REFERENCES

1. SAE J1116, Categories of Off-Highway Self-Propelled Work Machines. Society of Automotive Engineers, Inc., April 1978.
2. SAE J1057a, Identification Terminology of Earthmoving Machines. Society of Automotive Engineers, Inc., June 1975.
3. E. Cohen. Is Motion Needed in Flight Simulators Used for Training? Human Factors, Vol. 12, No. 1, 1970, pp. 75-79.
4. D.J. Gibino. Effects of Presence or Absence of Cockpit Motion in Instrument Flight Trainers and Flight Simulators. Aeronautical Systems Division, U.S. Department of Defense, Wright-Patterson AFB, OH, Rept. ASD-TR-68-24, June 1968.
5. ISO 2041, Vibration and Shock-Vocabulary. International Organization for Standardization, 1975.
6. R.L. Anderson and R.L. Brierly. Measuring Automatic Transmission Shift Performance. Society of Automotive Engineers, Inc., Paper No. 650465, 1965.
7. J.C. Barton and R.E. Hefner. Whole Body Vibration Levels: A Realistic Baseline for Standards. Society of Automotive Engineers, Inc., Paper No. 760415, 1976.
8. D. Simic. Beitrag Zur Optimierung Der Schwingungseigenschaften Des Fahrzeuges: Phsyiologische Grundlagen Des Schwingungskomforts. Technical Univ., Berlin, Germany, Dissertation D83, 1970.
9. H.E. Von Gierke. The ISO Standard Guide for the Evaluation of Human Exposure to Whole-Body Vibration. Society of Automotive Engineers, Inc., Paper No. 751009, 1975.
10. ISO/TC108, Guide for the Evaluation of Human Exposure to Whole-Body Vibration. International Organization for Standardization, International Standard ISO 2631-1978, 1978.
11. ISO/TC108, Evaluation of Exposure to Whole-Body z-Axis Vertical Vibration in the Frequency Range 0.1 to 1.0 Hz. International Organization for Standardization, Draft Addendum ISO 2631/DAD2, 1980.
12. D.E. Wasserman, T.E. Doyle, and W.C. Asburry. Whole-Body Vibration Exposure of Workers During Heavy Equipment Operation. DHEW (NIOSH) Publication No. 78-153, April 1978.
13. T.H. Milby and P.C. Spear. Relationship Between Whole-Body Vibration and Morbidity Patterns Among Heavy Equipment Operators. DHEW (NIOSH) Publication No. 74-131, July 1974.
14. R.C. Spear, C. Keller, V. Behrens, M. Hudes, and D. Tarter. Morbidity Patterns Among Heavy Equipment Operators Exposed to Whole-Body Vibration--1975. DHEW (NIOSH) Publication No. 77-120, Nov. 1976.
15. M.J. Smith, M.J. Colligan, and J.J. Hurrel, Jr. A Review of NIOSH Psychological Stress Research--1977. DHEW (NIOSH) Publication No. 78-156, March 1978.
16. A.M. Stave. The Effects of Cockpit Environment on Long-Term Pilot Performance. Human Factors, Vol. 19, No. 5, 1977, pp. 503-514.
17. D.A. Bernard, T.R. Ferragut, and D.L. Naumann. Production Efficiency Study on Large-Capacity, Rubber-Tired Front-End Loaders. Federal Highway Administration, Rept. No. FHWA-RDDP-PC-520, May 1975.
18. A.M. Stave. The Influence of Low Frequency Vibration on Pilot Performance (As Measured in a Fixed Base Simulator). Ergonomics, Vol. 22, No. 7, 1979, pp. 823-835.
19. ISO/TC108, Guide to the Evaluation of Human Exposure to Vibration and Shock in Buildings. International Organization for Standardization, Draft Addendum ISO 2631/DAD1, 1980.
20. SAE J1013, Measurement of Whole-Body Vibration of the Seated Operator of Agricultural Equipment. Society of Automotive Engineers, Inc., 1973.
21. SAE J1013 JAN 80, Measurement of Whole-Body Vibration of the Seated Operator of Off-Highway Work Machines. Society of Automotive Engineers, Inc., 1980.
22. ISO/TC108/SC2/WG4, N20, Mechanical Vibration-Land Vehicles-Reporting Measured Data. International Organization for Standardization, Working Draft, 1980.
23. SAE J1225, Development of a Frequency Weighted Portable Ride Meter. Society for Automotive Engineers, Inc., 1978.
24. F. Pradko, T.R. Orr, and R.A. Lee. Human Vibration Analysis. Society of Automotive Engineers, Inc., Paper No. 650426, 1965.
25. F. Pradko, R.A. Lee, and V. Kaluza. Theory of Human Vibration Response. American Society of Mechanical Engineers, Paper No. 66-WA/BHF-15, 1966.
26. R. Lee and F. Pradko. Analytical Analysis of Human Vibration. Society of Automotive Engineers, Inc., Paper No. 680091, 1968.

