Level of Service Concepts: Development, Philosophies, and Implications

ROGER P. ROESS

ABSTRACT

The concept of level of service and its use in highway capacity analysis are explored and discussed. Development of the concept is traced from the 1950 Highway Capacity Manual through the 1965 manual and more recent work published in Circular 212 and in the final reports of research efforts. The relative complexity of recent level of service applications and the changing interpretations, needed to understand the implications of level of service analyses, are discussed. Measures of effectiveness used in the definition and determination of level of service are also presented and discussed.

The first edition of the Highway Capacity Manual (HCM) (1) appeared in 1950 and presented a series of empirical procedures for the estimation of the traffic-carrying capabilities of a variety of traffic facilities. Although this manual did not specifically refer to levels of service, it treated capacity under a number of conditions. Practical capacity, for example, was defined as the maximum number of vehicles passing a point or segment of highway under prevailing conditions, when reasonable operating conditions are maintained.

Practical capacity was the first attempt to address the quality of traffic service provided at specified volume levels, and although reasonable operating conditions are fairly loosely defined, the concept of relating maximum volume levels to operating characteristics was firmly established in the 1950 HCM.

The concept of level of service was formally introduced in the 1965 HCM and was defined as follows:

> Level of Service is a qualitative measure of the effect of a number of factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and convenience, and operating cost (2,p.7).

The concept, as defined, relates solely to measures and characteristics that directly affect the quality of service provided to the driver. Measures included in the definition are those that are directly perceivable by the individual motorist, and are intended to describe, in relative terms, the quality of the driving experience.

It is from this straightforward definition that the level of service concept as currently used has arisen. The 1965 HCM defines six levels of service, with letter designations A-F. Level of service A designates the best quality of service and refers to virtually free flow in which the operation of an individual vehicle is not significantly affected by the presence of other vehicles. Level of service F designates the worst quality of flow and refers to conditions in which stop-and-go travel, long delays, and queued traffic exist.

Each level represents a range of operating conditions, and each is defined in terms of boundary values of appropriate parameters. Analysis procedures are intended to relate levels of service on specific types of facilities to the maximum volumes that can be accommodated at each level without causing the quality of service to fall below the defined limits for the given level of service. Such maximum volumes are referred to as service volumes.

LEVEL OF SERVICE IN THE 1965 HCM

The 1965 HCM deviated from the concept of level of service in one very important respect: Volumes, or surrogate volume measures, were often used to define level of service limits independent of the service quality parameters noted previously.

For uninterrupted flow facilities (freeways, multilane highways, two-lane highways), the 1965 HCM defined level of service in terms of two parameters, operating speed and volume-to-capacity ratio (V/C). The two were expressed as independent controls on level of service, and highway operations had to meet both the speed and V/C criteria to achieve a given level. Interestingly, the speed-V/C criteria defined were not consistent with the observed speed-volume relationships given in the manual. Particularly at levels C and D, minimum operating speed criteria given were lower than the speeds that regularly occurred for the given V/C criteria. This is discussed in detail elsewhere (3) and is important because virtually all level of service determinations made using the 1965 HCM criteria for uninterrupted flow facilities are controlled by the volume or V/C criteria.

Levels of service for other types of facilities are directly related to volume measures in the 1965 HCM. Signalized intersection level of service is based on the load factor, a parameter dependent on the demand volume. Criteria for ramp junctions are expressed only in terms of merge, diverge, and weaving volumes, and weaving area level of service is based on equivalent service volume estimates.

Only in procedures for arterials and downtown streets is level of service defined directly and solely in terms of performance measures--average overall travel speed. Despite the fact that volume is the single traffic parameter specifically excluded from the definition of level of service, it is the parameter most frequently used to define levels of service in the 1965 HCM. As a result, many public officials and practitioners have come to think of levels of service as being defined in terms of service volumes rather than in terms of the quality measures originally intended.

