THE IMPORTANCE OF RATIONAL AND COMPATIBLE PAVEMENT PERFORMANCE EVALUATION

R. C. G. Haas and W. R. Hudson, University of Texas at Austin

The planning, design, construction, and maintenance of pavements depend in part for success on the evaluation of their performance. Various procedures are available for this purpose, but they should be rational and compatible with each other for the best transfer of information. This paper is concerned with the importance of achieving these goals of rationality and compatibility in pavement performance evaluation. The rationale underlying pavement performance is discussed, and a general framework for developing a performance evaluation scheme is defined. The principal approaches to measuring or predicting pavement serviceability are reviewed, and their major deficiencies in application are pointed out. These deficiencies relate largely to a lack of understanding of the basic principles governing subjective rating procedures, as well as the possible errors involved. Suggestions for improvement of methods in use are presented. Several approaches to approximating present serviceability through condition surveys or roughness measurements are also discussed. It is pointed out that these can be relevant for a particular agency but only qualitatively useful to the data of another agency; further, significant errors can result from some of the transformations used in these approaches. Finally, several suggestions are presented toward achieving better compatibility within and among agencies. These relate to precise definitions, better understanding of subjective rating principles, rigorous error analyses, and carefully designed experiments.

The primary operating characteristic of a pavement at any particular time is the level of service that it provides to the users. In turn, the variation of serviceability with time is some measure of the pavement's performance. This performance and the cost and benefit implications are the primary outputs of the pavement and its overall management system.

Many words and methods have been used to describe the concepts of performance and serviceability. One of the best known procedures for defining and obtaining serviceability was established at the AASHO Road Test and reported by Carey and Irick (1). It was based on subjective evaluation, in terms of the road user's opinion, of the riding quality provided by the pavement at a given (present) time. To develop the method, correlations with physical measurements of the surface for a large set of test pavements were performed and the result termed a present serviceability index (PSI). This PSI has been extensively used, in its original form and in many modified forms, to predict pavement serviceability. The integration of PSI over time or over sum of load applications was termed as performance.

Sponsored by Committee on Pavement Condition Evaluation.

Although the serviceability-performance concept represented very real progress and is widely used by many agencies, the ensuing years have seen considerable confusion and lack of compatibility. This has resulted partially from a proliferation of modifications of the basic method but also from lack of appreciation and understanding of some of the fundamental considerations and possible errors involved with subjective evaluations. It has further stemmed from a seeming lack of appreciation that, while PSI is measured on an objective basis, its purpose is to estimate the subjective opinion of road users. The resulting incompatibilities mean that measurements by one agency often have little or no quantitative meaning to another agency. As a result there is considerable duplication of effort. It seems important for this reason alone that we strive to achieve better compatibility.

The general purpose of this paper is to define the rationale that underlies pavement performance evaluation, to attempt to clarify some of the concepts underlying serviceability measurements, and to suggest means for achieving a greater degree of compatibility in current methodology. More specifically, the objectives are (a) to define the role of pavement performance evaluation within an overall pavement management system; (b) to review the principal methods for measuring pavement serviceability, the underlying assumptions, and their limitations; (c) to discuss some of the incompatibilities among systems and to suggest means for achieving greater compatibility; and (d) to discuss some of the approximate means for estimating pavement serviceability and to classify their functional applicability.

PERFORMANCE EVALUATION AND THE PAVEMENT MANAGEMENT PROCESS

Performance as a Pavement System Output

The process of managing pavements consists of a variety of planning, design, construction, operation, and research activities. Attempts have recently been made by a number of investigators (3, 4, 5, 6, 7, 8, 9) to define portions (including design subsystems) or all of this process in terms of a formal systems framework. These ef-

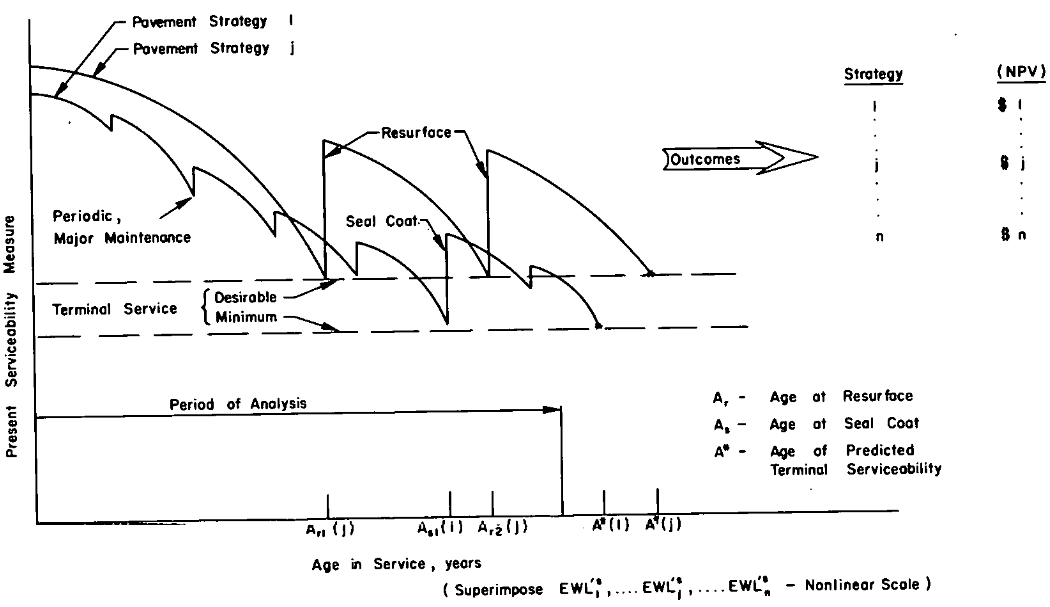

Figure 1. Gross output of a pavement system in terms of performance and value implications.

forts have explicitly recognized that one of the major activities involved must be that of performance evaluation or feedback.

If we accept the fact that the presently imperfect state of technology in the pavement field requires such performance evaluation, then we must first define what outputs of the system are to be evaluated. Figure 1 shows the gross output of alternative pavement strategies in terms of their serviceability-age histories (performance) and the associated value implications. A large number of traffic, materials, climatic, construction, maintenance, and other variables combine to produce any one such performance profile. These are all reflected in the overall pavement strategy that is adopted, and the performance achieved depends on this. (A pavement strategy includes items such as structural design, materials, construction processes and control, maintenance procedures, seal coats, and resurfacing.)

The distinction between serviceability and performance is important, as emphasized by Carey and Irick (1). Figure 1 shows that serviceability is a measure only of the pavement's ability to serve its function at a particular time (i.e., at the present). The past record or suspected future capacity of the pavement is not considered in a single PSI measure. Performance is the serviceability-age history of the pavement. Age rather than equivalent wheel loads (EWL's) carried has to be the primary abscissa in Figure 1 in order that value implications can be taken into account. Furthermore, it is not sufficient to know or predict only initial serviceability and terminal age. Without knowing the intermediate portion, it is not possible to adequately check design strategies, plan or program for maintenance and resurfacing, and explore the implications of raising the terminal serviceability level.