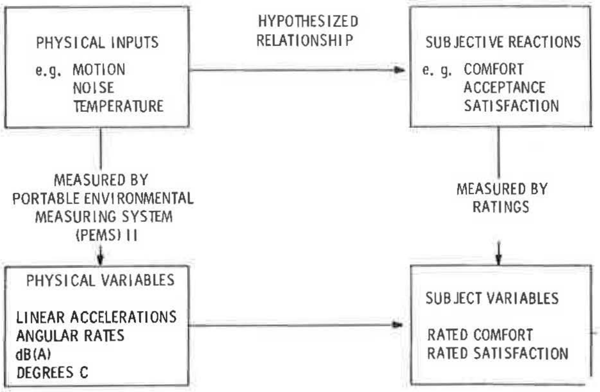
Subjective Ride Quality Testing Procedures

LARRY G. RICHARDS

This paper describes the problems of obtaining passenger ratings of their subjective reactions to different vehicle environments. The rationale for assessing subjective reactions is discussed. Several methods of measuring passenger comfort are noted, and two common methods are described in detail, along with the base for choosing between them. Practical and psychometric considerations governing the use of rating scales are reviewed, and the importance of individual differences is noted. Finally, a standard sequence of steps is presented for analyzing subjective data and deriving comfort models.

Since the first National Aeronautics and Space Administration (NASA) symposium on ride quality in 1972 (1), there has been concern about the inclusion of subjective factors in ride quality research. Much of this concern has been expressed as suspicion about the legitimacy of subjective factors and the desire to eliminate them from studies in this area. Thus, there have been attempts to build ride quality meters and attempts to establish ride quality standards. Neither of these efforts have proven com-

Figure 1. Research strategy for determining the passenger transfer function.



pletely satisfactory--largely because their results do not agree with those of human judges in test situations. Ride quality meters and standards should follow from an understanding of human reactions to vehicle environments, not precede it.

An essential component of ride quality research is the passenger transfer function (Figure 1). This is an equation or set of equations relating physical aspects of a vehicle's internal environment (motion, noise, pressure, etc.) to passenger reactions to that environment (comfort, acceptance, etc.). There has been substantial disagreement about which subjective factors (outputs) to measure and what measuring techniques to use. Many of these disputes are unnecessary because the choice of a testing method is often dictated by different test situations, subject populations, and the information requirements specific to the research problem. Furthermore, the results obtained by using different methods in different situations are generally compatible.

RATIONALE FOR STUDYING SUBJECTIVE REACTIONS

Ride quality is one of many factors that determine passenger acceptance of transportation systems. The viability of a transportation system or mode is largely dependent on how people react to it. Mode choice and decision processes are cognitive operations that depend on other subjective reactions, such as perceptions, preferences, motivations, evaluations, attitudes, and beliefs. Thus to understand transportation choice and use, many psychological concepts and principles must be involved.

The assessment of ride quality is basically a problem in complex psychophysics. Internal states of the person (subjective reactions) must be related to physical inputs (environmental variables). What a person perceives or feels may be assessed most directly by asking him or her about it. There are many ways to ask a person about their subjective reactions; several such methods will be described below.

Why not simply use behavioral or physiological measures instead of reports of subjective reactions? There are several reasons. The most important is that the relevance of either behavioral or physiological measures must be established by their correlation with the appropriate subjective states. What behavioral or physiological manifestations reflect a person's level of comfort? The only way to answer this question is to assess comfort independently and, thereby, see which variables or behaviors correlate with it. Thus, the subjective assessment of comfort is logically prior to, and necessary for, establishing the validity of behavioral or physiological measures. But if the subjective reactions must be assessed, there is no real reason to obtain the other indices.

In certain circumstances, behavioral or physiological measurements may be valuable, especially as related to health and safety. However, whether or not a passenger vomits is an insensitive measure of discomfort because it is an all-or-none measure. Whether or not he or she uses the vehicle is an insufficient measure of preference. An individual may be very dissatisfied with the local transport system, but still use it because it is all that is available. However, a designer who knows of this dissatisfaction might be able to develop an alternative system to attract a substantial portion of the ridership away from the existing system.

COMFORT AND RIDE QUALITY

Ride quality research has generally involved the measurement of passenger comfort. Comfort would seem to be the direct psychological correlate of ride quality. If all other conditions are acceptable and the ride is good, the passenger is comfortable; if it is bad, the passenger experiences discomfort. Thus, comfort reflects all the environmental factors that may act on a passenger.

