, a multiple regression could be used to compute, for example, the amount of variation in reaction accounted for by sensitivity after attitude is taken into account. However, this partial correlation of sensitivity and reaction would be affected by the increased correlation of reaction and the true component of attitude over and above the correlation of attitude (as measured) and reaction.

Nonetheless, examination of the data where such multiple regression has been carried out (*26,27*) may yield a guide as to the minimum amount of variation in reaction that is accounted for in the analysis. Bullen et al. (*27*) found that in a multiple regression of modifying variables against reaction, the modifying variables account for 59.4 percent of the variation in reaction. Noise exposure accounted for 13 percent of the variation in reaction. However, some small correlations between the modifying variable and noise exposure mean that these two percentages cannot be added together, as an estimate of the amount of variation in reaction accounted for by noise and modifying variables. This problem arises because the partial correlation for noise exposure and reaction could be expected to be less than 0.361 ($r\_$ = 13.0 percent). However, given the reliability of the reaction scale (r_{rr} = 0.931), only 86.7 percent of the variation in reaction is true variation to be explained. This amount still leaves somewhere between 14.3 and 27.3 percent of the variation in reaction unaccounted for. A significant but uncalculated proportion of this would be taken up by consideration of the lack of reliability of the modifying variables and the exposure measure. Hede and Bullen (*26*) also computed the multiple regression of the modifying variables against reaction, but on this occasion included noise exposure. They calculated that 65.5 percent of the variation in reaction is accounted for by sensitivity, attitude, and noise exposure in multiple regression. The reliability of the reaction scale in this study (r_{rr} = 0.92) indicates that 84.6 percent of the variance in reaction is true variance. This leaves less than 20 percent unaccounted for. More variation would be accounted for in consideration of the reliability of the modifying variables and noise exposure. However, these conclusions are based on the assumption, mentioned earlier, that the modifying variables influence reaction (rather than the correlations being based on some other connection). Although there is some justification for this assumption, conclusions based on it must remain tentative.

CONCLUSIONS

1. The data presented correct, at least partially, the dearth of data on the internal consistency of scales of reaction to noise. The analysis indicates that two scales (the general scale and the annoyance scale) can achieve reliabilities (internal consistencies) in excess of 0.9, with only three items. Complaint disposition and disturbance scales are not as reliable.

2. Analyses of the effects of the use of several items in a given scale of reaction indicate that the increase in reliability is worthwhile, in that the increased reliability may allow greater sensitivity in determining which measures and features of noise are the most important predictors of community reaction.

3. When noise exposure, modifying variables, and the percentage of error variance in the reaction measure are taken into account, the variation in reaction left to be explained is substantially reduced (to around 20 percent). This provides partial support for the suggestion that the variation in reaction left unexplained is not a problem (*25*).