More recent capacity analysis procedures, however, have tended to adhere more closely to the original definition of level of service. A wider range of measures of effectiveness has been incorporated into procedures. Of greater importance, how-

ever, is the fact that the interpretation of level of service is becoming more complex and is not always the same as, or even similar to, the levels of service defined in the 1965 HCM. Subsequent sections of this paper explore some of the more recent developments in highway capacity analysis techniques and the approaches to level of service embodied therein.

RECENT HIGHWAY CAPACITY ANALYSIS DEVELOPMENTS

In January 1980 the Transportation Research Board published Circular 212, Interim Materials on Highway Capacity (4). It contained the results of two research efforts undertaken to prepare draft materials for a third edition of the Highway Capacity Manual. The first was sponsored by the National Cooperative Highway Research Program and conducted by JHK & Associates (5). It resulted in the preparation of capacity analysis procedures for signalized and unsignalized intersections, pedestrians, and transit facilities. The second was sponsored by the FHWA and conducted by the Transportation Training and Research Center of the Polytechnic Institute of New York (6). It produced capacity analysis procedures for freeways and freeway components. The publication of these materials in Circular 212 was intended to allow broad use and testing of the procedures by professionals, so that substantial feedback could be obtained before the publication of a new HCM.

Since 1980, two additional documents of major importance have been completed. A major revision to the signalized intersection procedure of Circular 212 was prepared by JHK & Associates under NCHRP sponsorship (7), and new procedures for the analysis of two-lane highways were developed by the Texas Transportation Institute of Texas A&M University System (8).

These recent documents contain several novel approaches to the level of service concept.

Uninterrupted Flow--Freeways

Levels of service for freeway segments in Circular 212 are defined in terms of density, the number of vehicles occupying a unit length of freeway lane, and average running speed. The concept presented, in which density is the primary measure of effectiveness, deviates from the 1965 HCM in three major ways:

1. Density (passenger cars per mile per lane) is used for the first time as a level of service parameter. It quantifies the proximity of other vehicles and is directly related to freedom to maneuver within the traffic stream.
2. Volume is not used as a defining criterion. Service volumes are related to density and speed criteria in accordance with calibrated speed-flow-density curves observed on modern freeways.
3. Criteria are applied to uniform 15-min rates of flow (expressed as equivalent hourly volumes) not to actual full peak-hour volumes.

These points are critical. Figure 1 shows the basic concept of freeway level of service. Levels are defined in terms of density and speed. Volumes that actually occur when these speeds and densities exist are tabulated based on observed correlations. The combinations of density, speed, and volume shown as level of service criteria are not independently selected but represent combinations actually expected to occur under ideal conditions on uninterrupted freeway segments.

Figure 2 shows the speed-flow relationships used

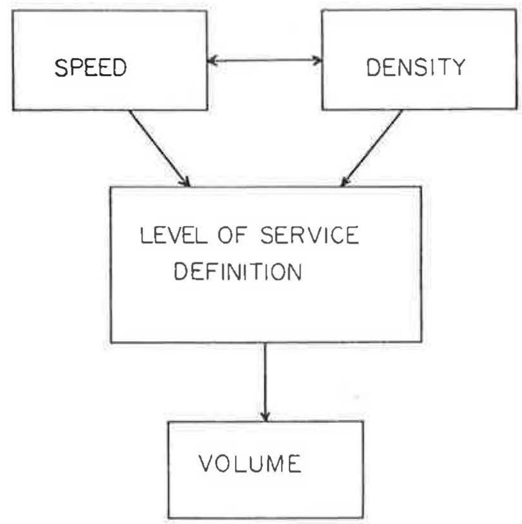

FIGURE 1 Level of service concept for uninterrupted flow.

in Circular 212, somewhat adjusted to reflect the higher average driving speeds observed under free flow conditions in recent years. Note that there is a substantial range of volumes [up to about 1,600 passenger cars per hour per lane (pcphpl)] over which speed is relatively stable followed by a range in which speed deteriorates rapidly with small increases in volume as capacity is approached. This characteristic led to the adoption of density as a primary parameter for defining level of service, because a speed definition of 50-55 mph for a given level of service, for example, would have covered the entire range of desirable volume levels.