Why not simply have the passenger assess the "ride quality" of the vehicle? What the passenger is asked to assess depends on what we wish to know. An individual considering the ride quality of a vehicle may make different assumptions about what he or she should attend to and assess than if he or she were considering a personal level of comfort. In the first situation, the individual may attend to the vehicle, but not to his or her reaction to the vehicle. Under these circumstances, the individual may assume that ride quality is directly related to the amount of motion present: The more motion there is, the poorer the ride quality. If taken to an extreme, the best vehicle would be the one with the least motion. Such a person would behave like an accelerometer in his or her judgments about the vehicle. The absence of motion would define good ride quality, and any detectable motion would degrade the quality of the ride.

But is this the kind of ride quality evolution we want? If the person judging ride quality behaves as an accelerometer, there is no further need to obtain judgments on ride quality since a real accelerometer would do a better job.

If human reactions to the vehicle are of primary interest to the researcher, it is necessary to first focus the passenger's attention on those reactions. Comfort reactions are the appropriate states to assess because passengers can be asked to report how comfortable or uncomfortable they are in various environments and the physical correlates of comfort can then be determined. Other subjective reactions may also be assessed and related both to physical variables and to comfort. The research strategy for determining the passenger transfer function is illustrated in Figure 1. Passenger reactions are assumed to be related to physical inputs from the vehicle environment. The vehicle environment is measured by using the appropriate instruments, and subjective reactions are assessed from passenger judgments and reports. These two sets of measurements are then related to each other.

Assessment of Comfort

Any of a variety of psychophysical methods could be used to assess comfort. For example, binary judgments could be obtained from passengers placed in various environments and asked to indicate simply whether they were comfortable or uncomfortable. Although some modeling is possible with such data, better models will result from more quantitative information concerning comfort. Rating scales have been used in field research to generate ride quality models for diverse vehicles (2,3). Magnitude estimation has been extensively used in laboratory studies of human comfort (4). Ranking methods and paired comparison techniques have been used only infrequently. They have not seemed to provide any notable advantage.

Rating scales require absolute judgments from the research subject involved, while ranking and paired comparison tasks involve comparative judgments. However, comparative methods present problems for domains in which the stimuli cannot be experienced simultaneously. For example, sequential presentation of ride segments requires the subject to compare his or her experience of one segment to the memory of his or her experience with the previous one. Yet the extent to which people can remember motion experiences is unknown.

The major ride quality research projects have used either rating scales or magnitude estimation to assess human comfort. When a rating scale is used, the person is asked either to place a stimulus (e.g., a ride segment) into a category or to assign a number to it reflecting the magnitude or level of an attribute displayed by (or resulting from) the stimulus. In ride quality research at the University of Virginia, passengers rated ride segments by using a seven-point rating scale as follows:

Verbal Comfort Rating	Scale Value
Very comfortable	1
Comfortable	2
Somewhat Comfortable	3
Neutral	4
Somewhat uncomfortable	5
Uncomfortable	6
Very uncomfortable	7

In the standard magnitude estimation task, the subject is given a stimulus and told that it represents some magnitude of a property (say 10 units of comfort). Then a second stimulus is presented and the person must assign a number to it reflecting how much of the property it displays. Magnitude estimation methods seek to produce numerical responses satisfying the properties of a ratio scale. Such a scale has a zero point and thereby permits the investigator to meaningfully compare the ratios of numerical judgments. This type of data also allows the investigator to determine the coefficients for the power law, which is presumed to be the psychophysical law (5).

Up to the present, the bases for choosing between these two methods have been largely determined by the research setting favored by the investigator. Laboratory studies, with captive or paid subjects who are available for large blocks of time, have generally used magnitude estimation procedures. Field studies, with regular commercial passengers or respondents hired for a brief time, have generally used rating scales. Thus, the choice of methods has been tied up with two types of distinctions, depending on whether a laboratory or field setting is involved, and whether subjects are captive (and relatively sophisticated about the experimental technique) or regular passengers. When regular commercial passengers are surveyed, it is necessary to use simple experimental techniques that capitalize on passengers' knowledge and use of everyday terms and that do not require much time to explain or accomplish. Thus for field studies, rating scales seem to be the most reasonable procedure to use.

The two approaches described above are complementary and each may be used to deal with specific types of questions. Field work is necessary to determine how people react to real-world vehicles and what physical factors influence their comfort and well-being. Laboratory work can be pursued to separate out various motion components, extend the range of motions studied, explore unique motion combinations, and simulate nonexistent vehicles. A great deal of work has been accomplished with each approach, and an analytic research effort should be undertaken to try to integrate the results from both.