Role of Performance Evaluation in the Pavement Management System

The measurement of the outputs of a pavement system during its time in service (i.e., the evaluation of its performance) has previously been noted as a major management activity. Figure 2 shows the principal elements of this activity as a portion of

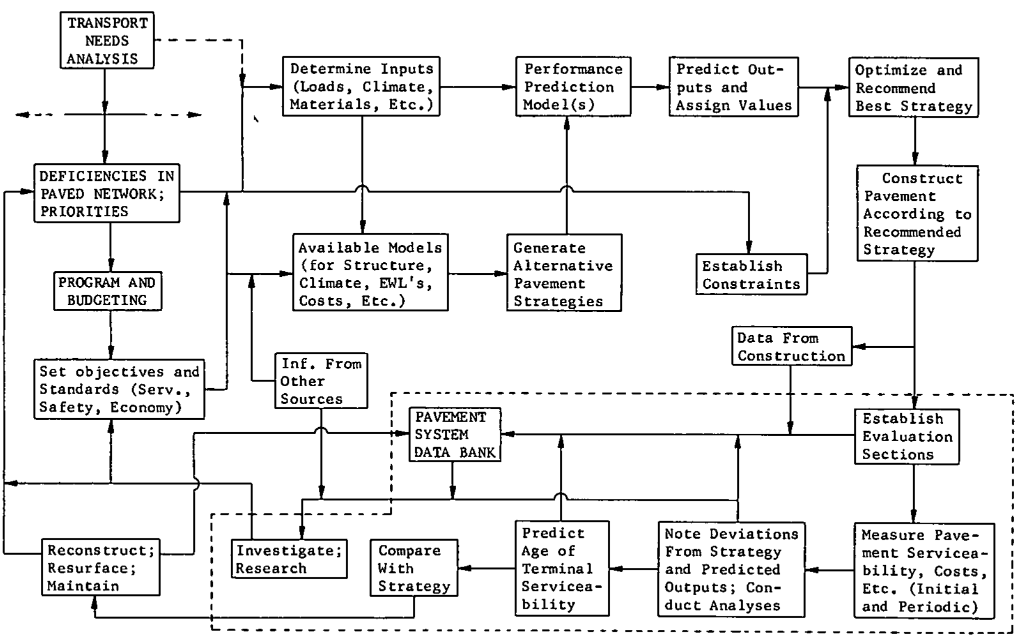

Figure 2. Role of performance evaluation in a generalized pavement management system.

the overall pavement management system and also shows the information flows that result in a continuous process of feedback. The development and implementation of the performance evaluation subsystem as a portion of the management system or of its components can be a comprehensive and major systems problem within itself. Several aspects of this are subsequently discussed in more detail.

DEVELOPMENT OF A PAVEMENT PERFORMANCE EVALUATION SCHEME

Any agency that builds and operates pavements has some method for evaluating their performance. The method may be relatively casual and loosely applied or it may be quite sophisticated in certain aspects. In any case, it is a major part or subsystem of the pavement management process. The rationale underlying the identification, organization, and coordination of such subsystems has been discussed by Haas and Hutchinson (3). They have also identified the major components of these subsystems. In order to develop and utilize a performance evaluation scheme on a comprehensive, unified basis, these components can be translated into a series of steps or phases, as follows:

1. Qualitative identification, classification, and interaction of all factors (such as those of a climatic, materials-associated, load, construction, and maintenance nature), that are important now and are predicted to become important in the future and that affect pavement performance or are outcomes of the performance achieved or the pavement strategy adopted or both;
2. Selection or development of techniques for quantitatively measuring these factors;
3. Development of a sampling plan for items 1 and 2 that is statistically based and progressively implemented with an initial set of key factors from item 1;
4. Design and implementation of a data storage and retrieval system for both present and future needs to handle inputs from item 3;
5. Development of techniques for analyzing stored performance data to check and update design and management models, establish sensitivity of performance factors, evaluate effectiveness of maintenance, and predict terminal serviceability or update predictions for programming or budgeting purposes; and
6. Investigations of the fundamental mechanisms controlling losses in pavement serviceability.

The foregoing phases are also shown in Figure 3 to illustrate the logic and information flows. Most major highway agencies have done considerable work on several of the phases. Few if any, however, have developed and implemented a comprehensive, automated information storage and retrieval system.

MEASUREMENT OF PAVEMENT SERVICEABILITY

General Performance Model

Figure 1 shows the gross outputs of a pavement system in terms of its serviceability-age history, or performance, and the value implications. This is defined by a large number of factors and interactions, and the general form of the underlying performance model for the whole pavement system may be expressed as follows:

$$P_A = f(C, P_o, A, S_p, L, W, R, S_r, S_a, M_d) \quad (1)$$

Where

P_A = measure of performance at age A;
C = construction process effects;
P_o = initial P, immediately after construction;
S_p = strength or stiffness of the pavement layers;
L = traffic loads and repetitions carried by the pavement to age A;
W = climatic factors;
R = roughness at age A;

S_r = slipperiness or skid resistance at age A;
S_a = structural appearance of the surface (i.e., cracking or patching); and
M_d = degree of maintenance to age A.

If all possible interaction terms were included, the model would become extremely complex. Moreover, it would be highly impractical if not impossible to acquire the data for fitting the complete model. Although the conceptual definition of this model is useful to us at this stage for understanding pavement behavior, the existing concept of PSI should not be taken as a limited version of it. Rather, PSI is a definable condition or serviceability level of a pavement. The change or rate of change of PSI with time and traffic provides an objective measurement or definition of performance, P. It is by defining P and relating it to the physical parameters in Eq. 1 that more rational methods of pavement design may be developed.

Principal Approaches

In rating pavements for serviceability, there are 2 basic approaches that are used: One is the evaluation of the present serviceability of the pavement surface (that is, how it is serving traffic today), and the other is appropriate to the more usual engineering approach and involves the mechanistic evaluation of the pavement structure with an eye toward future performance. That is to say, what is the current physical condition of the pavement, and what effect can I expect this condition to have on the future performance of the pavement?

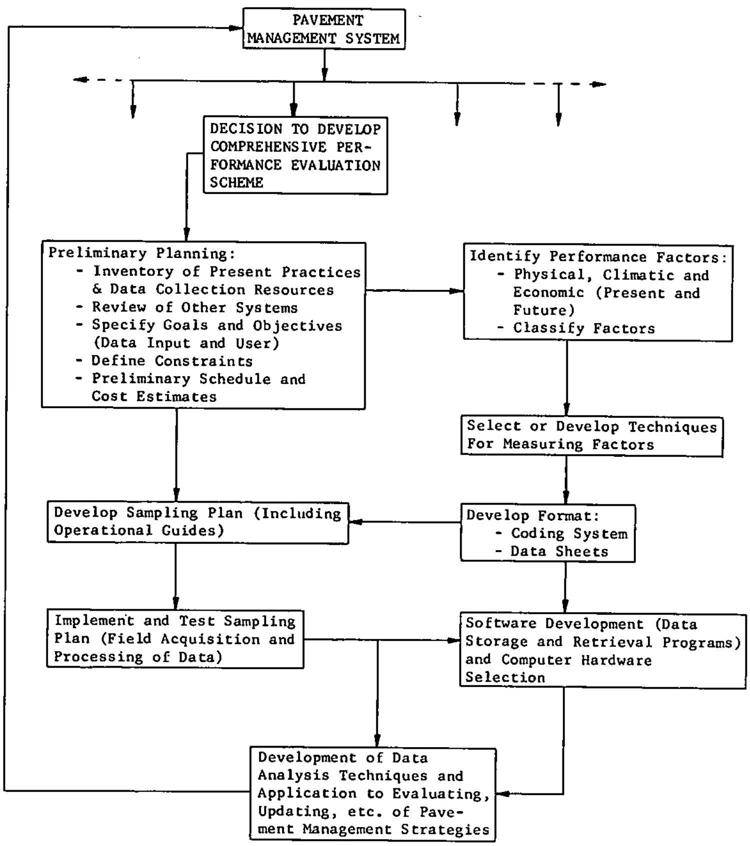

Figure 3. Major phases in developing and applying a pavement performance evaluation scheme.

