Measures of Pavement Performance must consider the Road User

W. Ronald Hudson, Ph.D., Professor Emeritus, University of Texas
Agile Assets, Inc.
3001 Bee Caves Road
Austin, Texas 78746-5560

Ralph Haas
Department of Civil and Environmental Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Eric Perrone
AgileAssets, Inc
3200 Atlantic Avenue, Suite 205
Raleigh, NC 27604

ABSTRACT
        In 1960, Bill Carey and Paul Irick developed the Present Serviceability Index (PSI) as a user-based performance measure to define pavement quality and failure at the AASHO Road Test, a controlled load experiment that cost $300 million in 2014 dollars (*1*). The Canadians used the same approach in creating their Riding Comfort Index, but on a 0 to 10 scale. The PSI method was adopted and used worldwide to define pavement quality until the early 1990's when FHWA arbitrarily adopted International Roughness Index (IRI). It was intended as a measure of quality for HPMS (Highway Performance Monitoring System) data because IRI was touted to be standard and universal by the World Bank. PSI is still used by many agencies around the world but most state DOTs felt forced to follow FHWA and adopt IRI.
        The IRI is not standard state-to-state and more importantly the levels of "acceptability" and "failure," which must be set to define performance, vary from state-to-state. The US Federal MAP-21 requires state DOTs to do broader "performance" management and develop acceptable pavement performance measures *(2)*. PSI is tied to road user response but IRI is not. This paper examines these indexes and how they derived. It contends that PSI can serve all levels of need while IRI does not, because it is not understood by highway users and legislators. PSI reflects human rider response and IRI does not close that gap.

INTRODUCTION
        Everyone is interested in good "performance." We want our children to "perform" well in school, we want cars with great performance, and we want our pavements to perform well. MAP-21 the new Federal highway legislation *(2)* requires that "performance" management be applied to all transportation assets.   The primary concern of this conference is pavements. It is interesting to note that while not totally understood by many, we have used a performance management approach for pavements since 1961 (*1*). If you are under 55 years of age, you may never have heard of the Present Serviceability Concept developed at the AASHO Road Test *(3)*. If you have the book *Modern Pavement Management* by Haas and Hudson *(4)* you could have read about it in Chapter 7.

HISTORY
        It may be hard to believe but prior to 1960 there was <u>no measure</u> of pavement quality or failure used for pavement design, maintenance, or rehabilitation. Seems unbelievable, right? If you go back and review any pavement design method published prior to 1960, the original CBR methods, the Texas Triaxial method, the Asphalt Institute and the PCA method. Look it up. Yoder's book *(5)* covers most of these and is a good reference.
        This phenomenon is perhaps best illustrated by a story told by Bill Carey, Chief Engineer on the Maryland, WASHO and AASHO Road Tests and later Executive Director of the Transportation Research Board to Dr. Hudson in 1960. The western states (WASHO) sponsored what was then the largest controlled road or civil engineering experiment ever conducted. Many historically great pavement engineers were on the advisory committee including Francis Hveem, legendary pavement engineer of CalTRANS who invented the Hveem Profilograph, one of the very first pavement roughness devices conceived; and AC Benkelman, one the best known early flexible pavement designers with the Bureau of Public Roads, later FHWA. The Benkelman Beam, the first device used for measuring pavement deflection was named for Mr. Benkelman. At one of the committee meetings, Rex Leathers, later Chief Engineer of FHWA, told the committee that one of the test sections had perhaps failed.

According to Bill Carey the group excitedly went out to look at the test section. After looking in person, Francis Hveem declared that the section had indeed failed. A.C. Benkelman told his friend Mr. Hveem, "You are crazy." This section will last another 50,000 load applications and maybe more if we apply a thin surface seal." A lively discussion ensued and it became clear that neither the research team nor the advisory committee had a clear definition of pavement failure. As a result, the project only recorded the rate of cracking and rutting versus time and ultimately let the states set their own level of failure *(6)*.

The lack of a failure measure concerned Mr. Carey and the AASHO Road Test statistician, Dr. Paul Irick of Purdue University. They considered the problem over the next five years. When the AASHO Road Test was funded at $30 million with 836 test sections, they knew they needed a well-defined measure of pavement performance and failure. Since the purpose of a highway or pavement is to "serve the riding public" Carey called their new quality index a Present Serviceability Index (PSI). The failure level on the index was also to be defined by the riding public. Carey and Irick noted that there are five fundamental assumptions associated with evaluating pavement:

1. Highways are built for the comfort and convenience of the traveling public. Stated another way a good highway is one that is safe and smooth.
2. User's opinions as to how they are being served by highways are by and large subjective.
3. There are however, characteristics of highways that can be measured objectively and that are directly related to the users subjective evaluation of how well the highway serves them.
4. The serviceability of a given highway or pavement may be expressed by the mean rating (evaluation) given by all highway users. Honest differences of opinion preclude the use of a single opinion in establishing serviceability ratings. The mean evaluation of all users however is a good measure of highway or pavement serviceability.
5. Performance is an overall appraisal of the serviceability history of a pavement. The performance of a pavement can be described if one can observe its serviceability from a given point, usually the time in which it was built until the time at which the performance evaluation is desired. In other words, the area under the serviceability curve, FIGURE 1.