Some Considerations on Use of Rating Scales

The first consideration in designing any rating scale is to determine what needs to be rated. It was argued above that, although ride quality may be the attribute of interest, passengers' ratings of comfort provide the best dependent variable for developing ride quality models. The second consideration is to ensure that the terms used to describe the rating scale are readily understood by subjects in the population to be studied. The terms should be unambiguous so that the subject will know what he or she is being asked to rate. Common terms, such as "comfort," are less ambiguous for most people than specialized terms such as "ride quality."

A third consideration is to ensure that the scale is unitary; the scale description should not combine dimensions, mix criteria, or confound various attributes. The seven-point comfort scale described above is such a unitary scale--it measures only one dimension of experience. Comfort is a unitary subjective reaction even though it is influenced by a large number of factors. If, on the other hand, the scale descriptions were more elaborate and complicated, they might involve multiple criteria. For example, if the endpoints of a scale were "very comfortable,...I could ride here all day," and "very uncomfortable,...I could not stand this ride for a minute," then the scale is confounded; it mixes assessments of comfort and judged duration of tolerance for the motion. These two attributes may be highly correlated, but they should not, without evidence, be assumed to be measuring the same thing. If the two attributes were assessed independently and found to be perfectly correlated with each other and similarly related to input variables, then they could reasonably be taken as measures of the same subjective reaction. However, without such prior study, separate judgments should be obtained for various criteria of interest: comfort, acceptability, judged tolerance for exposure duration, and judged ability to perform activities. In studies of aircraft passenger comfort, Richards and Jacobson (6,7) assessed both overall trip comfort and mode satisfaction in terms of willingness to use it again. They were then able to show the empirical relationship between these two variables and establish that comfort does in fact influence mode acceptability. Figures 2 and 3 show this relationship for two sets of aircraft data. Similar results have since been obtained for various ground-based vehicles. Likewise, the other subjective reactions and judgments mentioned above may be empirically separable from, but related to, the person's subjective experience of comfort. Discovering the appropriate relationships is a matter for research.

A fourth consideration is whether the rating scale should be bipolar (i.e., a scale proceeding in two directions from a neutral point) or unipolar. This issue may often be resolved by a linguistic analysis of the terms used for the scale. There is good reason, based on our use of the language, to assume that people can distinguish ordered levels of comfort from "extremely comfortable" through "neutral" to "extremely uncomfortable." Richards (8) has proposed that the passenger's experience of comfort/discomfort always involves effect, and, therefore, will correlate highly with the semantic dimension of evaluation (e.g., good, bad). While comfort is clearly bipolar, peoples' reactions to motion environments might be either all negative or a mixture of positive and negative experiences; that is, motion might produce only discomfort, or both comfort and discomfort. Whether passengers' responses to vehicle environments are unipolar or bipolar is an empirical question, which can only be answered by presenting bipolar rating scales to them. If a person experiences only discomfort, he or she may reveal that with a bipolar scale, but a person who has only a discomfort scale cannot reveal levels of comfort he or she may experience. The unipolar/bipolar scale issue may be resolved by examining the actual response distributions passengers produce during the field study or experiment. Distributions from two separate aircraft studies (6,7) are shown in Figure 4. Passengers in actual travel situations freely respond that they are comfortable in moving vehicles; some are even "very comfortable." "Very comfortable" of course could not be measured on a "unipolar" discomfort scale.

Figure 2. Percentage of aircraft passengers satisfied as a function of comfort level, 1975 data (18).

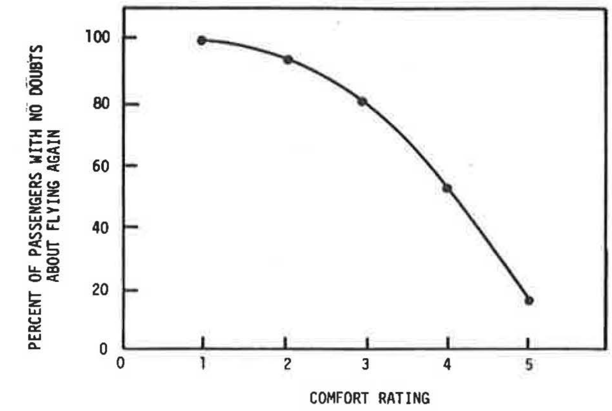


Figure 3. Percentage of aircraft passengers satisfied as a function of comfort level, 1977 data (18).