The difference between these 2 approaches can be illustrated by a simple example. A crack in the pavement surface may have little or no effect on how well the pavement is serving traffic today. On the other hand, the maintenance engineer who looks at this existing crack, in terms of a mechanistic evaluation, immediately thinks of it as a local failure. He knows that it will permit the intrusion of water, that deflections will increase, that spalling and additional cracking may quickly occur, and that the result can be a rapid deterioration of the pavement.

The foregoing might suggest that a variety of physical factors should be considered in the measurement of pavement serviceability. Although this is relevant to an engineering viewpoint of the structure itself, we should define serviceability relative to the purpose for which the pavement was provided, that is, to give a smooth, comfortable, and safe ride. In other words, the measurement should relate explicitly to the user, who is influenced by 3 distinct attributes of the pavement:

1. Response to motion as characterized by the particular pavement-vehicle-human interaction for a particular speed;
2. Response to appearance, as characterized by factors such as cracking and patching, color, and shoulder condition; and
3. Safety, as characterized primarily by slipperiness but also by factors such as color and texture as they relate to visibility and lane demarcation.

The user-oriented measurement of pavement serviceability, to which this paper is primarily addressed, should not obscure the importance of mechanistic evaluations of the pavement. It should be recognized, though, that serviceability is a current measure, while a mechanistic evaluation is mainly an indicator of action needed to maintain serviceability; or it can be an indicator of possible rapid loss in serviceability.

Prior to the AASHO Road Test, the evaluation of pavement serviceability in North America was primarily on a 2-category scale of satisfactory or unsatisfactory, as subjectively judged by the engineer, reinforced with condition surveys and some roughness data. However, as reported by Carey and Irick (1), the AASHO Road Test explicitly recognized the road user in formulating the concept of present serviceability rating (PSR) and correlating it to a set of physical measurement data called present serviceability index (PSI) as follows:

$$PSR = PSI + E \tag{2}$$

where E = error term (i.e., containing "residuals" not explained in the regression equation relating the physical measurement data to serviceability).

The original form of the equation relating the physical measurements to PSI is as follows:

$$PSI = C + (A_1R_1 + \ldots) + (B_1D_1 + B_2D_2 + \ldots) \tag{3}$$

where

- C = coefficient (5.03 and 5.41 for flexible and rigid pavements respectively);
- A_1 = coefficient (-1.91 and -1.80 for flexible and rigid pavements respectively);
- R_1 = function of profile roughness [$\log(1 + \overline{SV})$, where $\overline{SV}$ = mean slope variance from CHLOE profilometer measurements];
- B_1 = coefficient (-1.38 and 0 for flexible and rigid pavements respectively);
- D_1 = function of surface rutting ($\overline{RD}^2$, where $\overline{RD}$ = rut depth as measured by simple rut depth indicator);
- B_2 = coefficient (-0.01 and -0.09 for flexible and rigid pavements respectively); and
- D_2 = function of surface deterioration $\sqrt{C+P}$, where C + P = amount of cracking and patching, determined by procedures described in the AASHO Road Test report (10).

The actual numerical values noted for these coefficients were, of course, determined by Carey and Irick (1) by multiple regression techniques. It must be emphasized, however, that the PSI model represented by Eq. 3 is not an end in itself. Carey and

Irick made this quite clear in pointing out that it is intended to predict PSR to a satisfactory approximation. Unfortunately, this intention for the use of the concept has been forgotten by many engineers in the ensuing years. The engineer is somewhat inherently "hostile" to the concept of a completely subjective evaluation as represented by PSR. He prefers to evaluate his structure by physical criteria that he can measure objectively. Consequently, the PSI concept also largely served the purpose of making available to the engineer a tool that he was more familiar with and amenable to using. However, he has been somewhat unappreciative of some of the implications of employing this to predict subjective, user evaluations, as well as of some of the deficiencies in the subjective user evaluation concept itself. The consequence, as noted earlier, has been a certain degree of incompatibility and confusion in current methodology.

Another major concept of pavement serviceability has been the present performance rating (PPR) developed concurrently with the AASHO Road Test by the Canadian Good Roads Association (11, 12, 13, 14, 15). This method is similar to the panel rating (PSR) of the Road Test except that it uses a 10-point scale rather than a 5-point scale. There are 5 descriptive cues in each, and these are shown on the rating forms in Figures 4 and 5. However, the construction of the PPR scale in effect means that it has 10 categories. In addition, the PPR method emphasizes that only the descriptive words are to be given attention by the rater in judging a particular section and that an exact numerical rating will be scaled off later. (There are detailed rules or guides for doing PPR or PSR ratings, and these are contained in the references noted.) This involves a transformation of scales, where the linear numerical scale as shown in Figures 4 and 5 may or may not be valid, as subsequently discussed.

Another, and perhaps the most important, difference in the PPR and PSI approaches is that the PPR has essentially remained and continued to be used in its subjective panel rating form. Attempts were made, however, to develop a regression model, and the form of the best of these (15) was as follows:

$$PPR = C_O + a_1A + a_2D + a_3D(AT_r + A^2T_r^2) + a_4D(A + A^2) \pm \text{error} \tag{4}$$

where

C_O = coefficient (constant);
a_1, a_2, a_3, a_4 = regression coefficients;
A = age in years;
D = Benkelman rebound value, $\bar{x}$, plus $2\sigma_x$ for the fall, corrected for temperature;
T_r = AADT per lane; and
error = standard error of estimate.

Analyses of the extensive data collected in the Canadian work showed that, for the type of model in Eq. 4, age was the dominant variable, and surface strength was next. Although the analyses were relatively successful in explaining performance variations, the regressions

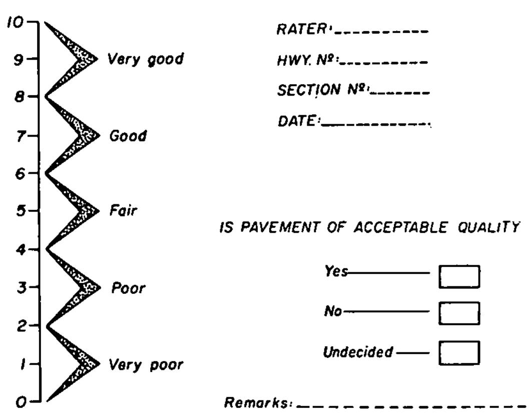

Figure 4. Present performance rating (PPR) form used by Canadian Good Roads Association.

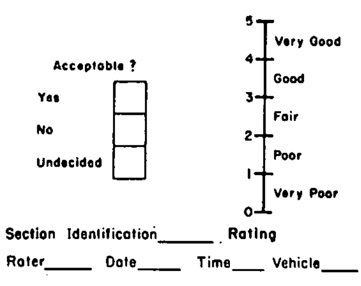

Figure 5. Individual present serviceability rating (PSR) form used at AASHO Road Test.

were not significant enough as a predictive tool for many pavement groups. As a consequence, most agencies have continued to make periodic, direct PPR ratings (some recent efforts have attempted to "mechanize" and approximate these with certain roughness measurements, which will be subsequently discussed) on portions of their paved network.

Such ratings have the advantage of being currently relevant for any particular agency, but the disadvantages of possible interagency incompatibility of results and certain errors, as subsequently discussed, also exist. On the other hand, the use of the original PSI model may result in considerable error in predicting current user opinions in any particular region. Furthermore, any valid attempts at updating, such as that conducted by Scrivner (16) to include a term for surface texture, have to be based on carefully designed and statistically analyzed experiments.

Another method of subjectively evaluating the present serviceability of pavements has been developed by the British Road Research Laboratory. Its application and details of the method have been illustrated by Chipperfield and Welch (17). An example is shown in Figure 7. The method has 6 categories of overall rating (from very good to bad) but is more detailed than PSR or PPR in that 5 areas of structural condition are included to make up the overall rating. These have varying point scales and are also shown in Figure 6. One of the advantages of this system is that structural condition evaluations, which the engineer is more amenable to conducting, are used to establish the overall subjective rating. However, the entire procedure is still subjective, as are PSR and PPR, and therefore may be open to the same types of incompatibility and error as the other methods discussed.