Note that the level of acceptability varies with the class of road and can be a policy or economic-based decision by the agency. The full development of the serviceability concept is reported in *(1) (4).*
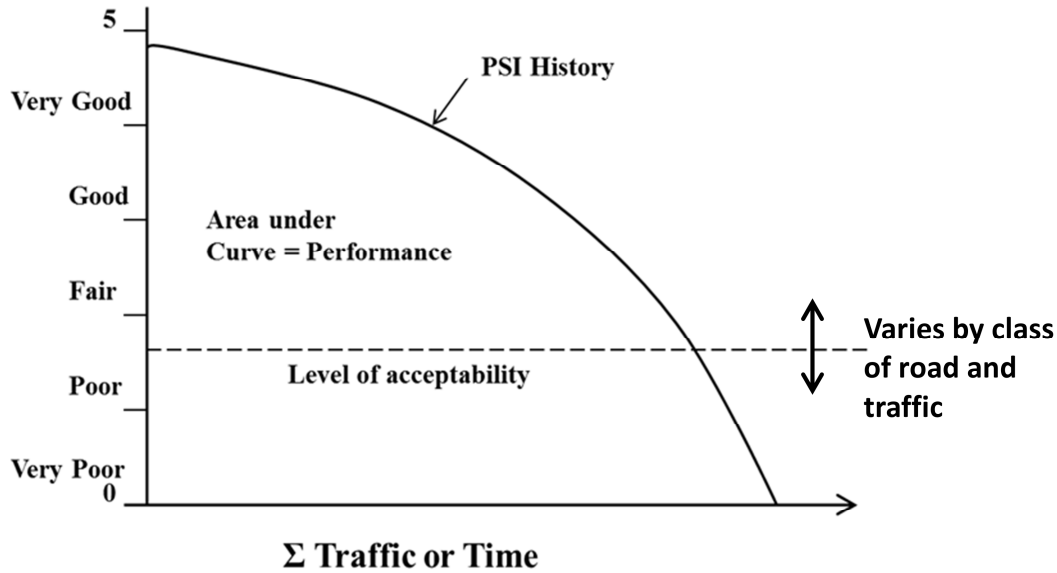
FIGURE 1 PSI – Performance Concept *(1)*

PRESENT SERVICEABILITY RATING (PSR) TO PSI

The PSI was developed at the AASHO Road Test using a rating panel of up to 50 typical highway users, men, women, young, old, pleasure drivers, truck drivers, long haul drivers, etc. All were included in the rating process. In the independent panel evaluation, all of these raters felt roughness or smoothness of the pavement was the most important characteristic to them at the 90% level. There were a few engineers included on two of the panels of raters and from their engineering "perspective" they felt that the pavement serviceability was also affected by cracking, patching and rutting. As a result Equation 1 was developed to predict present serviceability <u>index</u> (PSI) as a good estimate of the present serviceability rating (PSR). Thus PSI is based on objective measures to predict the subjective present serviceability <u>rating</u>. It must be remembered that a PSR is a subjective rating while a PSI is an objective way of estimating PSR. The following equation resulted at the Road Test.

Road Test Equation:

$$PSI = 5.41 - 1.80 \log (0.40R - 30) - 0.5 \sqrt{C + P} \qquad (\text{eq. 1})$$

Generic equation

$$(PSI = K_1 - K_2 \log (K_3 R - K_4) - K_5 \sqrt{C + P} \qquad (\text{eq. 2})$$

Where R = Roughness Measure which varies depending on equipment used, as do K_3 and K_4

C = % Cracking
P = % Patching

Where K_i are regression coefficients and K_3 and K_4 are equipment dependent.

At the Road Test roughness was measured with the Road Test profilometer and the BPR Roughometer. Equations were developed and used for both.

The PSI scale ranges from 0 to 5 with quality level keys or categories ranging from very poor, poor, fair, and good, to very good (Figure 1). While at the AASHO Road Test, roughness was measured with the Road Test Profilometer or the BRR Roughometer. It can be measured with any modern roughness device and correlated with your own user panel ratings.

This concept and the equations for both devices were used at the AASHO Road Test to evaluate all 836 test sections bi-weekly for the two year test period and to produce the final performance equations. The resulting performance equations are the basis for all of the AASHTO Pavement Design Guides from 1962 to 1993 and later. Although many pavement design engineers may not realize this because the equations are inherent in the core of the data and design manuals. In reality the cracking and patching affected the index by less than 5% and it is both scientifically and practically better (now common practice) to separate this distress into a separate distress index for purposes of maintenance and rehabilitation programming.