The shape of these curves also influenced the selection of criteria for level of service boundaries for various average highway speeds (AHS) given in Table 1. Note that as the levels go from good (level A) to poor (level F) the range of densities encompassed by each level increases, as does the range of speeds. The range of volumes encompassed, however, gets progressively smaller as the levels get poorer. Thus, the range of service volumes in level E is only 75 pcphpl, and the range of density covered is 20 passenger cars per mile per lane (pc/mi/ln). This is in accordance with the shape of the observed speed-flow curves (i.e., as capacity is approached, a small change in volume will bring about a radical deterioration in service quality, as measured by speed and density). Critical comment on the narrow range of service volume in level E may eventually lead to some revision of these criteria, but it should be noted that level of service is defined by the performance parameters of density and speed. It is therefore important to maintain reasonable ranges for these values, despite the narrow service volume ranges that result.

The application of criteria to 15-min rates of flow is another important change from the 1965 HCM. The 1965 HCM more or less applied an average level of service over the full peak hour of operation, with the exception of levels C and D for freeways, which explicitly consider the peak hour factor (PHF) and the peak 5-min flow period. The Circular 212 method recommends the analysis of uniform periods of flow. Thus, if operations were at level of service C for 0.5 hr and at level of service E for the rest of the hour, they would be separately analyzed and labeled. The 1965 HCM would essentially label the full hour as level of service E, even though the condition exists for only 0.5 hr. The emphasis on

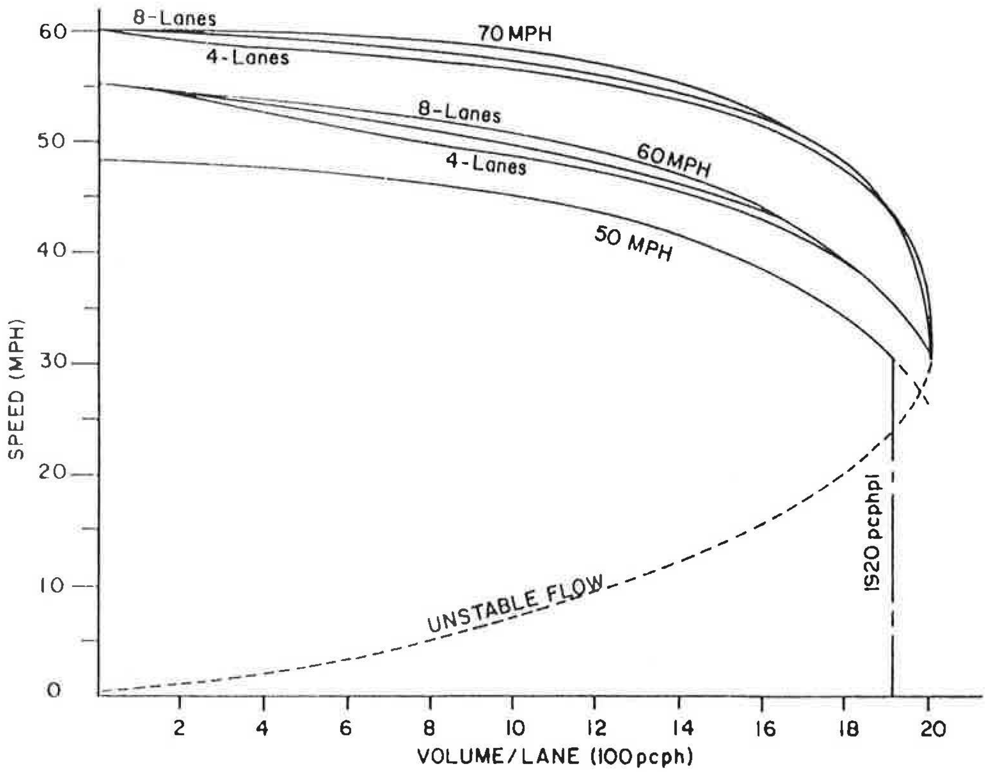

FIGURE 2 Speed-flow curves for freeway segments.