Figure 7 shows in a schematic manner the evolution of pavement performance evaluation methodology. It suggests that, although continual improvement toward achieving an (ideal) "ultimate" performance evaluation system has been made, the most significant improvement is the development of the present serviceability (i.e., PSI or PPR) concept. It further suggests that feedback data systems development may be one means of accelerating improvement in the 1970's.

Principal Limitations of Current Approaches

The engineer who evaluates pavement performance would like to be able to do this on a completely objective, mechanical, and accurate basis, as it is possible to do for certain other structures. However, he must translate any such measurements for pavements into a user evaluation, which is completely subjective. Because subjective rating methodology involves principles of psychophysical scaling, it has been developed

1	2	3	4	5	6	7	8		9
Section No.	Disintegrating ⟵⟶ Excess Binder 1, 2, 3, 4, 5, 6, 7	Deformation None ⟶ Very Severe 0, 1, 2, 3	Traffic Laning None ⟶ Very Severe 0, 1, 2, 3	General Variability None ⟶ Very Severe 0, 1, 2, 3	Texture R→M→S	Overall Condition VG⟶B	Average assessment as later agreed Overall Condition	Texture	Remarks
104	4	0	0	0	M	VG	G	MR	
105	4	0	0	0	M	VG	G	M	
106	5	0	0	0	M	G	G	M	
107	5	0	0	1	M	G	G	M	
108	6	0	1	1	S	FG	FG+z	MS	
109	6	0	2	1	S	FG	FG+z	MS	
110	5	0	1	1	M	G	G	M	

Figure 6. Procedure for individual and overall surface assessment used by British Road Research Laboratory.

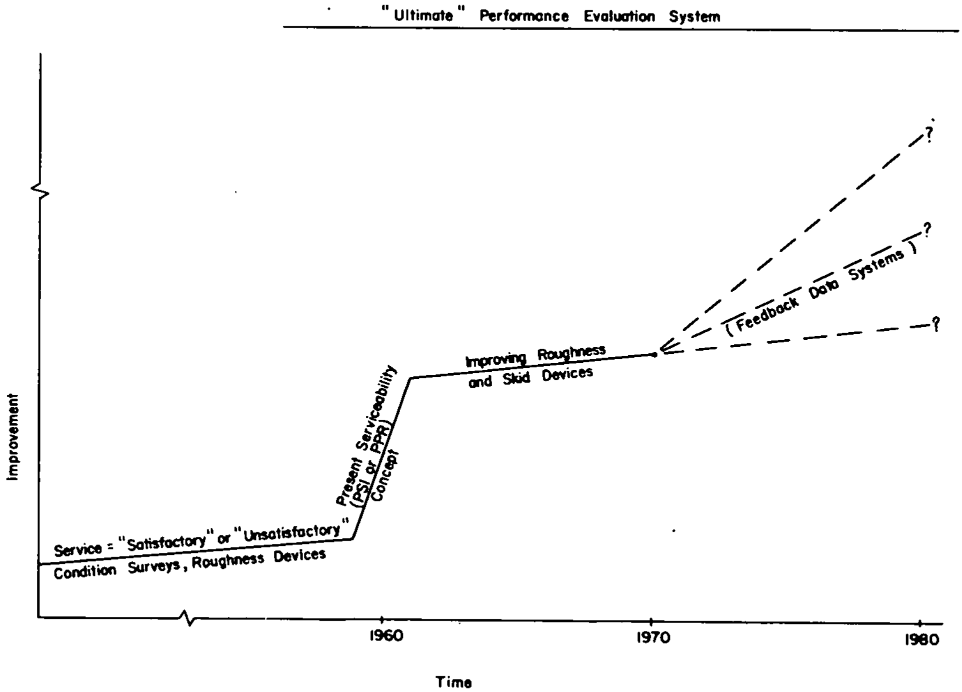

Figure 7. Evolution of pavement performance evaluation methodology.

primarily in the field of psychology. As a consequence, engineers have had little experience in its use or appreciation of its limitations.

The literature on this subject is quite extensive but certain works by Stevens (18), Torgerson (19), and Guilford (20) are particularly useful. In turn, the classification of scales of measurement developed by Stevens (18) is of particular importance to the pavement engineer. This is given in Table 1, including some examples from the paving field. Stevens devised this classification on the basis of the transformations that leave the scale form invariant. This means that certain types of statistical manipulations are not valid for certain scale classes, as discussed more fully by Hall (21). For example, the coefficient of variation for a set of PSI or PPR readings is no more valid or meaningful than expressing one PSI value as a ratio of another (i.e., to divide a PSI = 4 for one section by a PSI = 2 for another section to indicate that one is twice as serviceable as the other is meaningless). The reason, of course, is that the ratio of the standard deviation to the mean (i.e., the coefficient of variation) varies when the location of the mean varies from the zero point.

There are a number of such considerations, and others equally important in scaling techniques, that have been initially and extensively discussed for pavement serviceability evaluation by Hutchinson (22). Unfortunately, it appears that his discussion has either not been carefully considered, or understood, in developing some of the procedures that are in use today. He has explicitly listed the assumptions underlying the AASHO Road Test rating procedure (PSR). He has further pointed out and demonstrated that, in applying such subjective rating procedures to pavements, certain systematic errors can occur and these include the following:

1. Leniency error, i.e., a rater's tendency, for various reasons, to rate too high or too low;

TABLE 1

SCALES OF MEASUREMENT

Scale Class	Scale Function	Mathematical Group Structure and Allowable Transformations	Allowable Statistical Operations	Examples
Nominal	Determining identity or equality	Permutation group [any 1-to-1 substitution, f(x), i.e., $x' = f(x)$]	Frequency, mode	Street names, pavement surface types, car types on a parking lot, pavement evaluation by satisfactory or unsatisfactory
Ordinal	Ranking (i.e., greater or less than)	Isotonic group [any monotonic transformation, i.e., $x' = f(x)$, where f(x) = any increasing monotonic function]	Frequency, mode, median, centile, rank order coefficient of correlation	Ranking pavements or students as to very good, good, fair, poor, or very poor (e.g., ordered sets)
Interval	Determining equalities of intervals	Linear group (any monotonic and linear transformation, i.e., $x' = ax + b$, where $a > 0$)	Frequency, mode, median, centile, rank order coefficient of correlation, mean, standard deviation, skewness, product-moment	Temperature, pavement roughness index, numerical PSI or PPR (e.g., ordered sets of real numbers with an arbitrary zero)
Ratio	Determining equalities of ratios	Similarity group (any linear transformation retaining the zero origin, i.e., $x' = cx$, where $c > 0$)	All statistical operations including coefficient of variation	Time, stress, strain, mass (e.g., an absolute zero exists)

Source: Stevens (18).

2. Halo effect, i.e., a rater's tendency to force a particular attribute rating toward his overall impression of the object; and
3. Central tendency error, i.e., a rater's hesitation to give extreme judgments, thereby tending ratings toward the mean of the rating panel.

