Visual Scales of Pavement Condition: Development, Validation, and Use

DAVID T. HARTGEN, JOHN J. SHUFON, FRANK T. PARRELLA, AND K.-W.P. KOEPPEL

This paper describes the procedures used by the New York State Department of Transportation to develop visual scales for assessing the condition of pavements in New York State. The paper describes the use of a psychometric scaling method known as Q-Sort to develop visual scales of pavement surface condition and the condition of base material that underlies the pavement. A panel of eight experts from the department's operating units reviewed more than 50 photographs of pavements and rated them on a scale of 1-10 for surface and base condition. Subsequent analysis produced a set of photographs representative of each position on the scales. Test-retest reliability was conducted by using the same eight judges two months later to demonstrate the replicative nature of the scales and their stability over time. These scales were then used to train scorers from the department's 11 regional offices, and scoring of the entire state highway system was conducted. Results were available within several weeks and summaries were made of the condition of the highway system in 1981. The paper concludes that the use of visual scales in highway scoring is a practical and accurate method of assessing pavement condition. It also concludes that, although the use of such scaling procedures has lagged in pavement analysis, their use should be expanded in the future as continuing assessments of highway condition are needed.

Of all the issues that confront transportation analysts in the 1980s, perhaps the most significant and possibly the most difficult to deal with is the issue of the deterioration of transportation systems. The condition of highway pavements in the United States is deteriorating as these systems age and are subjected to greater and more severe traffic loads (1). In some cases pavements constructed in the 1930s and 1940s have deteriorated to the point where their structural integrity is now impaired and traffic is slowed. To make matters worse, Interstate highways constructed in a relatively short time span in the 1950s and 1960s are now approaching the end of their design lives and will require significant rehabilitation to protect the initial investments. Funding for restoration and rehabilitation of the Interstate system is not enough to keep up with needs. This shortage comes at precisely the same time as increases in construction costs, declining revenues, and inflation have constrained governments' ability to finance the necessary improvements (2). Given projected deterioration rates of highways (1) and the likelihood that current revenues will decline as gasoline use falls (3), significant attention must be paid to this problem soon.

A periodic accurate assessment of the condition of various types of highway pavements is of critical importance in determining highway needs. The primary reason for accurate and timely information on condition is that it leads us to better decisions on investment policy. These decisions are crucial to developing a program that will provide the most service to the public per dollar invested. To achieve this goal, such assessments must be conducted periodically so that we may evaluate the changes in highway condition over time. Such assessments must be conducted and reported rapidly if their use is to be maximized.

The purpose of this paper is to describe recent activities at the New York State Department of Transportation (NYSDOT) to develop and implement straightforward windshield survey procedures for assessing pavement condition in a rapid, consistent, and accurate fashion. These procedures parallel the visual assessment procedures developed by Arizona and Ontario (4) and the federal highway performance

monitoring system scoring methods (5), as well as the early American Association of State Highway Officials (AASHO) scaling procedures (6). This paper describes how such visual scales can be developed, validated, and used. Results of the 1981 survey are also presented. The primary conclusion is that such methods are reliable and capable of producing the necessary information in rapid fashion with sufficient accuracy.

DEVELOPMENT OF VISUAL CONDITION SCALES

A scale is a sequence of items such that each item, or the total group, has a specific numerical level associated with it. This is not an arbitrary assignment but one in which certain properties of the scale (e.g., origin, distance, and order) are preserved. Although we are all familiar with common examples of certain scales (e.g., rulers and thermometers) other scales can be constructed from words, photographs, and statements. Procedures for developing scales are well developed (7,8). A complete discussion of such procedures is outside the scope of this paper; the reader should realize that numerous methods of construction of such scales are treated extensively in the psychology literature.

In the area of transportation, visual and verbal scales have been used for a number of years. Verbal scales have been developed to measure the characteristics of transportation services (9), such as comfort, convenience, and cost. In a recent review of such studies, Tischer (10) found more than 100 examples of the use of transportation attitude scales constructed from verbal assessments. These scales relate primarily to public perceptions of the characteristics of transit services but have also been applied to visual assessments of highway environment, perceptions of highway impacts, and opinion studies on highway financing. The New York State Department of Transportation has also developed and used many such scales.