TABLE 1 Level of Service for Basic Freeway Segments (4)

LEVEL OF SERVICE	PERFORMANCE CRITERIA FOR LEVELS OF SERVICE: SPEED MPH (Km/h)	DENSITY PC/mi/LN (PC/km/LN)	MAXIMUM SERVICE VOLUMES (ONE DIRECTION) FOR LEVELS OF SERVICE DURING UNIFORM PERIODS OF FLOW (PCPH): 4-Lane (2 ea.dir)	6-Lane (3 ea.dir)	8-Lane (4 ea.dir)	EA. ADD LANE
AHS = 70 MPH (112 km/h.)						
A	≥50(80)	≤15(9.4)	1600	2400	3280	820
B	≥50(80)	≤25(15.6)	2500	3900	5400	1350
C	≥48(77)	≤35(21.9)	3400	5100	6800	1700
D	≥40(64)	≤47(29.4)	3850	5775	7700	1925
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-
AHS = 60 MPH (96 km/h.)						
A	*	*	*	*	*	*
B	≥45(72)	≤25(15.6)	2300	3525	4800	1200
C	≥43(69)	≤35(21.9)	3050	4575	6100	1525
D	≥38(61)	≤47(29.4)	3600	5400	7200	1800
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-
AHS = 50 MPH (80 km/h.)						
A	*	*	*	*	*	*
B	*	*	*	*	*	*
C	≥40(64)	≤35(21.9)	2800	4200	5600	1400
D	≥35(56)	≤47(29.4)	3300	4950	6600	1650
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-

* Level of Service not achievable due to reduced safety on highways with restricted AHS

peak 15-min rates of flow carries through to other procedures in the Circular and to the more recent procedures for two-lane highways and signalized intersections as well.

Freeway Components

Level of service criteria for weaving sections and ramp junctions are generally keyed to those for uninterrupted flow freeway segments in Circular 212.

The circular introduces a totally new approach to weaving area analysis in which levels of service for weaving and nonweaving traffic are separately rated. Because speed has been found to be sensitive to volume throughout the full range of volumes in weaving sections, speed is used as a direct measure of effectiveness for nonweaving vehicles. The level of service for weaving vehicles is based on the differential between the average speed of nonweaving vehicles and the average speed of weaving vehicles. It is assumed that weaving vehicles would tolerate somewhat slower speeds than nonweaving vehicles for any given level of service because of the lane-changing required.

Merge and diverge criteria at ramp junctions are keyed solely to volumes. Volume levels, however, have been set to allow freeway operations in the vicinity of the ramp to continue as defined by general freeway criteria. In general, the merge and diverge volume limits of the circular are lower than the corresponding criteria in the 1965 HCM. As do the general freeway criteria, weaving area and ramp junction criteria refer to peak 15-min flow rates and performance criteria.

Two-Lane Highways

The recently completed two-lane highway methodology prepared by Texas Transportation Institute for NCHRP (8) introduces a new measure to the level of service arena. As was the case for freeways, speed has been found to be relatively constant over a broad range of volumes. Although the methodology retains speed as a principal level of service parameter, it also introduces a secondary measure: percent of time delayed. This parameter is a measure of the proportion of time drivers must spend in a queue because of the inability to pass a leading vehicle or vehicles. Although this parameter is difficult to measure directly, the method assumes that headways of 5 sec or less indicate an involuntarily queued vehicle.