Hutchinson (22) has summarized some of the pertinent guidelines that should be observed in constructing rating scales. One of the most important is that cues should apply to a very short and particular range on the continuum to provide raters with definite anchors; further, cues of a general character such as very good, fair, and so on should be avoided. He has also pointed out that the origin and unit of the subjective continuum used by the raters (i.e., the anchoring effects) are functions of the particular experimental situation. This has important implications with regard to interagency compatibility. It also implies that, because a highway department may now usually resurface at a PSI ≅ 3, instead of PSI ≅ 2.5, this does not necessarily mean that the minimum serviceability level has been raised. Rather, depending on the way in which serviceability is determined, it could mean that the origin of the scale has now become 1, instead of 0, to give a rating scale of 1 to 5. There has been some similar Canadian experience where pavements are being resurfaced at a PPR ≅ 6. This in turn could mean that the scale has effectively become 2 to 10. Because the principles of psychophysical scaling tell us that the terminal serviceability should be at about the midpoint of the scale (which was, of course, experimentally verified for both PPR and PSI), the assumption of a shifted origin is certainly possible. We might further expect from well-known observations of economic values (which indicate that utility scales, of ordinal scale status, do no remain stable over time) that subjective rating scales for pavements are not stable with time.

Another problem concerns the number of categories used (i.e., 5 for the PSI approach, 10 for the PPR method, and 6 for the British method). This has also been discussed by Hutchinson (22), and he suggests that, because the optimum varies considerably with the nature of the trait being rated, experimental evaluation is needed to determine this number. He further suggests that one objective criterion is a reliability coefficient, obtained through an analysis of variance, which can be expressed as f follows:

$$r_m = \frac{V_p - V_e}{V_p} \qquad (5)$$

where

r_m = reliability of the mean ratings of a group, where the reliability for a single rater, r_s, is given by Eq. 6;
V_p = variance among pavements; and
V_e = variance of residuals.

$$r_s = \frac{V_p - V_e}{V_p + (K - 1)V_e} \quad (6)$$

where K = number of raters.

Another significant problem concerns the status of the scale representing a particular measurement. For example, the PPR or PSR approach in its basic subjective form only achieves the ranking or ordinal scale status given in Table 1. The subsequent transformation to interval scale status assumes, by its linearity, that each category has equal spacing. Whether such a transformation without the corrections discussed by Hutchinson (22) is valid for pavement evaluation in general or in a particular situation is open to question. For example, Chong (23) has performed a limited experiment in which he indicates that the PPR scale may be somewhat nonlinear for the conditions of his tests. He has also achieved a very high degree of correlation between mean panel ratings by nontechnical people (i.e., truly representative users) and pavement ride as measured by the profilometer of the Ontario Department of Highways.

Not all measurements fall into a single scale class, and those of a more "complicated" nature require separate scales to specify the components of the attribute being measured. Failure to account for such combinations of dimensionality can lead to some simple traps, and consequent faulty conclusions, as illustrated in an excellent example given by Hall (21). A similar trap could exist in pavement serviceability evaluation unless it is explicitly recognized that current methodology is essentially "forcing" a multidimensional vector onto a linear scale. There is, of course, nothing wrong with this when ratings are made by human observers because the weightings of the components are implicit. If the components are all known, the operation of multidimensional scaling, as described by Guilford (20), can sometimes be used to find the weights. Alternatively, if all the components are measurable on interval or ratio scales, as assumed in the PSI model, a multiple regression analysis can be applied to find the composite scale. A comprehensive consideration of some of the principles of multidimensional scaling may lessen some of the problems in basing serviceability on pavement condition-structural surveys or combined condition-structural-ride analyses. Otherwise, the effort may be analogous to mixing apples, oranges, and bananas without realizing that the resultant fruit salad could turn out to be a mess.

In summary the subjective rating approach for pavement serviceability evaluation is apparently simple in concept, easy to use, and fundamentally valid in relation to the user for which the pavement is provided. However, current methodology seems to have evolved largely without proper appreciation of the necessary psychophysical scaling background and possible errors involved. One of the consequences seems to have been certain incompatibilities between and within the results of various highway agencies. The discussion by Hutchinson (22) is still most applicable to this problem.

Approximating Present Serviceability

The present serviceability of pavements can be approximately predicted by several means including condition surveys, roughness measurements, and combined condition and roughness measurements. If we accept the definition of serviceability as directly relating to the user, then the measures can only be considered as approximations. They may, of course, be excellent estimates for many purposes, but it is important to remember that, if used for serviceability evaluations, they are not perfect. The literature is replete with information on roughness and condition survey work, and a review of this material is beyond the scope of this paper. Nevertheless, it is possible to summarize some of the more important aspects as they concern serviceability predictions and possible incompatibilities.

Condition surveys have been performed by many agencies for many years. Their importance in providing information for mechanistic evaluations of structural adequacy and the causes of deterioration cannot be overemphasized. Hudson et al. (24) have pointed this out and have also pointed out that correlation coefficients between pavement ratings and serviceability indexes are only increased by about 5 percent by adding condition data. In other words, detailed condition surveys may be warranted for mechanistic studies but not for serviceability studies. This conclusion, especially where no condition data are used, is not completely accepted by many engineers, as again pointed out in an HRB subcommittee meeting (2). They feel that any serviceability measure cannot be complete without certain "normal" engineering measurements of structural condition or faults.

Hudson et al. (24) have also pointed out that an extremely wide variation exists in the manner in which condition surveys are made, recorded, and analyzed. This is well known by those in the pavement field, who further know that many agencies have filing cabinets full of condition data that never have and likely never will be used. The necessity of developing comprehensive data feedback systems, as previously discussed in this paper, seems to be given added importance by the situation. Standard definitions of items for condition surveys have been suggested by Walker and Hudson (25), but the proliferation of methods in existence has not diminished. It does not seem likely that they will because every agency has its preferences and needs, and perhaps its own biases. If we accept this as a fact, we should also accept as a fact that there is a considerable if not complete lack of compatibility between methods. Consequently, any measure of serviceability predicted solely from condition data, or from combined condition-rideability data, will similarly lack compatibility with other measures. This is not to say that a system evolved by an individual agency, where various condition factors are, say, weighted to produce a composite scale, may not be perfectly workable for that agency. It is meant to say, though, that such a system will only be qualitatively relevant to others.

The use of roughness or ride measurements to predict serviceability can result in much better systems. It can also result in some very significant errors.

Equation 3 demonstrates that roughness was by far the dominant variable in estimating PSI. This has been originally pointed out by Carey and Irick (1) and by others and implies that satisfactory approximations may be obtained by roughness measurements alone. This fact, coupled with the desire of many agencies to obtain mass inventory data on serviceability, has resulted in extensive work over the past few years toward developing rapid, efficient, simple, and low-cost roughness measuring equipment. While work has also continued on the more sophisticated profile data gathering equipment (25), a result has been the development of such devices as the PCA road meter (26) and the Mays road meter (27). A comparison of these 2 devices with 2 other common devices has been made by Phillips and Swift (28). It would be useful to extend their very excellent comparative tabulations of operating features to all commonly used devices. This would likely include for North American practice the following: BPR roughometer, CHLOE profilometer, GMR surface dynamics profilometer (24, 25), RRL profilometer, and rolling straightedge.

A number of agencies have embarked on a program of evaluating these various devices on a systematic basis, especially the PCA road meter, as subsequently discussed. The approach in such efforts should initially ask the questions of purpose of measurement, applicable facility, use of data and so on, and whether the primary interest is in estimating serviceability or in some other purpose. Table 2 gives data for 5 devices to illustrate this approach. The table could be extended, if desired, to include the relationship between the output or transformed output of the device and the end objective of the measurement, i.e., to estimate serviceability.

One of the sources of very large possible errors in current methods of present serviceability evaluation has escaped the attention of many users. The reason is that the errors are "hidden" by using previous correlations or assuming that they are perfect. To illustrate this situation, we can recall that the initial present serviceability equations (1) as previously discussed are multiple regression equations, with correlation coefficients of about 0.8 and 0.9 and a standard error of ± 0.3 to 0.4 PSI units.