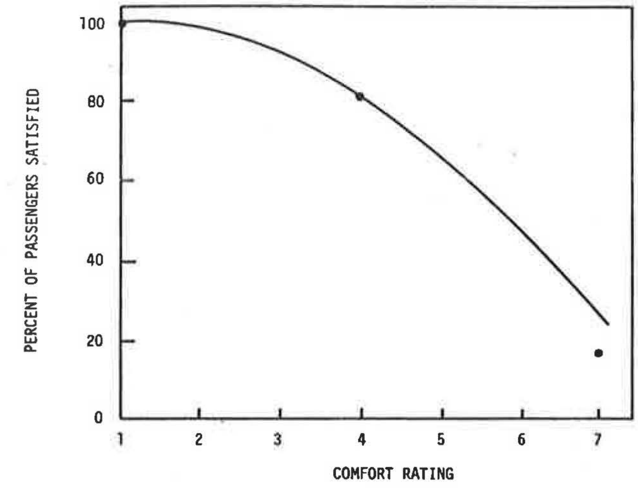
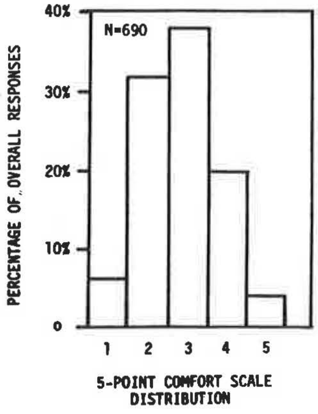


Figure 4. Distribution of comfort responses on 5-point scale for two aircraft studies (18).



A fifth necessary consideration in the development of a rating scale is to determine how many scale steps (levels) to use. For reasons that will be more fully developed below, seven categories is the minimum that should be used with a bipolar attribute. More levels are preferable in most situations since subjects tend not to use the end points of a rating scale. Thus, the effective length of the scale is often less than the actual number of categories. Obviously, a very large number of points might intimidate the respondent.

Psychometric Properties of Rating Scales

Psychological measurement techniques are usually evaluated in terms of their reliability, validity, norms (response distributions), and meaningfulness (theoretical and empirical relevance). The issue of reliability may be dealt with in two ways: (a) there are general psychometric results that determine the possibility of attaining reliable judgments, and (b) there is the empirical question of whether subjects in a particular experimental situation can or do use a rating scale reliably. The possibility of obtaining reliable judgments is related to the number of levels (steps) in the rating scale; from the perspective of psychometric theory, the greater the number of response categories, the more accurate is the scale. Increasing the number of scale steps creates the possibility of more reliable judgments at least up to a degree. There is considerable literature regarding this point. From the perspective of information theory Eriksen and Hake (9), Garner (10), and Bendig (11) have shown that the amount of information transmitted by a subject who uses a rating scale increases with the number of categories for scales up to 20 steps. The theoretical reliability of individual rating scales is a monotonically increasing function of the number of scale steps (12,13). This increase is very rapid in the range from 2 to 7 steps; it is much less in the range from 7 to 11 steps; and it is negligible beyond 11 steps. From the framework of mental test theory, as the number of scale steps increases, both true score variance and error variance increase. However, initially the increase in true score variance is very rapid in comparison with error variance. The increasing contribution of true score variance is greatest in the range up to 7 steps; beyond 11 steps the relative contribution of error variance is usually greater. In terms of correlation theory, reliabilities based on a few scale steps are more influenced by the number of steps and by the marginal distributions of judgments than are those based on more scale steps (14).

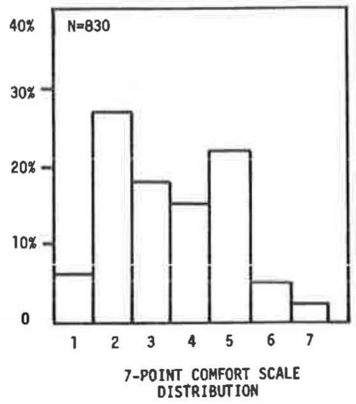
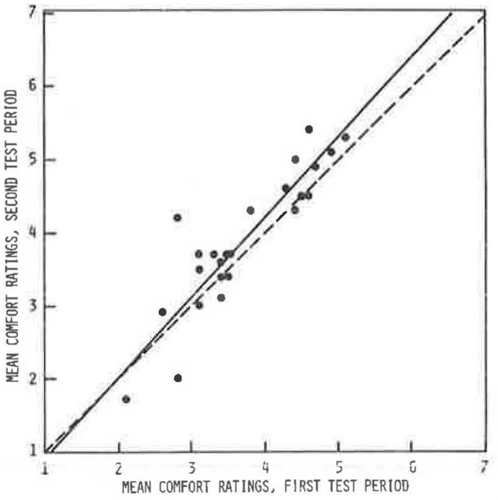


Figure 5. Distribution of comfort responses on 7-point scale for two aircraft studies (18).