The procedure used to develop visual scales for highway condition is straightforward. Scales were developed by using department experts (judges). Regional personnel were then trained in their use, condition assessment was conducted, and summaries were prepared. Scales of pavement condition focused on two measures of pavement quality:

1. The condition of the pavement surface and
2. The condition of the base that underlies the pavement surface, described as the rupture and displacement rating.

The scales were developed by using a modification of a technique known as Q-sort (8). This is a standard method in psychometrics that has seen wide use in the development of visual scales. Basically, the method involves sorting or rating a number of preselected items (e.g., photographs and words) by some criteria (in this case, photographs of highway pavements are rated according to condition). The initial rating process is done by judges who apply expert knowledge to the sorting process. From a large number of sorted photographs, a few photographs are then selected, by statistical or other

Figure 1. Photographic scales of pavement condition.
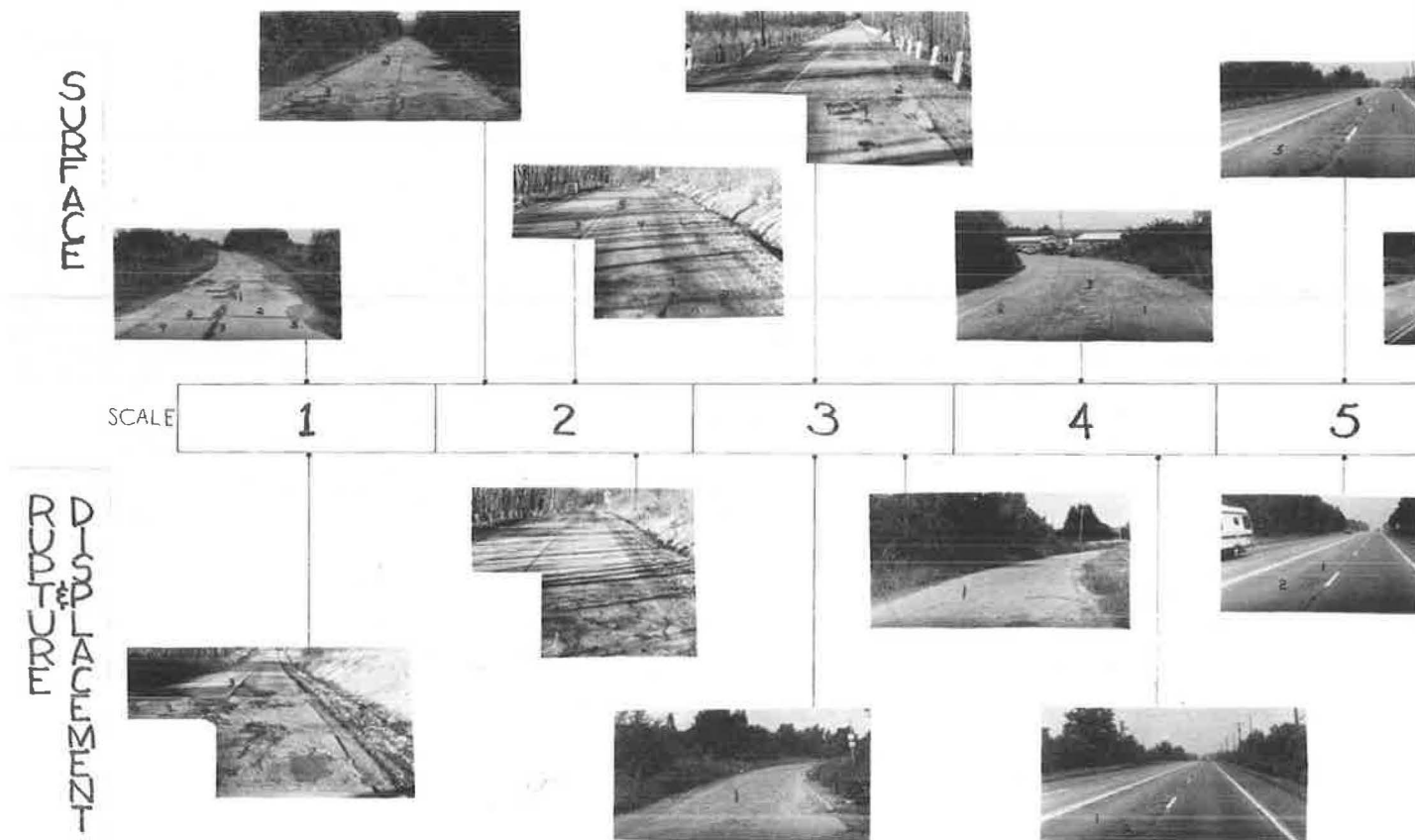
means, to form the final scale. The procedure is then repeated (usually after several months) and test-retest reliability estimates are prepared. Once developed and validated, the resulting scale, which consists of a small set of photographs to represent the scale points, is then used to train regional scorers in rating highways. If constructed carefully, the scales may then be used with confidence by others in scoring the condition of highways.