TABLE 2

RELATIONSHIPS BETWEEN AREAS OF APPLICABILITY AND PURPOSE OF MEASUREMENT FOR SOME CURRENT ROUGHNESS DEVICES FOR USE IN ESTIMATING PRESENT SERVICEABILITY

Applicability	Classes of Measurement by Purpose[a]		
	Initial Ride	Periodic Ride	Terminal Ride
Facility			
Expressway or primary highway	BPR, GMR, PCA (RRL, CHLOE)	PCA, GMR (RRL, CHLOE)	PCA, GMR, CHLOE (RRL)
Secondary, rural highway	BPR, PCA (GMR, CHLOE, RRL)	BPR, PCA (GMR, RRL, CHLOE)	PCA, GMR (RRL, CHLOE)
County or local rural highway	(BPR, PCA)	PCA	PCA
Use of Data			
Construction monitoring	Yes[b]	-	-
Maintenance programming	-	Yes	-
Inventory and network programming	-	Yes	Yes
Research	Yes	Yes	Yes

Note: The devices listed represent significant current usage; however, many exceptions exist for individual agencies.
[a]Parentheses denote usual application only for control sections or special purposes, although again exceptions exist.
[b]Primary applicability of class of measurements (e.g., initial ride measurements are primarily applicable to construction monitoring).

But this correlation is valid only for the original AASHO Road Test profilometer. That device has been used only for research because it is too big to be of practical use on highways. Consequently, the most widely used PSI equation involves the CHLOE profilometer. This equation was obtained by correlating the roughness estimate from the CHLOE device with the AASHO Road Test device on several pavement sections.

Because there is error in both measurements, the true correlation coefficient of CHLOE profilometers with the panel ratings becomes more erroneous. Recently, users of the new instruments, as subsequently illustrated, have gone one step further by correlating their devices with a CHLOE device, which is not the original but a later model. Thus, if they use the original PSI equations, as most do, they are 3 or 4 steps away from the original correlation data. The true correlation, if it could be evaluated properly, would be quite low and the probable error quite high. The best way to eliminate this is to form a new rating panel and to compare it, over a number of sections, with the instrument of interest. This has been done in Texas for both the CHLOE profilometer (16) and the GMR surface dynamics profilometer (29), and in Canada as subsequently discussed.

The PCA road meter is one of the roughness measuring devices that has recently received considerable and relatively enthusiastic attention by many North American highway agencies. This has largely been a result of the following advantages:

1. Simplicity, ease of operation, and relatively low cost (i.e., it can be produced and installed into a standard automobile for less than $1,000, excluding the price of the automobile);
2. Operating speeds at or near normal traffic speeds;
3. Relatively good repeatibility, although agreement is not unanimous on this; and
4. Feasibility of mass inventory of the pavement network on a recurring basis (i.e., annually, including seasonal surveys).

This device has been well described in the literature, including that previously noted (26, 27) and in other sources (28, 30). It is shown in schematic form in Figure 8.

Some work with the PCA meter has involved a correlation of its sum (Σ) count output with slope variance as measured by the CHLOE profilometer (26, 30), which in turn is used to calculate PSI. This PSI may then be plotted directly versus Σ count of the PCA device and used for pavement serviceability evaluation. Figure 9 shows such an example derived from Brokaw (26), for both portland cement concrete pavements and bituminous flexible pavements. This type of relationship, or a similar one, has received a fair degree of use in estimating pavement serviceability from the PCA road meter.

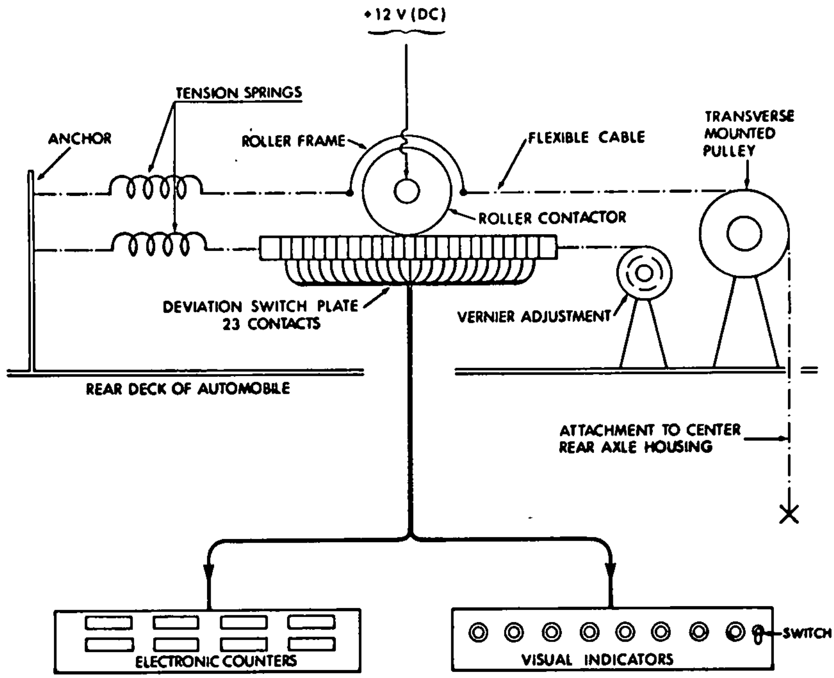

Figure 8. Schematic representation of the PCA road meter.

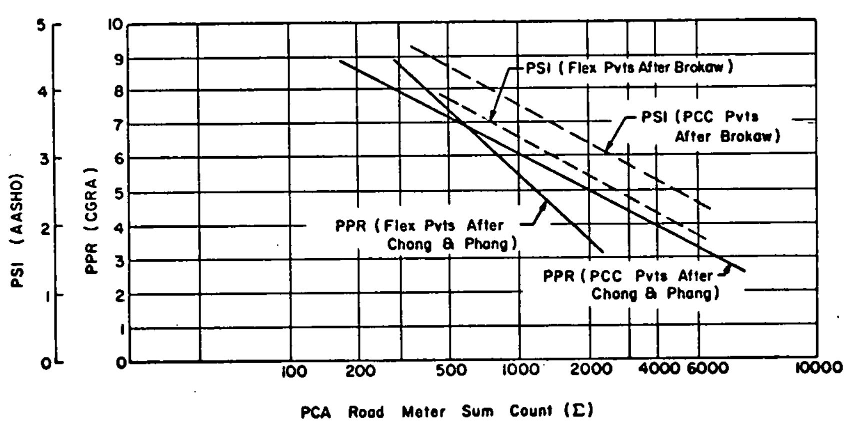

Figure 9. Trend comparison of pavement serviceability as related to PCA road meter Σ count at 50 mph.

Because a series of transformations of this sort may be "filtering out" some of the possibly significant elements of pavement serviceability (or decreasing the reliability of the serviceability estimate), the approach of the Canadian Good Roads Association has been to concentrate on correlating PCA Σ count directly to the present performance rating and to use the CHLOE or RRL devices for calibration. Much of the results of this are yet unpublished but the correlation of PPR with Σ count by Chong and Phang (28), which is indicative of the type of work being conducted, is superimposed on data shown in Figure 9. It is, of course, not valid to compare these directly, but the trend comparison between the two, especially at higher Σ counts, is interesting. Observations such as this, accumulated data, and experience have led the various researchers to at least tentatively conclude that the following limitations or precautions or both should be noted in using the PCA meter:

1. Although roughness is the major factor in PPR, any roughness measurement by itself is inadequate in completely describing PPR; therefore, the PCA meter and similar devices cannot become absolute indicators of PPR. This does not, however, detract from the device's usefulness for approximate mass inventory work.
2. A universally applicable relationship between serviceability and PCA road meter output is not feasible. Rather, each agency must develop its own relationship based on the naturally varying interpretation of pavement performance for its particular jurisdiction. However, for purposes of compatibility (and its implications for widespread use of the results), it may be possible to develop universally acceptable guidelines for operation, including calibration, of the PCA road meter.
3. Frequent calibration of the device is necessary, and this should be done against an instrument of high repeatability.
4. Certain operating conditions regarding the vehicle, such as tire pressure, suspension system wear, gasoline tank level, and number of passengers, and regarding the environment, such as wind and temperature, can adversely affect the results. These have been discussed in more detail elsewhere (27, 28, 29) and are emphasized as quite important.
5. Without certain modifications, the PCA meter as currently used by many agencies (i.e., in its essentially "original" form) may be "too slow" in detecting short-duration, high-amplitude roughness inputs, i.e., those occurring on pavements of poor performance. Consequently, its output may be significantly low in this range.