RIDE NUMBER

In the 1980's Janoff [7] developed a concept of ride number (RN) on a 0-5 similar to PSI as what he felt was a more scientific rating than PSI. He never had enough research funding to complete his work. In reality RN was a reasonable concept and as shown in Figure 2 [8] was 34% better correlated with mean panel rating than was IRI and the standard error was only 0.20 vs 0.43 for IRI. In these plots, RN and IRI would be two options for the variable R in equations (1) and (2) but in the data shown, RN is clearly a better measure than IRI.
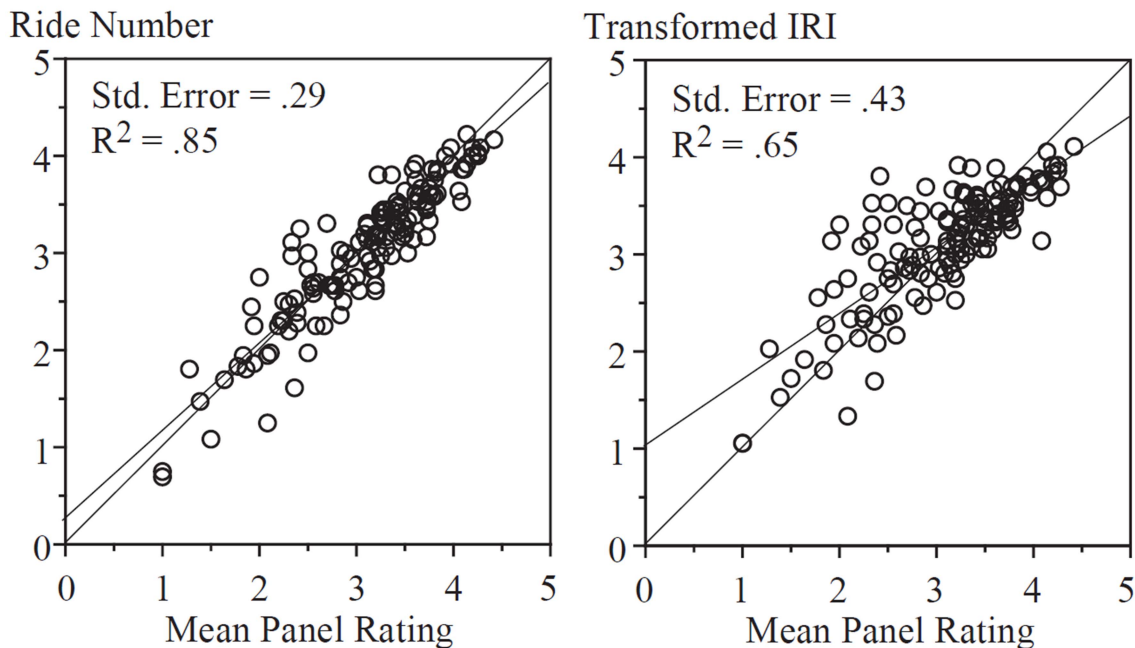

Figure 2 Correlation between Ride Number (RN) and Mean Panel Rating (MPR) [8]

PSI (PSR) was used by FHWA in the highway performance monitoring system (HPMS) data collection effort by states until the mid 80's or early 90's. Unfortunately they allowed states to "estimate" PSR if they did not have an objective PSI index. This resulted in haphazard reporting by the states, some of whom failed to measure the roughness data required.

ADVENT OF IRI USE ON US HIGHWAYS

FHWA looked for a way to solve this dilemma of poor data. They opted to change to reporting roughness as the international roughness index (IRI). The World Bank was touting the IRI which actually was <u>not</u> universally used internationally but was developed by the University of Michigan and adopted by the World Bank. It gained support because it was reportedly a "universal standard no matter how measured." This concept of a universal standard should be questioned since from the outset two scales were used, one being meters per kilometer and the other inches per mile. A conversion made by some between the two units is not always valid because the data is also captured in different units. In some cases it is filtered twice for moving average because of equipment configuration *(9) (10)*.

The Sayers reference *(8)* is an excellent treatment of roughness equipment, analysis, and of potential ways to use roughness measurements as a method of evaluating pavement performance. He provides over 50 graphs, figures, and equations of various renditions of profile and roughness measurements, equipment details, analysis summaries and the international roughness index (IRI). He does an excellent job of relating all the parts of this complex puzzle in the *Little Book of Profiling (8)* but he gives only a few pages out of over 100 to user opinions and pavement ratings. He, like many other engineers, expresses an "opinion" that user ratings are not useful for pavements. But that in itself shows the value and prevalence of "subjective opinions", in this cast <u>his</u> opinion on how things truly get evaluated in the real world.

There are errors (variations) in measurements and there are variations among road user opinions but the mean ratings of a panel of 10-25 road users was shown at the AASHO Road Test to be very stable *(1) (3)*. Such ratings reduced through a statistical equation are the best way to evaluate pavement PERFORMANCE.

FIGURE 3 shows the range of IRI which Sayers and the World Bank feel represent six different types or classes of roads found in the world. The scales show IRI for damaged pavements ranging from 250 to 675. How would highway users (the riding public) or for that matter a maintenance foreman, differentiate such a large numerical range?

<u>PROLIFERATION OF PROFILERS AND IRI VARIABILITY</u>

With the proliferation of profilometer users and producers since 1990, the accuracy and precision of profile data became a concern. The FHWA initiated a pooled fund study, "Improving the Quality of Pavement Profiler Measurements," study TPF-5(063) in 2002 *(11)*. An extensive effort has been and continues to be made to calibrate and verify profiler results. The essense of the calibration process for roughness is described in FIGURE 4 *(12)*.