In the NYSDOT application of Q-sort, a panel of experts from the department's staff rated preselected photographs. A total of eight judges participated, representing the major program areas of the department [planning, project development, programming, soil mechanics, engineering research (2), highway maintenance, and highway construction]. The judges' qualifications included engineering degrees, engineering licenses, and 10-20 years of pavement-related experience.

The first step in the process was to develop a set of verbal scales of highway conditions. NYSDOT had previously developed 10-point verbal scales of highway condition that had been used for many years in the periodic sufficiency series of pavement condition evaluation. Through discussions with the judges and other experts, these verbal scales were carefully reviewed and refined. The results of these technical discussions are given in the list below, which gives the actual verbal descriptions prepared for surface rating.

10--Visually, pavement should show no deviations from a smooth surface. Pavement probably was recently constructed or reconstructed within the past year or two.

9--Facilities should have no cracks or patches.

Pavement probably was recently resurfaced within the past year or two.

8--Pavements in this category give an excellent ride and exhibit only a few signs of surface deterioration. Flexible pavements may start to show slight evidence of rutting and fine random cracks. Rigid pavements may start to show evidence of surface deterioration such as minor cracks, slight joint spalling, or scaling.

7--Pavements in this category still give a good ride but are starting to show definite signs of surface deterioration. Flexible pavements show evidence of slight rutting, random cracking, and possibly some raveling. Rigid pavements show evidence of joint spalling, scaling, or minor cracking.

6--The riding qualities of pavement in this category are noticeably inferior to those of new pavements. Surface defects of flexible pavements may include rutting, cracking, and possible raveling; patching is also apparent. Surface defects of rigid pavements may include slight joint spalling, faulting, or cracking, and patching is apparent.

5--The riding qualities of pavements in this category are noticeably inferior to those of new pavements and may be barely tolerable for high-speed traffic. Surface defects of flexible pavements may include moderate rutting, cracking, and raveling and could have frequent patching. Surface defects of rigid pavements could include joint spalling, some faulting, moderate-to-heavy cracking, and frequent patching.

4--Pavements in this category have deteriorated to a point where resurfacing may be required. Rideability, even at slow speeds, is impaired. Surface defects of flexible pavements could include such signals as frequent rutting, cracking, raveling, and