In summary, though, it appears that the PCA road meter is providing a valuable tool to many highway agencies in obtaining comprehensive performance evaluation information. However, it must be realized that several transformations of PCA data to estimate PSI can also produce some significant errors in the estimate.

TOWARD ACHIEVING BETTER COMPATIBILITY

Highway agencies are increasingly becoming aware of the importance of pavement performance evaluation. These agencies have put considerable effort into developing, applying, and analyzing serviceability measuring schemes. This is certainly encouraging; however, it has also led to a proliferation of methods and data, much of which is unfortunately incompatible with other data. A recent discussion (2) has indicated that this lack of compatibility is essentially dual in nature: external, relating to correlating the results of one agency's work quantitatively with those of another agency; and internal, relating to correlating results and achieving repeatibility within an agency.

It seems apparent from various conferences, papers, and discussions, and from engineering reason based on experience with other structures that better compatibility in pavement performance evaluation is desirable. Consequently, the following suggestions may be useful to highway agencies toward achieving this goal.

1. Performance evaluation of pavements should be established on a planned basis to become an integral part of the overall pavement management system, as previously discussed in this paper.

2. An automated data feedback system is a most useful and perhaps necessary component of the performance evaluation scheme.

3. The existing definitions and underlying assumptions of serviceability and its components should be clearly stated. Moreover, it should be explicitly recognized that serviceability measures, such as those developed at the AASHO Road Test, in Canada, or in Britain, are not ends within themselves; they exist to estimate the road user's opinion.

4. There are a variety of possible errors in subjective evaluations of serviceability. These can be significant, and it is important that the principles underlying subjective rating scale design and analysis are well understood. Because such principles have not been a "normal" part of engineering analysis, they have been somewhat neglected in much of our current methodology. It seems necessary, though, that such understanding be achieved for significant progress toward achieving better compatibility. This paper has presented some pertinent discussion on this problem area and has noted the major references that should be examined by those involved with pavement serviceability analyses.

5. Serviceability measures can be conveniently approximated, for many practical purposes, by condition surveys, roughness measurements, or a combination of both. However, it must be realized that any serviceability predicted from a "unique" method of surveying condition is only qualitatively compatible with any other measure. Predictions from roughness measurements can be quantitatively compatible; but it must be recognized that, because of the nature of subjective evaluations, the interpretation and use of a serviceability measure are unique to the particular region.

6. Roughness data can be used to predict serviceability by the following steps: (a) correlating the output of a simple, fast roughness device to that of a second (perhaps more accurate but more expensive and less efficient) device; (b) correlating the output of the second device to present serviceability ratings (by panels), which themselves have some replication and error; and (c) predicting present serviceability by the easily obtainable output of the first device, which can result in a compounding of errors and perhaps not too reliable a prediction. It seems that better compatibility could be achieved by correlating a device of the first type directly with panel ratings for each region, and then periodically calibrating it with an accurate or repeatable device such as the surface dynamics profilometer.

7. The problems of internal compatibility seem to be often related to lack of correlations and replications. These can perhaps be largely controlled by carefully designed experiments, so that proper statistical analyses may be conducted.

CONCLUSIONS

This paper has been based on the premise that, because a pavement performance evaluation scheme is a vital part of the overall management of pavements, it must be systematically and comprehensively developed. In turn, there are requirements for rationality and compatibility in this development. The paper has been directed toward these considerations and the principal points may be summarized as follows:

1. Performance evaluation can be formally structured as a major phase of the pavement management system. Some general recommendations on the form of this development have been presented in the paper.

2. The distinction between present serviceability of a pavement and performance, as the serviceability-age history, is important. Performance and its value implications have been defined as the principal outputs of a pavement and its management system.

3. The principal approaches to evaluating or measuring present serviceability of pavements have been reviewed in the paper. It has been emphasized that these all are intended to predict the road user's opinion and are not ends within themselves. It has been further pointed out that purely subjective evaluations are not completely acceptable to many engineers, who wish to include in serviceability certain structural condition parameters.

4. The limitations and assumptions of current approaches to evaluating pavement serviceability have been discussed. It was suggested that some of the problems, errors, and lack of compatibility among various serviceability measures are due to a lack of understanding of the basic principles of subjective rating procedures.

5. Procedures for approximating present serviceability through condition surveys, roughness measurements, or a combination of both are widespread. However, it has been pointed out that many of these are only qualitatively relevant or compatible to others and that significant errors are possible in some of the transformations that are made.

6. Several suggestions were made toward achieving better compatibility, both interagency and intra-agency, among various serviceability-measuring techniques and results. These relate to precise definitions, better understanding of subjective rating principles, rigorous error analyses, and carefully designed experiments.

ACKNOWLEDGMENTS

The advice and research of many of our colleagues have been used in this paper, and we sincerely appreciate their contributions. Acknowledgment and appreciation must also be extended to our own research sponsors who have provided funds over the years; these include Gulf Oil Canada, Ltd., the Texas Highway Department, the Federal Highway Administration, the National Research Council of Canada, and the Ontario Department of Highways.

REFERENCES

1. Carey, W. N., Jr., and Irick, P. E. The Pavement Serviceability-Performance Concept. HRB Bull. 250, 1960, pp. 40-58.
2. Hudson, W. R., and Haas, R. C. G. Compatibility Between Pavement Roughness and Serviceability Measuring Systems. Minutes of Joint Subcommittee Meeting of HRB Committees DB-4 and DB-5, Jan. 12, 1970.
3. Haas, R. C. G., and Hutchinson, B. G. A Management System for Highway Pavements. Paper presented to Australian Road Research Board, Sept. 1970.
4. Hudson, W. R., et al. Systems Approach to Pavement Design. NCHRP Project 1-10, Interim Rept., March 1968.
5. Scrivner, F. H., McFarland, W. F., and Carey, G. R. A Systems Approach to the Flexible Pavement Design Problem. Texas Transportation Institute, Research Rept. 32-11, 1968.
6. Wikins, E. B. Outline of a Proposed Management System for the Canadian Good Roads Association Pavement Design and Evaluation Committee. Proc., Canadian Good Roads Assn., 1968.
7. Hutchinson, B. G., Haas, R. C. G. A Systems Analysis of the Highway Pavement Design Process. Highway Research Record 239, 1968, pp. 1-24.
8. Haas, R. C. G., and Anderson, K. O. A Design Subsystem for the Response of Flexible Pavements at Low Temperatures. Proc., Assn. of Asphalt Paving Technologists, 1969.
9. Haas, R. C. G. A Systems Framework for Roadway Materials Problems. Paper presented to Second Inter-American Conf. on Materials Technology, Mexico, Aug. 1970.
10. The AASHO Road Test: Report 5—Pavement Research. HRB Spec. Rept. 61E, 1962.
11. Manual on Pavement Investigations. Canadian Good Roads Assn., Tech. Publ. 11, 1959.
12. Proc. Symposium on Pavement Design and Evaluation. Canadian Good Roads Assn., 1960 and 1961.
13. Pavement Evaluation Studies in Canada. Proc., Internat. Conf. on Structural Design of Asphalt Pavements, Univ. of Michigan, Ann Arbor, 1962.
14. A Guide to the Structural Design of Flexible and Rigid Pavements in Canada. Canadian Good Roads Assn., Sept. 1965.