REFERENCES

1. S. S. Stevens. *Psychophysics: Introduction to its Perceptual Neural and Social Aspects*. John Wiley, New York, 1975.
2. S. Fidell, T. Shultz, and D. M. Green. A Theoretical Interpretation of the Prevalence Rate of Noise-Induced Annoyance in Residential Populations. *Journal of the Acoustical Society of America*, Vol. 84, 1988, 2109–2113.
3. J. M. Fields. The Effect of Numbers of Noise Events on People's Reactions to Noise: An Analysis of Existing Survey Data. *Journal of the Acoustical Society of America*, Vol. 75, 1984, pp. 447–467.
4. T. Yoshida and S. Nakamura. Effect of Noise Events on Inhabitants' Reactions to Railway Noise. *Journal of the Acoustical Society of Japan (English)*, Vol. 10, 1989, pp. 339–348.
5. R. B. Bullen and R. F. S. Job. Re-analysis of data presented in "Community Response to Blasting," by Fidell et al. [J. Acoust. Soc. Am., 74, 888–893 (1983)]. *Journal of The Acoustical Society of America*, Vol. 78, 1985, pp. 799–800.
6. S. Fidell, R. Horonjeff, T. Shultz, and S. Teffeteller. Community Response to Blasting. *Journal of the Acoustical Society of America*, Vol. 74, 1983, pp. 888–893.
7. R. F. S. Job and A. J. Hede. Community Reaction to Noise from Power Stations. *Proc., Inter-Noise 1989*, Newport Beach, California. Noise Control Foundation, Poughkeepsie, N.Y., 1989, pp. 865–893.
8. J. Lambert, F. Simonnet and M. Vallet. Patterns of Behavior in Dwellings Exposed to Road Traffic Noise. *Journal of Sound and Vibration*, Vol. 92, 1984, pp. 159–172.
9. R. F. S. Job. Community Response to Noise: A Review of Factors Influencing the Relationship Between Noise Exposure and Reaction. *Journal of the Acoustical Society of America*, Vol. 83, 1988, pp. 991–1001.
10. J. M. Fields and J. G. Walker. The Response to Railway Noise in Residential Areas in Great Britain. *Journal of Sound and Vibration*, Vol. 85, 1982, pp. 177–255.
11. J. S. Bradley and B. A. Jonah. The Effects of Site Selection Variables on Human Responses to Traffic Noise, Part 1: Type of Housing by Traffic Noise Level. *Journal of Sound and Vibration*, Vol. 66, 1979, pp. 589–604.
12. R. B. Bullen, R. F. S. Job, and D. H. Burgess. *Reaction to Aircraft Noise on R.A.A.F. Bases*. National Acoustic Laboratories Commissioned Report No. 7. Australian Government Publishing Service, Canberra, 1985.
13. R. M. Edwards. A Social Survey to Examine the Variance of Aircraft Noise Annoyance. *Journal of Sound and Vibration*, Vol. 41, 1985, pp. 41–50.
14. R. F. S. Job and R. B. Bullen. The Effects of a Face to Face Interview Versus a Group Administered Questionnaire in Determining Reaction to Noise in the Workplace. *Journal of Sound and Vibration*, Vol. 116, 1987, pp.161–168.
15. F. L. Hall and S. M. Taylor. Reliability of Social Survey Data on Noise Effects. *Journal of the Acoustical Society of America*, Vol. 72, 1982, pp. 1212–1221.
16. A. Barah and C. F. Cannell. The Effects of Interviewer Voice Intonation on Reporting. In *Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971–1977*, C. F. Cannell, L. Oksenberg and J. M. Convere, eds., Survey Research Center, University of Michigan, Ann Arbor, 1979.
17. R. L. Kahn and C. F. Cannell. *The Dynamics of Interviewing*, John Wiley, New York, 1957.
18. G. Kalton, M. Collins and L. Brook. Experiments in Wording Opinion Questions. *Applied Statistics*, Vol. 27, 1978, pp. 149–161.
19. H. Shuman and S. Presser. *Questions and Answers in Attitude Surveys*. Academic Press, New York, 1981.
20. I. D. Griffiths and F. R. Delauzun. Individual Differences in Sensitivity to Traffic Noise: An Empirical Study. *Journal of Sound and Vibration*, Vol. 55, 1977, pp. 93–107.
21. I. D. Griffiths, F. J. Langdon, and M. A. Swan. Subjective Effects of Traffic Noise Exposure: Reliability and Seasonal Effects. *Journal of Sound and Vibration*, Vol. 71, 1980, pp. 227–240.
22. J. Nunnally. *Psychometric Theory*. McGraw-Hill, New York, 1978.
23. M. N. Moreira and M. E. Bryan. Noise Annoyance Susceptibility. *Journal of Sound and Vibration*, Vol. 52, 1972, pp. 449–462.
24. M. E. Bryan and W. Tempest. Are Our Noise Laws Adequate? *Applied Acoustics*, Vol. 6, 1973, pp. 219–232.
25. F. J. Langdon. Reliability of Estimates of Annoyance With Road Traffic Noise. *Proc., 3rd International Congress on Noise as a Public Health Problem*, Freiburg, West Germany, 1978.
26. A. J. Hede and R. B. Bullen. Community Reaction to Noise from a Suburban Rifle Range. *Journal of Sound and Vibration*, Vol. 82, 1982, pp. 39–49.
27. R. B. Bullen, A. J. Hede, and E. Kyriacos. Reaction to Aircraft Noise in Residential Areas Around Australian Airports. *Journal of Sound and Vibration*, Vol. 108, 1986, pp. 199–225.
28. R. B. Bullen, A. J. Hede, and R. F. S. Job. Community Reaction to Noise from an Artillery Range (in preparation).
29. B. J. O'Loughlin, R. B. Bullen, A. J. Hede, and D. H. Burgess. *Community Reaction to Noise from Wiliamtown Rifle Range*. Commissioned Report No. 9. National Acoustic Laboratories, Australian Government Publishing Service, Canberra, 1986.
30. C. H. Hull and N. H. Nie. *SPSS Update*. McGraw-Hill, New York, 1979.
31. P. Guilford. *Psychometric Methods*. Tata McGraw-Hill, Bombay, India, 1954.

32. B. J. Murray and B. Avery. *Survey of Community Reaction to Overpressure from Blasting*. Wilkinson-Murray Report W4244, Linfield, Australia, 1984.
33. R. Rylander, M. Bjorkman, V. Ahrlin, S. Sorensen, and K. Berglund. Aircraft Noise Annoyance Contours: Importance of Overflight Frequency and Nose Level. *Journal of Sound and Vibration*, Vol. 69, 1980, pp. 583–595.
34. S. Sorenson and J. Magnusson. Annoyance Caused by Noise from Shooting Ranges. *Journal of Sound and Vibration*, Vol. 82, 1979, pp. 437–442.
35. A. C. McKennell. Annoyance from Concorde Flight around Heathrow. *Proc., 3rd International Congress on Noise as a Public Health Problem*, Freiburg, West Germany, 1978, pp. 562–566.
36. F. J. Langdon. Reliability of Estimates of Annoyance with Road Traffiç Noise. *Proc., 3rd International Congress on Noise as a Public Health Problem*, Freiburg, West Germany, 1978, pp. 567–570.
37. A. C. McKennell. *Aircraft Noise Annoyance Around London (Heathrow) Airport*. U.K. Government Social Survey Report SS337, 1963.
38. A. C. McKennell. Psycho-Social Factors in Aircraft Noise Annoyance. *Proc., International Congress on Noise as a Public Health Problem*, Dubrovnik, 1973, pp. 627–644.
39. R. B. Bullen and A. J. Hede. Assessment of Community Noise Exposure from Rifle Shooting. *Journal of Sound and Vibration*, Vol. 82, 1982, pp. 29–37.
40. F. J. Langdon. Noise Nuisance Caused by Road Traffic in Residential Areas: Part 1. *Journal of Sound and Vibration*, Vol. 47, 1976, pp. 243–263.

Publication of this paper sponsored by Committee on Transportation-Related Noise and Vibration.