Sayers is one of the developers of IRI. He does a good job describing the basics of roughness and profile measuring *(8)*. In this reference he disparages subjective ratings and strongly prefers measurements, mathmatical filtering, and calculations. His discussions of measurement and analysis are excellent as far as they go but 60-70% of road roughness/smoothness measurements are used to define how well roads serve that pesky customer, the road user, who by the way also pays for the roads. The road users opinion is clearly subjective, no doubt about it, but we must accept this challenge and Carey and Irick did that in developing the PSI/PSR concept *(1)*. Sayers makes a mistake when he says the raters used at the Road Test were "expert engineers." In fact the vast majority were private motorists ranging from housewives to lawyers to truck drivers. W.R. Hudson ws full time staff at the Road test. Only 30-40% of the time are roughness/smoothness measurements used to help maintenanc forces define how to maintain the pavement. The rest of the time they are used to define

performance. Both uses should be kept in mind. PSR/PSI should now be put back into use as the best availabe measure of pavement performance.
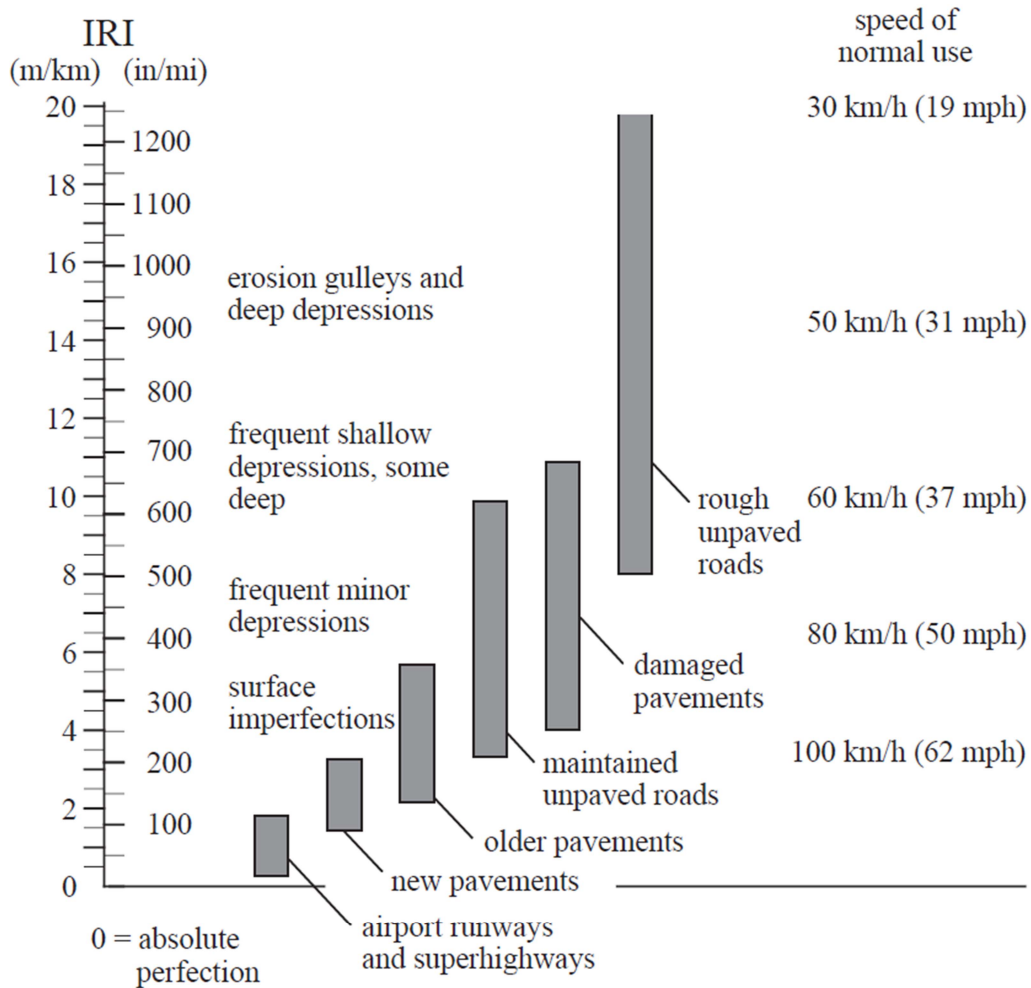

FIGURE 3 IRI range for six types and condition of road *(8)*

As shown in FIGURE 4, profilers can be evaluated based on a roughness summary statistic or based on the profiles measured by the profiler as compared to reference profiles. Three things are needed for the calibration/verification process:

- Measurement of "true" profile.
- An analytical process for comparison of the reference profile to the data from the profiler.
- Criteria for determining an acceptable difference in the profiler and reference results.
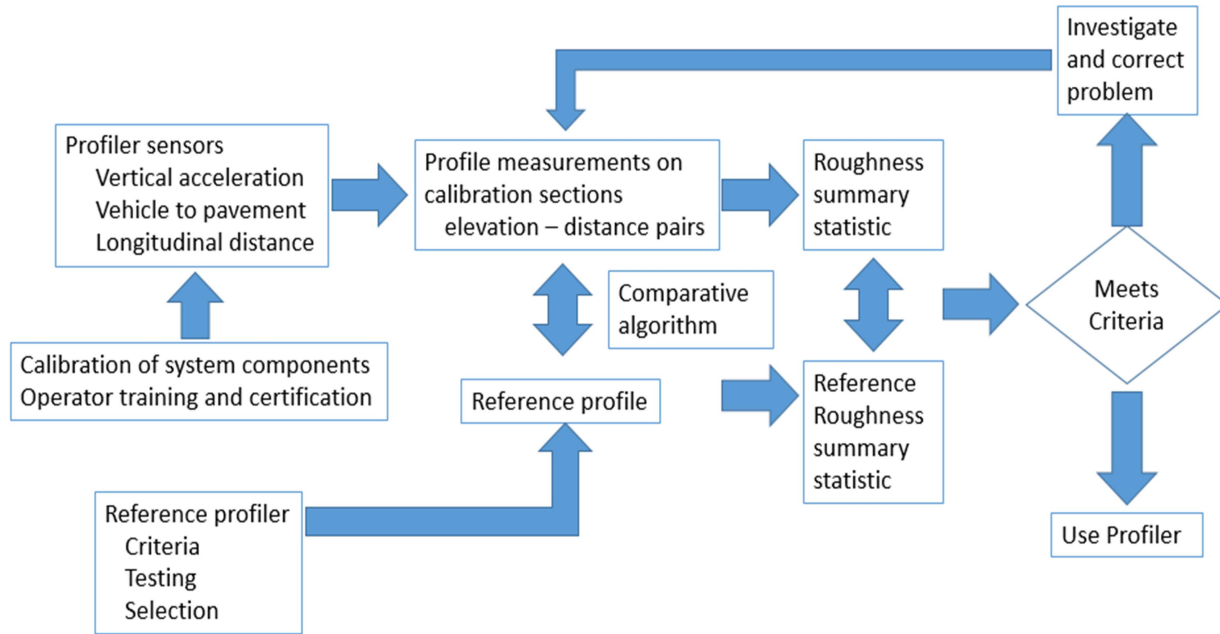
FIGURE 4   Profiler calibration and verification process (10)

A 2010 project compared different profiling equipment. TABLE 1 identifies the equipment that participated in the study *(12)(13)*, The International Roughness Index, IRI, was used as the roughness summary statistic.

The high-speed profilers were evaluated for survey and construction monitoring data collection; light-weight profilers for construction monitoring; and the walking profilers were evaluated as reference profiling devices.

FIGURE 5 shows the variability of IRI from various profilers in tests reported by Sayer *(8)* can range from 25% for the two KSLaw devices to 48% among four ICC ultrasonic devices. This is not meant to denigrate measurements but rather to point out that large variability can exist in measurement, at least as great as for panel ratings. Sayers *(8)* provides over 50 graphs, figures, and equations of various profile measurements, equipment details, and analysis. While he does an excellent job of relating all the parts of this complex puzzle *(8)*, he gives only a few pages out of 100+ to discussing road user opinions and pavement ratings. He like many other engineers expresses his "opinion" that user ratings are not useful for rating pavements. But that in itself shows the value and prevalence of "subjective opinions," in this case his, in how things truly get evaluated in the real world.