6   7   8   9   K
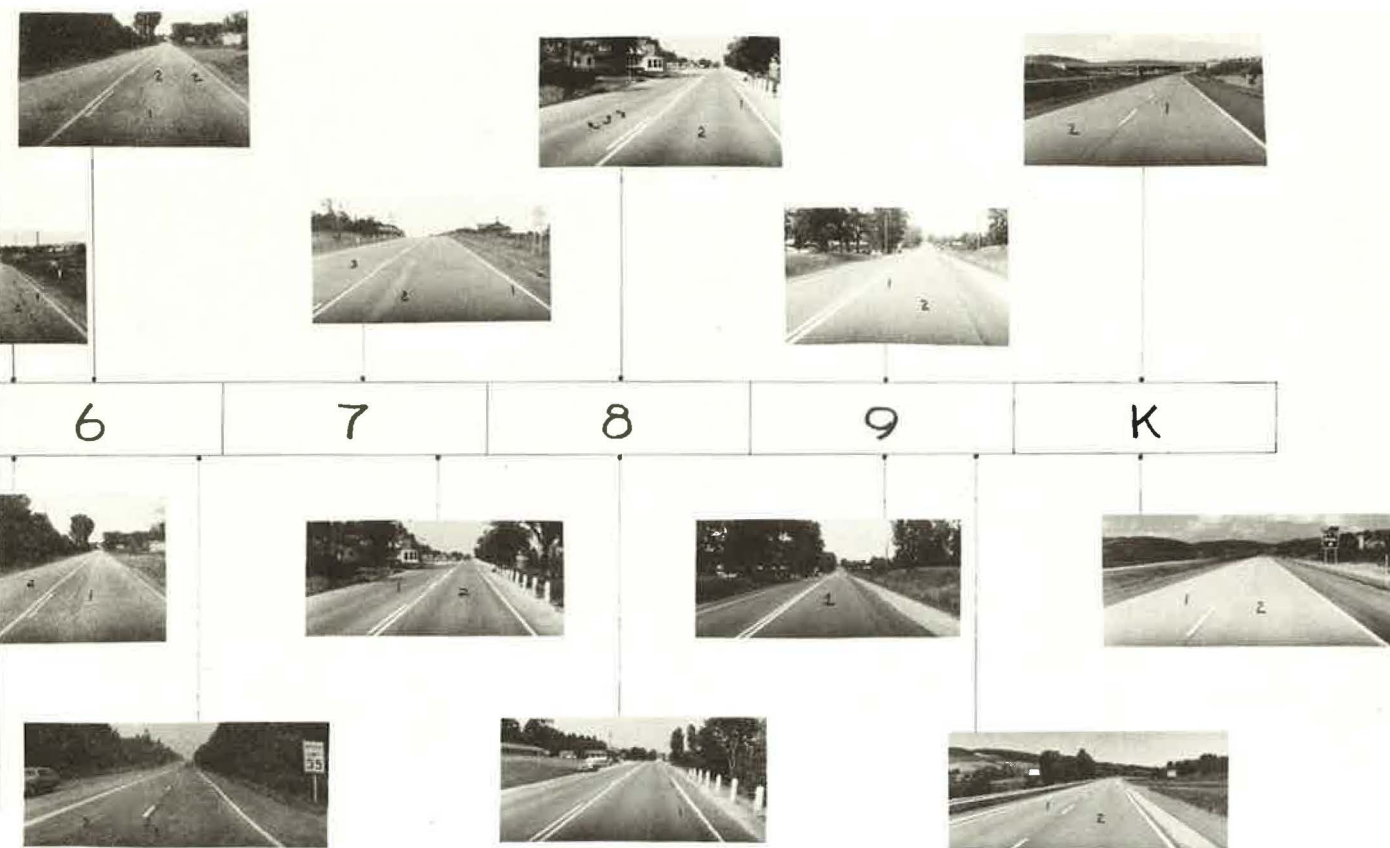
patching. Surface defects of rigid pavements might include frequent joint spalling, moderate faulting, severe cracking, and frequent patching.

3--Pavements in this category have deteriorated to a point where resurfacing is required immediately. Rideability at any speed is so impaired that the motorist will experience discomfort. Surface defects of flexible pavements may include severe and frequent rutting, cracking, raveling, and patching. Surface defects of rigid pavements will include severe and frequent scaling, joint spalling, faulting, cracking, and patching.

2--Pavements in this category are in an extremely deteriorated condition and may even require complete reconstruction. Motorists experience discomfort and traffic will slow down.

1--Pavements in this category are in an extremely deteriorated condition and are in need of some immediate corrective action. These facilities could be considered impassable at posted speeds.