Figure 6. Comparison of mean comfort ratings to 24 motion segments during two test periods (18).



Ideally, a rating scale should allow a subject to make all the discrimination he or she is capable of making. If too few scale steps are used, we lose available and potentially valuable information. Finer categories can always be collapsed into coarser ones if it seems that the subjects failed to use the additional categories, or perhaps used them unreliably. But there is no way to extract finer discriminations from coarse categories.

People making judgments sometimes complain about the subjective difficulty of making fine discriminations. According to Johnson, one could argue that all the discriminations a judge makes should be difficult, because only then do we benefit from all of his or her discriminative power (15). In addition, Johnson argues that the subjective difficulty of making judgments is usually not reflected in the actual data. Even if the judgments were difficult to make, they may still be reliable.

The empirical question then is whether the comfort ratings given by passengers experiencing segments of a ride are reliable. Various early studies with ground-based and in-flight simulators established that subjects could reliably rate their comfort in response to motion inputs. The data of Brown (16), shown in Figure 5, provide one example.

The issues of validity and meaningfulness are generally related in psychological studies. In our cases, comfort ratings are taken as indicator responses (indices) of the hypothesized subjective reactions. The validity of comfort ratings is established by the fact that they behave as our theoretical notions about comfort predict they should (8). In particular, comfort ratings relate to motion inputs, noise levels, and seating and space variables in meaningful and consistent ways; this permits the development of ride quality models (2,17,18). Furthermore, comfort ratings are related to passenger's ratings of their ability to perform certain tasks in motion environments and to their actual performance of those activities (19). And, as noted above, comfort ratings and passengers' judgments of mode acceptability are consistently and reasonably related to each other. Both are also meaningfully related to input variables.

Finally, the response distributions of comfort judgments behave in reasonable ways across different vehicles and situations. (See Figure 6.) Large planes are rated more comfortable than small ones, compact cars are more comfortable than subcompacts, and luxury buses with an improved suspension system are rated more comfortable than regular buses (3). When street segments were driven over repeatedly in automobile comfort studies with a different set of three passengers each time, the same segments were rated poorly by each group. The distribution of ratings over segments was consistent with the several sets of subjects.

Thus, comfort ratings display the appropriate properties of reliability, validity, distributional consistency, and theoretical relevance that we demand of any measure of a psychological state or process.

Individual Differences and Sampling Subjects

Ratings depend on (a) the object to be rated, (b) the scale used for ratings, and (c) the person doing the rating. Ratings are a type of human judgment, and so may vary with the person making the judgment. Thus, we may find systematic differences in ratings as a function of who is doing the rating. This is illustrated by the fact that judgments vary as a function of experience and familiarity with the objects to be rated and with the rating scale. Ratings will thereby depend on the amount of information an individual has about the object to be rated. Fortunately, responses to ride quality are often relating homogeneous with demographically defined subgroups.

In studies of passenger comfort, the distribution of comfort judgments may depend on various psychological, situational, social, and physical factors. If different groups of people respond to the same environment in different ways, then separate models must be derived for each group, and therefore different passenger transfer functions are necessary.

In studies of airplane comfort, Richards and Jacobson found that attitude toward flying was an extremely potent variable influencing comfort ratings (6). People who said they "love to fly" were generally more comfortable than those who said they "fly because they have to." Furthermore, evidence of the role of attitude on comfort ratings has been found for ground-based vehicles (3). A group of test subjects on the Port Authority Transit Corporation (PATCO) system gave predominantly comfortable responses that had no relation to the motion of the vehicle, whereas a group of subjects riding a luxury bus gave very negative (uncomfortable) responses in the presence of extremely smooth ride segments. This luxury bus study was preceded by an unpleasant 1.5-h wait during which subjects were confined to the bus. Whereas the PATCO study was very pleasant for the subjects, the bus study was a negative experience. Thus, the effects of passenger preconditioning were evident in the respective ratings of the vehicles.

The sex of the respondent also influences comfort responses. Women in airplanes are generally more comfortable than men. Jacobson and Richards (18) attribute this to the fact that the women are usually more satisfied with their seats than men. Yet in ground-based vehicles, women appear to be more sensitive to the angular rates of motion than men (20). Thus, ride quality research should (a) include both men and women as experimental subjects and (b) analyze the resulting data for differences due to the sex of the respondent.

The nature of the subject population is another variable whose effects should be examined in detail. That is, are the individuals doing the rating commercial passengers, paid respondents, or special test subjects? How familiar are they with the vehicle and test situation? By virtue of their common experiences, special circumstances, and training, laboratory subjects may come to differ from people in the real world and may fail to show response patterns that are commonly found for passengers of commercial vehicles. This possibility has not been systematically examined, but it could account for some of the discrepancies found in the results of field and laboratory studies of passenger comfort.