TABLE 1 Profiling devices participating in FHWA pooled fund study *(12)*

| Class	Profiler Make/model	Number in study
High-Speed	ROSAN	3
	ICC	13
	ARAN	6
	MGPS	1
	Custom	3
	RSP five0five1	2
	ROADMAS	1
	Pathway	2
	K.J. Law	2
	MHM	2
	Digilog VX	1
	Starodub/DHM	1
	SSI	1
	Ames	1
Light-Weight	Starodub/ULIP	1
	ICC	6
	SSI	4
	Dynatest/Law T64	1
	K.J. Law	2
	Custom	1
	Transtology	1
	Ames	2
Walking-Speed	SuPro 1000	3
	R/D-Meter	2
	ARRB WP	3
	Rolling Rod and Level	1
	YSI RoadPro	1
	COMACO GSI	1
	ROADMAS Z2	1
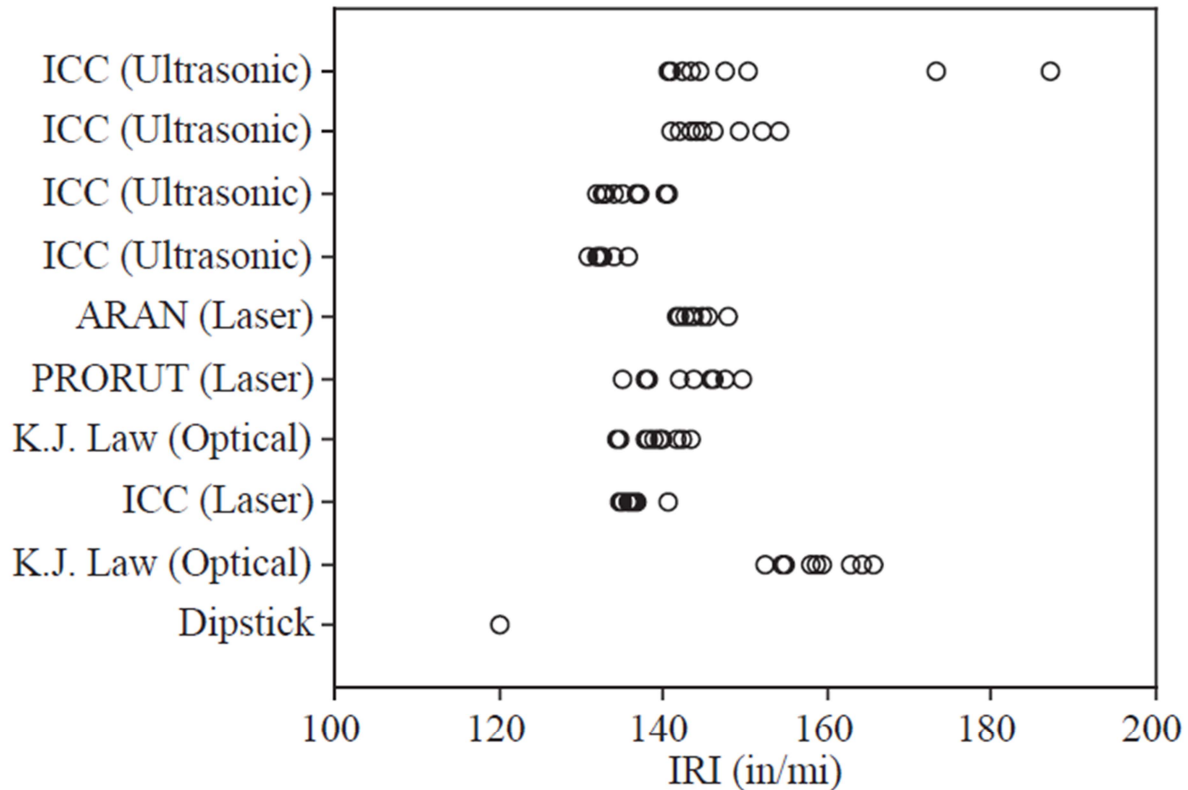
FIGURE 5   Repeatability of various profiles *(8)*

There are errors (variations) in measurements and there are variations among road user opinions but the mean ratings of a panel of 10-25 road users was shown at the AASHO Road Test to be very stable *(1)(2)*. Such ratings reduced through correlation produce an objective index (PSI) which is the best way to evaluate pavement performance.

While the IRI statistic is defined as a roughness summary statistic; technically it is a profile filter that uses a quarter-car simulation to mimic a response type road roughness measurement system. Profiler manufacturers process profile data differently and the software from the different manufacturers can produce different IRI results from the same road data. Transtec, under contract with the FHWA developed the ProVal software *(14)* in an effort to provide a common methodology for computing IRI.

Once a consistent scale is established then the user of the measure must decide what is an acceptable level of roughness, similar to the level of acceptance for PSI in FIGURE 1. TABLE 2 compares Wisconsin DOT's (WSDOT) and FHWA's different assessment of pavement quality based on IRI *(15)* . There are several reasons for these differences but clearly there are large differences and little uniformity in usage.

TABLE 2   Different interpretations of pavement quality based on IRI *(15)*

IRI Categories of Roughness	WSDOT	FHWA
Very Good	<= 95	<= 60
Good	96 - 170	61 - 95
Fair	171 - 220	96 - 120
Poor	221 - 320	121 - 170
Very Poor	> 320	> 170
	Shaded areas deemed unacceptable cells by the respective agency.	

Relation to Pavement Management

As a result of being required by FHWA to provide IRI for HPMS, the concept of using IRI as a performance measure in pavement management began to grow. This is a questionable development, but since it was administratively decreed, it had a major impact on technical developments.