The verbal descriptions of rupture and displacement rating are given in the list below.

10--Riding quality of roadways in this category should be excellent, and there will probably be no indications of any subsurface shifting. Facilities newly constructed within the last year or two generally will fall into this category. Close attention should be directed to evaluation of any riding quality that could be attributable to a displacement of the base material, which can happen even on recently constructed facilities and would preclude such a facility from receiving a 10 rating. These modifiers also apply to the 9 rating as well.

9--Riding quality is also excellent, with little

or no indications of subsurface problems. Facilities reconstructed or rehabilitated within the last year or two fall into this category.

8--Facilities fall into this category if there is no evidence of base or subbase deterioration, as reflected by the pavement surface.

7--Roadways in this category are beginning to show signs of rupture and displacement caused by roadbed movement. Flexible and rigid pavements may show slight evidence of longitudinal or transverse cracking.

6--Roadways in this category show definite but infrequent signs of distress (e.g., cracks) caused by roadbed movement or inadequate roadbed support. Flexible pavements show evidence of moderate longitudinal and transverse cracking. Rigid pavements show signs of moderate longitudinal, transverse, or random cracking.

5--Roadways in this category show occasional signs of distress caused by roadbed movement or inadequate roadbed support. Flexible pavements show evidence of severe longitudinal or transverse cracking and may be beginning to show evidence of block or reflection cracking. Rigid pavements show evidence of severe longitudinal, transverse, or random cracking and may start to show signs of corner or diagonal cracking caused by a loss of the foundation material under the slab.

4--Roadways in this category show frequent signs of distress caused by roadbed movement or inadequate roadbed support. Flexible pavements may show signs of moderate block or reflection cracking and slight alligator cracking. Rigid pavements show evidence of moderate corner or diagonal cracking caused by a loss of foundation material under the slab.

3--Roadways in this category show frequent signs

Table 1. Test-retest correlations of judges' ratings of pavement condition from 50 test photographs.

Judge	Surface			Rupture-Displacement		
	r	Slope	Inter-cept	r	Slope	Inter-cept
A	0.974	0.953	0.294	0.968	1.002	-0.320
B	0.975	0.933	0.522	0.911	0.805	0.908
C	0.972	1.030	-0.287	0.784	0.741	1.240
D	0.962	1.003	0.122	0.969	0.987	-0.190
E	0.956	1.230	-2.420	0.871	0.700	2.190
F	0.953	1.024	-0.839	0.937	0.836	0.599
G	0.947	1.053	-0.365	0.852	0.901	0.874
H	0.942	0.910	0.528	0.853	0.834	0.789
Overall	0.950	0.975	0.009	0.902	0.851	0.722

Table 2. ANOVA results of test-retest of surface and base scales by expert judges.

Source of Variation	Sum of Squares	Variance (%)	df	Mean Square	F-Tests[a]	
Surface						
Photograph	5321.9	86	9	591.3	0.00	
Judges	147.0	2	7	21.0	0.00	0.01
Date	4.2	<<0.1	1	4.2	0.02	0.07
Interaction terms	234.8	4	142	1.7		
Error	520.0	8	640	0.8		
Total	6227.9	100	799			
Base						
Photograph	4353.8	67	9	483.8	0.0	
Judges	239.2	4	7	34.2	0.0	0.01
Date	4.5	<<0.1	1	4.5	0.16	0.07
Interaction terms	365.1	6	142	2.6		
Error	1479.7	23	640	2.3		
Total	6442.3	100	799			

[a]Tail probabilities.

of distress and may even show occasional roadway failures due to roadbed movement or inadequate roadbed support. Flexible pavements show signs of severe block and reflection cracking, and alligator cracking is especially visible through wheel-track rutting. Rigid pavements show evidence of severe corner and diagonal cracking and occasional pumping or faulting due to uneven roadbed support.