15. Field Performance Studies of Flexible Pavements in Canada. Proc., Second Internat. Conf. on Structural Design of Asphalt Pavements, Univ. of Michigan, Ann Arbor, 1967.
16. Scrivner, F. H. A Report on a Modification of the Road Test Serviceability Index Formula. Texas Transportation Institute, Research Rept. 32-1, May 1963.
17. Chipperfield, E. H., and Welch, T. R. Studies on the Relationships Between the Properties of Road Bitumens and Their Service Performance. Proc., Assn. of Asphalt Paving Technologists, Vol. 36, 1967.
18. Stevens, S. S. Measurement, Psychophysics, and Utility. In Measurement: Definitions and Theories (Churchman, C., and Ratoosh, P., eds.), John Wiley and Sons, New York, 1959.
19. Torgerson, W. S. Theory and Methods of Scaling. John Wiley and Sons, New York, 1960.
20. Guilford, J. P. Psychometric Methods, McGraw-Hill, New York, 1954.
21. Hall, A. D. A Methodology for Systems Engineering. Van Nostrand, New York, 1962, ch. 8.
22. Hutchinson, B. G. Principles of Subjective Rating Scale Construction. Highway Research Record 46, 1964, pp. 60-70.
23. Chong, G. A Road Rideability Rating Experiment. Ontario Department of Highways, Rept. IR19, March 1968.
24. State of the Art of Pavement Condition Evaluation. HRB Spec. Rept. 95, 1968, pp. 49-68.
25. Walker, R. S., and Hudson, W. R. A Road Profile Data-Gathering and Analysis System. Highway Research Record 311, 1970, pp. 36-54.
26. Brokaw, M. P. Development of the PCA Road Meter: A Rapid Method for Measuring Slope Variance. Highway Research Record 189, 1967, pp. 137-149.
27. Phillips, M. B., and Swift, G. A Comparison of Four Roughness Measuring Systems. Highway Research Record 291, 1969, pp. 227-235.
28. Chong, G., and Phang, W. A. The PCA Road Meter—Measuring Road Roughness at 50 Miles per Hour. Ontario Department of Highways, Rept. IR 26, 1968.
29. Roberts, F. L., and Hudson, W. R. Present Serviceability Equations Using the Surface Dynamics Profilometer. Paper presented at Western Summer Meeting and published in this Special Report.
30. Law, S. M., and Breckwoldt, E. J. PCA Road Meter Correlation Study. Louisiana Department of Highways, Interim Rept. 2, Oct. 1969.

Discussion

M. P. Brokaw, Portland Cement Association

The authors have provided an excellent service by outlining the important parts of a rational and compatible system for evaluating pavement performance.

Their references to the PCA road meter should have included all devices that can be related to serviceability ratings, either directly or through an intermediate transformation, such as the CHLOE profilometer. Data from several sources, much of which was presented in my paper (31), show a high level of repeatability and an excellent correlation with both the panel serviceability ratings and the part of serviceability index dependent on pavement roughness.

An intermediate correlation with the CHLOE profilometer, and hence with the AASHO profilometer and AASHO serviceability ratings, was used for the PCA road meter as a temporary step in lieu of having a direct correlation with serviceability ratings. This is a choice for the operating agency; and some have found it desirable to make new correlations with their own rating panels while others have been satisfied with the results per se.

Apprehension expressed by the authors may have been the result of faulty interpretation of data presented in Figure 7 of their paper. First, the PPR curves attributed to Chong and Phang have been reversed as to identification between rigid and flexible pavements; and second, PSI (both rigid and flexible) attributed to Brokaw is only that part of PSI that can be related to pavement roughness. Both rigid and flexible PSI's should have been reduced by appropriate deductions for cracking, patching, and rutting before they were compared with Chong and Phang. This difference was made clear in Figures 10 and 11 of the Chong and Phang report (28).

The authors are also concerned by "hidden" errors resulting from an intermediate transformation between road meter Σ count and serviceability ratings. In a strict statistical sense there should be a small increase in standard error of estimate. This is true because individual deviations from regression are random rather than systematic, and thus chances for accumulation of maxima are minimal.

Data from the Chong and Phang report (28) are used to illustrate this point. In Ontario, PPR was related to profilimeter log 'q', log 'q' was related to road meter Σ count, and PPR was related to road meter Σ count.

Thus (from Chong and Phang),

$$\begin{aligned} \text{PPR} &= 16.2037 - 5.9878 \log \text{'q'} \\ \text{Log 'q'} &= 1.0126 \log \Sigma \text{ count} - 1.2795 \end{aligned}$$

transformed,

$$\text{PPR} = 23.8651 - 6.0632 \log \Sigma \text{ count}$$

or direct (from Chong and Phang),

$$\text{PPR} = 25.0720 - 6.4873 \log \Sigma \text{ count}$$

When observed Σ count values are used in these equations and computed PPR values are compared with panel PPR's, the standard error of estimate for transformed PPR is 0.64 and for direct PPR it is 0.62.

The small difference in standard errors agrees with the hypothesis and does not support the idea of hidden error that can be consequential in practical pavement evaluation.

Reference

31. Brokaw, M. P. A 5-Year Report on Evaluation of Pavement Serviceability With Several Road Meters. Paper presented at the Western Summer Meeting and published in this Special Report.

Closure

R. C. G. Haas and W. R. Hudson

We would like to thank Mr. Brokaw for the corrections he indicates are due in Figure 7. This in no way changes the import of the comparison as he indicates.

Concerning "hidden" (cumulative) errors, it must be reemphasized that the intermediate transformations are estimated intermediate transformations. Brokaw insists that there will be a small increase in the standard error of estimate because the deviations from the estimated transformations are random; on the contrary, the increase in the standard error of estimate is directly proportional to the magnitude of the deviations from the estimated transformations and only indirectly related at best to the randomness of these deviations. Furthermore, any "lack of fit" in the data will result in systematic not random errors. In his example, Brokaw compares correlations of identical sets of sections and data where the deviations among the observed log 'q' and the predicted log 'q' may be expected to be small. It does not follow that such deviations will be small for all sets of experimental data, particularly on data taken at different times on different pavements.

The idea of hidden cumulative errors can be expressed most simply by looking at the mathematics involved, taking the estimated relationships for PSI, and moving through a correlation between the AASHO profilometer and the CHLOE profilometer to a third device as follows: From

$$PSI = A_0 + A_1 \log(1 + \overline{SV}) \pm e_1$$

and a correlation between CHLOE and AASHO profilometers,

$$\log(1 + \overline{SV}) = B_0 + B_1 [f(CHLOE)] \pm e_2$$

it follows, by direct substitution, that

$$PSI = C_0 + C_1 [f(CHLOE)] \pm A_1e_2 \pm e_1$$

It can be seen that the importance of the terms A_1e_2 and e_1 is not their randomness but their magnitude. If in addition we correlate a third type of roughness meter,

$$f(CHLOE) = D_0 + D_1(\overline{RM}) \pm e_3$$

it follows that

$$PSI = F_0 + F_1(RM) \pm [C_1e_3 \pm A_1e_2 \pm e_1]$$

The total error is that in brackets, $[C_1e_3 \pm A_1e_2 \pm e_1]$, not e_3 alone.

Contrary to the Brokaw implication, we feel that the PCA meter is a very useful device when properly calibrated and applied. A 3-stage calibration is still an undesirable approach.