For two-lane highways, this is a most interesting parameter and is indicative of one of the most vexing driver frustrations encountered on such roads. Although the parameter is used more in a qualitative than a quantitative sense in the procedure, its introduction is an indication of developing philosophies of level of service.

Signalized Intersections

Circular 212 formalizes a U.S. analysis procedure based on critical movement analysis. The critical movement approach is hardly new. Greenshield's original work on signal timing is based on critical movement analysis. Work on applying critical movement techniques to capacity analysis has been extensive both in the United States (9,10) and abroad (11). Circular 212 relates level of service directly to the saturation ratio, which is the V/C ratio.

The updated procedure prepared by JHK & Associates for NCHRP (8), however, bases level of service on delay, specifically the average individual stopped delay. The V/C ratio does not enter into the criteria for level of service. This is consistent with the concept of level of service (LOS) but nevertheless introduces an entirely new interpretation of level of service labels. The following table gives the criteria presented in the JHK procedure.

Level of Service	Average Individual Stopped Delay (sec/veh)
A	< 10.0
B	10.1-20.0
C	20.1-30.0
D	30.1-40.0
E	40.1-60.0
F	> 60.0

Research conducted during development of the procedure has shown that delay and V/C ratio are not strongly correlated on a one-to-one basis. Signal progression, for example, has a major influence on delay. The JHK procedure essentially predicts average individual stopped delay assuming random vehicle arrivals. A correction factor is provided to account for progression and the type of signal controller. The factor, however, ranges from 0.5 to 1.5. Thus an intersection with a random arrival delay of, for example, 42 sec/veh could be adjusted to a low of 21 sec (LOS C) or to a high of 63 sec/veh (LOS F). The V/C ratio would remain unchanged, because it is not affected by progression.

Two interesting cases can arise when the level of service criteria in the JHK procedure are used:

1. A delay of > 60 sec (LOS F) may result when the V/C ratio is less than 1.00 and operations are entirely stable and
2. A delay of < 60 sec (LOS A-E) may result when the V/C ratio is marginally higher than 1.00 (i.e., marginal oversaturation exists).

Thus, level of service F does not necessarily connote "forced" or "breakdown" flow conditions, but it is an indication of excessive or unacceptable levels of delay. Further, V/C ratios marginally greater than 1.00, a condition that identifies breakdown flow, will not necessarily be rated as level of service F.

This concept of level of service is considerably different from that used throughout the 1965 HCM. The lower boundary (maximum service volume) for level of service E, for example, is no longer synonymous with capacity. In fact, its relationship to capacity is somewhat variable. In essence, the JHK procedure separates the consideration of level of service from capacity. Both the LOS and the V/C ratio are critical parameters in the overall analysis of an intersection, and both must be considered in evaluating present or future operations.

This will require a change in the way LOS terminology has been commonly used and understood in the past. Delay is a more relative measure than V/C ratio; what is unacceptable in a small urban area may be quite acceptable in a major metropolis. In midtown Manhattan, for example, delays of 60 sec/veh may be quite tolerable as long as oversaturation is avoided (i.e., V/C < 1.00). Traffic engineers and planners will have to defend cases in which a poor LOS is deemed acceptable and explain cases in which level of service F exists but in which excess capacity still remains. This may be a difficult task when dealing with groups accustomed to the more absolute LOS designations currently in use.

Unsignalized Intersections

The unsignalized intersection methodology presented in Circular 212 is a translation of a German procedure developed in 1974. In one of the few cases where a direct indexing of level of service criteria to volume still exists, the procedure uses the parameter "unused capacity" (i.e., the capacity of an approach or lane minus the volume on the approach or lane). Although not consistent with the general philosophy of LOS as defined in the 1965 HCM, unused capacity is presumed to generally correlate with approximate delay ranges that are verbally defined. Because this correlation is not numerically expressed or statistically verified, delay is clearly not a principal parameter in this methodology, nor is it used in defining the level of service boundaries.