2--Roadways in this category are in an extremely deteriorated condition and may require complete reconstruction. Flexible pavements show frequent evidence of severe alligator cracking with rutting and possibly occasional bumping. Rigid pavements show evidence of frequent and severe cracking, pumping, and faulting.

1--Roadways in this category are in an extremely deteriorated condition. Flexible pavements show frequent evidence of severe alligator cracking with rutting or severe bumping. Rigid pavements show evidence of frequent and severe cracking, faulting, and pumping.

Technical definitions of terms (e.g., random cracks) were drawn from TRB's glossary (12); TRB's glossary was also used to illustrate specific distress signals of pavement deficiencies.

While these verbal scales were being developed, a set of photographs that represents a wide variety of highway conditions was prepared. The data base for this set of photographs is the department's photolog system, which is a visual record of the condition of state highways. Out of the very large number of photographs available for selection, 50 were selected by the researchers to span the range of highway condition, ranging from 1 to 10, as defined in the above verbal scales. No attempt was made to make the photographs representative (in a statistical sense) of the entire state highway system, since the purpose of the exercise is to develop a scale that is then used by others to score road condition.

Each of the judges was then provided with copies of the final verbal scales and copies of the 50 photographs. Each judge was asked to rank each photograph on the 1-10 scales, once for pavement surface condition and once for rupture and displacement (base) condition. No instructions were provided concerning how this ranking or scoring should be done other than that each photograph must be assigned a position on the scale. The resulting distribution of the photographs for each judge represents his or her perception of the degree to which the technical descriptors of highway conditions, as described in the verbal scales, are reflected in the visual photographs that represent the pavements. On completion of this assignment by each judge, the assigned category for each photograph was recorded.

The photographs were then reviewed by the analysis team and arranged according to the score position in which the greatest number of judges ranked the photograph (i.e., modal score). This resulted in each of the 10 categories containing a small group (3-7) of candidate photographs. The photographs within each category were then carefully reviewed by the analysis team for consistency of rating, low variance of rating, as well as an ability to show clearly typical problems with the pavement. Based on this assessment, the analysis team selected final photographs to best represent the categories. The positioning of the photographs and the final layout of the scales are shown in Figure 1.

VALIDATION

The scale established in this way needs some further analysis before it can be applied with confidence in the field. Questions to be answered include,

1. Are eight judges sufficient to establish a scale?
2. Are there significant differences between judges? and
3. Are the scales stable over time; i.e., are the judges able to reproduce their original ratings some time after the pictures were rated initially?

Research in psychology has shown that the number of judges required to establish a scale is surprisingly small. A recent study of highway condition in England (11) tested scales developed by different numbers of judges and found that scales developed with as few as 8 judges were capable of estimating the position of a given highway pavement to within one scale position on a seven-point scale with 95 percent confidence. Further, it was shown that once the number of judges exceeds about 15, little improvement in reliability is subsequently obtained.

To answer the questions concerning scale reliability over time and differences between judges, the eight experts were asked to rescore the 50 selected photographs two months after the initial scoring. The exact same procedures were used in this retest. A simple correlation-regression analysis of the two scoring results is given in Table 1 (a perfect line would have a correlation of 1.0, a slope of 1.0, and an intercept of 0.0). The overall correlation coefficient (r) for both retests is high, 0.95 for surface, 0.90 for base (rupture-displacement). The linear regressions show a good

Table 3. New York State touring routes, 1981 system condition.

Condition	Level	Surface		Base	
		Lane-Miles	Percent	Lane-Miles	Percent
Excellent	10	1 188		1 115	
	9	1 439	6.6	1 442	6.4
Good-to-fair	8	8 381		6 473	
	7	13 487	80.4	10 610	72.6
	6	10 012		11 712	
Poor	5	3 828		5 641	
	4-1	1 325	13.0	2 668	21.0
Total		39 661		39 661	

degree of conformity among judges. The consensus of literature in psychological scaling (8) is that such results are considered excellent and indicate high reliability in the use of the final scales.