Analysis of Subjective Response Data

Once rating scale data are obtained, the investigator can try to develop ride quality models. The adequacy of the resulting models will depend on the adequacy of both the environmental (physical) variables and the subjective measurements. A standard sequence of steps has been developed to assess the reasonableness of data to derive comfort models (21). All of these steps are necessary to ensure that a modeling effort is reasonable and to guarantee the adequate interpretation of the resulting model. These steps are briefly presented below.

Step 1

Histograms are obtained for both the comfort ratings and the physical variables. A distribution of comfort ratings is generated for each ride segment, as well as for the vehicle (aggregated over subjects and segments). Standard descriptive statistics (mean, median, mode, range, standard deviation, standard error, skewness, and kurtosis) are computed in each case. This information is then examined for peculiarities within the data, such as lack of variability or unusual response distributions. This step is essential because the possibility of developing regression models depends on the adequate variability of the data. Three sets of data mentioned above lacked this necessary variability, and this in turn precluded the development of appropriate ride quality models. For example, the PATCO and luxury bus data mentioned above each displayed restricted ranges of comfort values. Similarly, test flights on the Concorde provided little variation in either physical variables or comfort ratings (22).

Another problematic data set was obtained on an HM2 Hoverferry. When the response distributions for comfort were compiled for each segment, it was discovered that most of the distributions were bimodal since approximately equal numbers of subjects found the segments comfortable as found them uncomfortable. Under these circumstances no meaningful model can be derived by using the mean comfort ratings; the mean is not representative of the behavior of most of the respondents.

Step 2

Matrices of the intercorrelations of physical and subjective variables are computed, and scatterplots are made for each pair of variables. These graphs are then examined for non-linearities in the relations between variables, outliers or discrepant data points, and clusters of observations. Each of these phenomena suggests that the resulting correlations are either not meaningful or mean something different than they would with more typical data.

Step 3

Principal component analyses are done on the intercorrelation matrices for the physical variables to assess the number of independent dimensions of variation in the data (23). While the six degrees-of-freedom of motion may be conceptually independent, they are often correlated in practice. For most of the vehicles studied to date, several motion variables define a single principal component. For example, in these data, roll, pitch, and vertical acceleration are strongly intercorrelated. Such dependencies must be known by the investigator in order to interpret the results of the modeling effort.

Step 4

For each vehicle, stepwise, simultaneous, and constrained multiple regression techniques are used to isolate the best model for the data set. The environmental measurements are used as predictors, and the criterion is the mean comfort response (for a segment) taken over all subjects who rated the segment.

Step 5

Various partitions of the data by subject variables are examined to detect systematic effects due to individual differences (passenger characteristics). Thus, the data for men and women would be processed separately to see if separate models were needed. Other subject variables such as age, driver versus passenger status, and familiarity with the vehicle should also be examined.

Step 6

Usually a validation study is done to see if the model developed in step 4 holds for an independent group of subjects riding in the same vehicle.

Step 7

Data from any particular study are compared with the entire data base generated to date, and proposed composite ride quality models are assessed for their fit to the new data.

REFERENCES

1. Symposium on Vehicle Ride Quality. NASA, TMX-2620, 1972.
2. I.D. Jacobson, L.G. Richards, and A.R. Kuhlthau. Models of Human Comfort in Vehicle Environments. International Conference on Ergonomics and Transport, Swansea, Wales, Sept. 1980.
3. L.G. Richards, I.D. Jacobson, and R.D. Pepler. Ride Quality Models for Diverse Transportation Systems. TRB, Transportation Research Record 774, 1980, pp. 39-45.
4. T.K. Dempsey, J.D. Leatherwood, and