In the late 60's the PMS concept of integrating planning, design, maintenance, preservation, rehabilitation, and reconstruction of pavements was introduced *(4)*. FIGURE 6 shows the basics of pavement performance in pavement management.
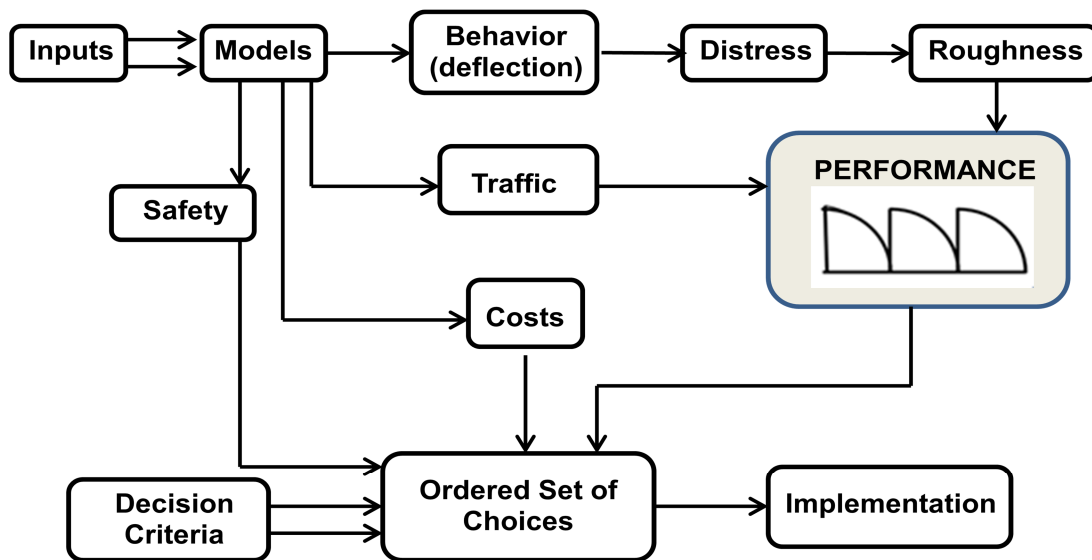

FIGURE 6   Major components of project-level PMS showing importance of "Performance" *(4)*

As shown, design models whether empirical, mechanistic, or mechanistic/empirical predict behavior, not performance. This is discussed in many references including Yoder and Witczak *(5)* and in the development of the Mechanistic-Empirical Pavement Design Guide *(16)* where it became necessary to empirically modify the mechanistic equations to make reasonable

predictions of pavement behavior. Behavior usually measured by deflection is an early indicator of distress (potential cracking, deformation and disintegration) of the pavement. It is therefore "a leading indicator of future pavement response." Behavior taken to a limit creates such distress. These distresses accumulate as a function of traffic and create roughness in an otherwise smooth road. Smoothness/Roughness is the characteristic most important to pavement users and driver reaction and was dubbed Serviceability by Carey and Irick *(1)*.

In a time-sequence, higher than expected deflection is an early indicator that distress will begin to occur and increasing distress indicates that the serviceability level, PSI/PSR may begin to decrease. This rate of decrease is an indicator of performance or future pavement failure. Some people have developed behavior indexes; distress indexes, roughness indexes, and remaining service life equations. All are useful in different ways and at different times in pavement life for predicting damage and indicating the need for pavement strengthening. Distress indexes define needed preservation, maintenance, or rehabilitation actions. Roughness most importantly expressed as PSI, the "quality" of a pavement at any given point in time and thus overtime shows the PSI history. The area under the PSI curve is a true measure of "Pavement Performance."

Understandable Nomenclature

IRI history does not provide information useful to all levels of concern, which range from the 1) road user through 2) engineers, both design and maintenance, to 3) administrators, and 4) legislators who make the funding decisions. On the other hand, all interested people can understand when you tell them that a pavement has a PSI of 3.5 in the "good category," whereas the IRI level of 87 inches per mile or 1 meter per kilometer means little to anyone except the engineers involved. It cannot in most cases be translated from one agency to another because of the variable ways in which the IRI is measured and calculated. The PSI curve is also used with a standard PSI level of acceptability to the road users for pavements on each classification of highways and speed limits. For example, it has been found that most users find a PSI below 2.5 to be unacceptable on the interstate highway system, but they will accept pavements down to a level 2.0 or 1.5 on farm-to-market or secondary roads. Furthermore, it is not possible to integrate an IRI history curve to generate a measure of performance, whereas it is easy to integrate the area under the PSI curve to produce a measure of performance (FIGURE 1). The step of translating pavement roughness to serviceability and a measure of acceptability for highway users, administrators, and legislators is absolutely essential.

SUBJECTIVE RATING

Subjective rating does require the step of having a road user's panel provide their subjective evaluations; e.g., "This feels like a good ride or a 'very good' ride." While the original present serviceability ratings at the AASHO Road Test were performed using a group of highway users ranging from five to 25 in number, it was also found that average ratings from groups of six or more provided stable estimates of the larger group of 25 and thus a good estimate of the riding public in general. To develop a PSI equation in your agency, a rating team of 6 to 12 is ample. The original teams used truck drivers, men, women, old, young, drivers, riders, and passengers of all sizes, shapes, and kinds. They were allowed to rate in their own vehicle and/or a better vehicle, truck drivers in Cadillac's, etc. In general after a very short conditioning period in any vehicle they all rated the pavements within a reasonable range of each other. The development of a PSI equation is a straightforward process but like any good process

requires some effort and attention. After the first AASHTO Interim Pavement Design Guide was produced in 1962, several states including Texas *(17)* conducted pavement rating sessions and developed a PSI equation for their state.

Many engineers disdain subjective evaluations; why is not totally clear since all human beings deal with subjective evaluations every day throughout their life. The world is loaded with subjective evolutions:

- Parents judge their children as good/bad
- School grades as A, B, C, D or pass-fail.
- Hiring: I like this person "best."
- Buying cars: I like the BMW "better" than the Mercedes, but John likes the Mercedes better than the BMW.
- Food: This steak is not as "good" or as tender as Angus beef.
- Sports Teams: The 1958 New York Yankees are the "best" baseball team of all time.
- Drafting of Players: I think Peyton Manning will be a "better" quarterback than his brother Ely Manning.
- Pavement type: "I chose Rigid over Flexible."