Table 1 also shows that the correlations are generally higher and more consistent for the surface ratings than for the rupture and displacement ratings. This suggests that it is more difficult to infer the state of the base from a photograph that primarily shows the surface. Therefore, judgment and experience would enter more heavily in assessing base conditions than surface ratings. In view of these concerns, a second analysis of the ratings was performed to test for differences attributable to differences among judges. This analysis was conducted by using analysis of variance (ANOVA), and the results are given in Table 2. Quite clearly, for both surface and base ratings, we find (expectedly) that most of the variance is explained by the photographs themselves. The variation attributable to judges or to time between tests (date) is considerably smaller. A standard F-test for the comparison of sources of variation shows all three sources to be significantly larger than the residual variation left unexplained. The same test shows that the variation between photographs, in turn, is significantly larger than the variation attributable to judges or time elapsed. Finally, for both scales the level of significance of other sources of variation drops off sharply. These results mean that, in actual use, the differences among scorers is likely to be small and can probably be eliminated with joint training of field scorers. The scales are also stable over time, and hence can be used (with training) in subsequent scoring with confidence that changes of pavement condition with time will not be confused with changes in the scale itself.

The first of these implications was fully addressed through training. The potential time consistency problem has not been fully addressed yet; however, it appears that this problem is slight for the eight expert judges undertaking this effort as a two-times only event. We are thus led to believe that it may be even slighter for the scoring teams that actually rate the state's road system, as these teams often have several years of experience at this task.

USE OF VISUAL SCALES IN HIGHWAY SCORING

The primary purpose of these visual scales is to enable a rapid and consistent assessment of the condition of New York State's highways. It was important, therefore, to ensure that all individuals conducting the 1981 scoring, from each of the department's 11 regional offices, be trained to assess highways in a consistent manner. To achieve this, a training session was held in the main office of NYSDOT in late April 1981. As part of the training, personnel were instructed in scoring by actually

working with films of highways in various stages of condition.

A series of photolog films were shown that simulated travel over routes of the state highway system. Each film was scored and then discussed, and the individual scorers were then instructed on how to improve their ability to make judgments by using the visual scoring materials. Additional films of highways were then shown, and convergence of scoring among the different regional individuals was then evaluated. Training was repeated until convergence of scoring was reasonable.

At the conclusion of the training session each survey team was given computer-generated field sheets with locational information and physical characteristics of each highway link to be assessed listed contiguously by route. The raters traveled each highway section and coded the appropriate surface and base scores on the preprinted forms. Once completed, the field scoring sheets were transmitted to the main office for processing. Although the regional offices were given 14 weeks to complete the inventory (May-mid-August 1981) most of the regional administrators allowed the scoring crews about a one-month window in existing workloads to perform the survey.

Regional staff cooperation was excellent in reporting the results of the survey. In return, main office staff edited, processed, and tabulated the data as they came in and immediately provided each region with computer printouts in various formats. Statewide summaries were available by October 1981.

SURVEY RESULTS

Table 3 gives the results of the 1981 scoring. The overall pavement condition of the state touring route system is generally quite good, and road surfaces are in better condition (an average of level 6.82) than road bases (an average of level 6.53). Approximately 87 percent of road surfaces are in fair or better condition compared with about 79 percent of road bases. This is expected because recent department focus has centered on resurfacing projects rather than on more costly reconstruction activities. Of particular concern are highways in the lower end of the good-to-fair category. About 26 percent of the system is at level 6, just above the poor level. Without adequate attention, these highways will soon deteriorate to the poor range. Thus, although the overall system is in good shape, a significant share of facilities needs attention. Rigid pavements are in better shape, on the average, than flexible or overlay pavements (Table 4). The Interstate system is in the best overall condition (Table 5). The surfaces of the Interstate system average 7.51, bases 7.34, about a point above averages for the other federal-aid systems. Most of these Interstate highways (Thruway is excluded) are rigid and were constructed more recently than roads in the other systems. The average condition of urban roads is slightly better than that of rural roads.

CONCLUSIONS

The use of visual scales for assessing pavement condition is both practical and feasible. The development of such scales is straightforward and scoring may be completed rapidly. When constructed according to psychometric methods, the reliability of such scales is generally good.