S.A. Clevenson. Discomfort Criteria for Single-Axis Vibrations. NASA, Tech. Paper 1422, 1979.
5. S.S. Stevens. Problems and Methods of Psychophysics. Psychological Bulletin, Vol. 54, 1958, pp. 177-196.
6. L.G. Richards and I.D. Jacobson. Ride Quality Assessment I: Questionnaire Studies of Airline Passenger Acceptance. Ergonomics, Vol. 18, 1975, pp. 129-150.
7. L.G. Richards and I.D. Jacobson. Ride Quality Assessment III: Results from a Revised Questionnaire. Ergonomics, Vol. 20, 1977, pp. 499-519.
8. L.G. Richards. On the Psychology of Passenger Comfort. International Conference on Ergonomics and Transport, Swansea, Wales, Sept. 1980.
9. C.W. Eriksen and H.W. Hake. Absolute Judgments as a Function of Stimulus Range and Response Categories. Journal of Experimental Psychology, Vol. 49, 1955, pp. 323-333.
10. W.R. Garner. Rating Scales, Discriminability, and Information Transmission. Psychological Review, Vol. 67, 1960, pp. 343-352.
11. A.W. Bendig. Reliability and the Number of Rating Scale Categories. Journal of Applied Psychology, Vol. 38, 1954, pp. 38-40.
12. J.P. Guilford. Psychometric Methods. McGraw-Hill, New York, 1954.
13. J.C. Nunnally. Psychometric Theory. McGraw-Hill, New York, 1967.
14. J.B. Carroll. The Nature of the Data, or How to Choose a Correlation Coefficient. Psychometrika, Vol. 26, 1961, pp. 247-372.
15. D.M. Johnson. A Systematic Introduction to the Psychology of Thinking. Harper and Row, New York, 1972.
16. L.G. Richards. Time Dependence and Temporal Information Integration for Human Reaction to Motion. Ergonomics, Vol. 21, 1978, pp. 913-923.
17. I.D. Jacobson and L.G. Richards. Ride Quality Assessment II: Models of Subjective Responses. Ergonomics, Vol. 19, 1976.
18. I.D. Jacobson, L.G. Richards, and A.R. Kuhlthau. Models of Human Reaction to Vehicle Environment. Applied Ergonomics, Vol. 9, 1978, pp. 169-172.
19. D.E. Sussman, J.S. Dumas, A.M. Wichansky, and C.N. Abernethy. Advanced Ground Vehicle Ride Quality. United States-Federal Republic of Germany Cooperative Study, Rept. R5006-PM-79-31, Nov. 1979 (Rev.).
20. L.G. Richards, I.D. Jacobson, and A.R. Kuhlthau. What the Passenger Contributes to Passenger Comfort. Applied Ergonomics, Vol. 9, 1978, pp. 137-142.
21. L.G. Richards, I.D. Jacobson, and R.D. Pepler. Ride Quality Models in Diverse Vehicles. Dunlap and Associates, Darien, CT, June 1979.
22. L.G. Richards and I.D. Jacobson. Concorde: Ride Quality and Passenger Reactions. Aviation, Space and Environmental Medicine, Vol. 49, 1978, pp. 905-913.
23. H.H. Harmon. Modern Factor Analysis. Univ. of Chicago Press, Chicago, 1967.

Ride Quality Evaluation in Transport Aircraft

A. ROBERT KUHLTHAU AND IRA D. JACOBSON

A complete procedure is described for obtaining an estimate by transport aircraft passengers for the vehicle's comfort rating and its effect on overall acceptance of the flight. Passenger comfort is shown to depend on the state of the air through which the vehicle moves, the design characteristics of the vehicle, and the flight maneuvers involved. The comfort ratings are stated in terms of the percentage of passengers who have a certain probability of being sufficiently satisfied with their experience so as not to object to repeating it. Analytic expressions are included for the factors contributing to a passenger comfort evaluation. An interactive FORTRAN program is presented that will allow repeated computations of acceptance to be easily made. The procedure thus becomes a valuable tool for designers and operators to use in the study of the effects of vehicle configurations and flight maneuvers on passenger satisfaction.

The quality of the ride as experienced and assessed by the passengers has a significant influence on the use and acceptance of a particular vehicle to achieve a particular transport mission. In this context ride quality is defined as the impact on the passenger of all aspects of the physical environment of the vehicle that have been found to influence acceptance. This paper presents a summary of a systematic quantitative procedure for evaluating the quality of a ride in a transport aircraft and of determining the effects of this ride quality on the passenger's satisfaction with the ride. It should be mentioned that the basic approach of the process is general and applicable to all modes of transportation. It is the specification of quantitative values that is peculiar to the air transport mode.

The evaluative procedure is not directly dependent on the choice of aircraft or the specific characteristics of the passengers, although these factors are accounted for indirectly, as they do indeed influence the results. Those aspects of the vehicle environment that influence acceptance can be grouped into three general categories: inputs to the vehicle from its surroundings, flight operations, and inherent vehicle design configurations. Inputs from the surroundings depend on characterization of the state of the air through which the vehicle moves. This results in a set of six degree-of-freedom motions caused by the normal response of the vehicle in flight. Flight operations consist of motions induced by maneuvers of the aircraft such as turns, climbs, descents, etc., and of resulting pressure changes that might occur in the cabin. The aircraft design inputs include such things as seating and passenger space limitations, noise caused by aircraft engines or control motors, and cabin temperature extremes caused by inadequate or improperly adjusted air conditioning equipment.

BACKGROUND

A general description of the evaluative procedure can be found in the literature (1,2) and is illustrated conceptually in Figure 1 (25). It requires