Yes, many of these subjective decisions are supported by evaluations, test scores, grades, productivity levels, scoring, rebounds, touchdowns, yards gained, etc. but the final decision is subjective, supported by the objective data. In the same way, objective data is used to produce an objective Present Serviceability Index instead of the subjective Present Serviceability Rating and in that case PSI is an objective predictor of the subjective evaluation, PSR, of the riding public.

Once a PSI index is generated from the PSR data, it becomes an objective measure of collected subjective opinions. The AASHO Road Test and all pavement design equations produced there-from were based on the PSI data concept including all previous AASHO Pavement Design Guides. FHWA in its early HPMS program made a regrettable decision when they allowed as default, the states to estimate or guess at the PSR of their HPMS sections.

It is unfortunate that the HPMS team did not require the states to measure the PSI rather than guess at it. But perhaps an engineering mentality" took over and they opted for a measurement which was erroneously portrayed in the "subjective opinion" of the World Bank as "the best, most stable measurement of roughness." (That pesky "opinion" crops up again.)

The use of PSI has another very practical use in pavement management. It allows pavement management engineers to look at alternative levels of pavement quality which can be obtained within their existing budgets and allows them to adjust those levels on a PSI scale and thus calculate the change in performance that will result from a change in funding for their overall highway network. This makes it easier to present a strong case to administrators and legislators for future budget changes.

DEFINING PERFORMANCE INDICATORS REQUIRED BY MAP-21

As work is done to define performance indicators for performance management in response to MAP-21, we would all be well-served to go back to our roots and use the present serviceability concept with a 0 to 5 scale, very poor to very good, which everyone, road users, designers, maintainers, administrators, legislators, and the general public can relate to. You can start the ball rolling by changing the approach in your own agency, by recognizing how difficult it is for the other stake holders in your highways to understand IRI and how difficult it is to establish meaningful service levels on an IRI scale. All you need to do is go the extra step of

panel rating and data analysis to produce your PSI. It is time to go back to a proven PSI concept as a better measure of Pavement Performance and Performance Management to fulfill MAP-21.

REFERENCES

1. Carey, W.N. and P. W. Irick, "The Pavement Serviceability-Performance Concept," Highway Research Board Bulletin 250, 1960.
2. MAP-21, "Moving Ahead for Progress in the 21st Century Act (MAP-21)," (P.L. 112-141), Federal Highway Administration, Washington, D.C., 2012.
3. Highway Research Board, "The AASHO Road Test: Report 5 – Pavement Research," HRB Special Report 61-E, 1962.
4. Haas, R., W.R. Hudson, and J.P. Zaniewski, *Modern Pavement Management*, Krieger Publishing, Florida, 1994.
5. Yoder, E.J. and M.W. Witczak, *Principles of Pavement Design*, John Wiley and Sons, 1975.
6. WASHO, "Evaluation of Pavement Systems of the WASHO Road Test by Layered System Methods," Transportation Research Board bulletin, Issue # 177, Washington, D.C., 1958, Order URL: http://www.trb.org/Publications/Pages/262.aspx
7. Janoff, M.S., "Pavement Roughness and Rideability Field Evaluation," National Cooperative Highway Research Program Report 308, 1988.
8. Sayers, Michael and Steven M. Karaamihas, "The Little Book of Profiling," The Regent of the University of Michigan, Copyright 1998.
9. Fernando, E.G., and R. S. Walker, "Impact of Changes in Profile Measurement Technology on AQ Testing of Pavement Smoothness: Technical Report," Federal Highway Administration Report No. FHWA/TX-13/9/6610-1, March 2013
10. Federal Highway Administration, "Quantification of Smoothness Index Differences Related to Long-Term Pavement Performance Equipment Type," FHWA Publication Number: FHWA-HRT-05-054, September 2005.
11. Federal Highway Administration Flyer, Update of TPF- 5(063) "Improving the Quality of Pavement Profiler Measurement," 2003.
12. Wang, H, "Road Profiler Performance Evaluation and Accuracy Criteria Analysis", Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2006.
13. Wang, H. and G.W. Flintsch, Profiler Performance Evaluation and Accuracy Criteria Analysis, TRB 86th Annual Meeting Compendium of Papers CD-ROM, 2007.
14. Transtec, ProVal software web page, http://www.roadprofile.com/, accessed 2014.
15. WSDOT, Pavement Roughness (IRI) REPORT: 2010, Wisconsin Department of Transportation, 2012.
16. NCHRP, L. M. Pierce and G. McGovern, Consultants, "Implementation of the AASHTO Mechanistic-Empirical Pavement Design Guide and Software," NCHRP Synthesis 457, Washington, D.C., 2014.
17. Roberts, Freddy L., *Pavement Serviceability Equations Using the Surface Dynamics Profilometer,* PhD Dissertation, University of Texas at Austin, Texas, May 1970.