Training is an integral part of the development and use of such scales. Training can be organized in such a way that the individual scorers from

Table 4. Condition of New York State touring routes by pavement type, 1981.

Condition	Level	Flexible		Overlay		Rigid	
		Surface	Base[a]	Surface	Base[a]	Surface	Base[a]
Excellent	10	345	313	420	388	423	415
	9	561	549	696	699	183	193
Good-to-fair	8	2 911	2 143	3 397	2 464	2072	1867
	7	4 158	3 153	5 998	4 822	3330	2635
	6	3 363	3 797	4 753	5 494	1896	2421
Poor	5	1 178	1 966	1 774	2 481	876	1184
	4-1	460	1 056	551	1 232	313	380
Total lane-miles		12 976	12 976	17 589	17 589	9093	9093
Mean score		6.83	6.46	6.77	6.44	6.90	6.73

[a] Differences in base lane-mile totals are due to rounding.

Table 5. New York State touring routes—1981 condition by federal-aid system.

System	Roadway	Lane-Miles	Avg Condition	
			Surface	Base
Interstate	Urban	2 155	7.46	7.34
	Rural	1 917	7.56	7.34
	Avg		7.51	7.34
Primary	Urban	6 251	6.69	6.55
	Rural	14 132	6.85	6.50
	Avg		6.80	6.52
Federal-aid urban system	Urban	4 645	6.80	6.56
Secondary	Rural	8 574	6.62	6.25
	Avg		6.68	6.35
Non-federal-aid	Urban	155	6.89	6.72
	Rural	1 831	6.49	6.02
	Avg		6.52	6.07
All urban		13 207	6.85	6.68
All rural		26 454	6.80	6.44
Overall		39 661	6.82	6.53

different regions of the state can use the scales with consistency over time.

The application of such methods to visual assessment of highways can be conducted in rapid fashion, thereby speeding up quantification of rehabilitation needs. Summarization of the results can provide a quick means of assessing network-level condition of pavements at the statewide, regional, or county level.

The condition of New York State's touring route system is generally good--87 percent of road surfaces and 79 percent of road bases are in fair or better shape. However, a significant portion of roadway is in poor condition and must be given attention in the near future. Rigid pavements in New York State are generally in better condition than flexible or composite pavements. The Interstate system is in significantly better condition than other federal-aid systems. The non-federal-aid rural portion of the system is in the worst condition.

REFERENCES

1. The Status of the Nations Highways: Conditions and Performance. Office of the Secretary of Transportation, U.S. Department of Transportation, Jan. 1981.
2. Price Trends for Federal Highway Construction. Federal Highway Administration, April 1981.
3. A.T. Reno, B.J. Ritchey, and J.R. Stowers; System Design Concepts, Inc. State Transportation Finance Within the Context of Energy Constraints. NCHRP, Rept. 231, April 1981, 86 pp.
4. Collection and Use of Pavement Condition Data. NCHRP, Synthesis of Highway Practice 76, July 1981, 74 pp.
5. Highway Performance Monitoring System (HPMF): Analytical Process and Applications. Office of Highway Planning, Federal Highway Administration, 1981.
6. The AASHO Road Test. HRB, Spec. Rept. 73, 1962, 438 pp.
7. W.S. Torgerson. Theory and Methods of Scaling. Wiley, New York, 1963.
8. J.C. Nunnally. Psychometric Methods. McGraw Hill, New York, 1967.
9. Charles River Associates. Data Inventory of Attitudes Toward Transportation Systems. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, CRA Rept. No. 347, Oct. 1977.
10. M.L. Tischer. Attitude Measures: Psychometric Modeling. Paper presented at 4th International Conference on Travel Modeling, Grainau, West Germany, 1979.
11. A.O. Grigg. Rating Scales--Central Tendency and Sample Sizes. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplemental Rept. No. 647, 1981.
12. Standard Nomenclature and Definitions for Pavement Components and Deficiencies. HRB, Special Rept. 113, 1970, 29 pp.