Transit Facilities

Circular 212 presents a comprehensive overview of transit capacity and level of service. This is basically a new area for formal highway capacity analysis and introduces some interesting concepts. Transit level of service is a two-dimensional issue, dealing with the internal environment of the transit vehicle and the traffic environment of the vehicle itself. A bus, for example, can be traveling at free flow speed on an uncongested highway but be loaded with many standees, thereby providing a low level of service. Conversely, a lightly loaded bus can provide an excellent internal environment but be stuck in a traffic stream operating at level of service F. The transit material of Circular 212 emphasizes the internal environment of the transit vehicle as the principal LOS measure, using load factor as the determining parameter. Load factor is defined as the total number of passengers on the vehicle divided by the seating capacity of the vehicle.

The transit methodology of Circular 212 also introduces a new concept of capacity. Capacity is defined as the maximum loading of the transit vehicle including a "reasonable" number of standees. Crush load capacity is a larger number of people who can be accommodated under "crush" conditions--loading that is severely uncomfortable. Capacity is also defined in terms of persons per hour passing a point or segment of facility, rather than the more traditional vehicles per hour. The concept of person capacity as opposed to vehicle capacity is critical in the analysis and justification of various bus transit priority schemes.

Pedestrians

The treatment of basic flow characteristics of pedestrians in Circular 212 is quite similar to that of uninterrupted flow vehicular characteristics. Whereas for uninterrupted flow, density was used as the principal defining parameter for LOS, the pedestrian methodologies use the parameter "space" (square feet per pedestrian), which is the inverse of density. These criteria assume that pedestrians are most severely affected by their proximity to others, which has been shown to have a drastic impact on freedom to maneuver and on walking speed. Pedestrian capacity is approached in terms of person capacity.

Other Facilities

Materials on capacity analysis of other types of facilities are being developed. Indeed, revisions to the materials discussed previously are also under way. LOS criteria for suburban and rural multilane highways will be similar to those for freeways and will rely on density as a primary measure of effectiveness. Arterial LOS criteria are currently projected to be based on average overall travel speed. Different criteria for different classes or categories of arterials may be defined.

SOME OVERALL ISSUES

Relationship to Other Standards

Whether fortunate or unfortunate, a large number of formal federal, state, and local standards for such divergent areas as highway design, highway operations, noise, and air quality are written in terms of specific level of service criteria. Thus, a state may require that rural highways be designed to level of service B or that highways be redesigned and reconstructed when they reach level of service E. A planning board may require that air quality analysis be done for operations at a specific level of service.

Most of these standards have legal significance. As new procedures for capacity analysis are introduced and used, it is vital to remember that level of service definitions and criteria have changed, sometimes radically. Level of service C for a signalized intersection, for example, may bear little resemblance to level of service C as defined in the 1965 HCM. Because most LOS references in existing standards are to the 1965 HCM, it is critical that analysts and users make the proper adjustments in the application of new procedures. In time, as new procedures are included in a new Highway Capacity Manual, it is to be expected that these standards will be revised. Such revisions, however, may be slow in coming, and administrators, policymakers, legislators, and others involved will have to be carefully re-educated in the interpretation of LOS criteria and their meaning.

The level of service concept, as defined in the 1965 HCM, has become such a useful tool in describing traffic operations, and so generally accepted as an absolute quality scale, that changes in its framework will require a concentrated re-education effort.

Re-Education

When the 1965 HCM was published, it was followed by a multitude of professional short courses, offered by numerous consultants and universities, aimed at training practitioners in its use. Much the same can be expected with a third edition HCM. Such courses have already been offered on Circular 212. Given the relative complexity of newer techniques compared with the 1965 HCM, and given the rather substantial revisions in the use of and criteria for level of service, the re-education effort will be most important to a smooth transition. There are also more users to consider: Since the 1965 HCM, more and more professionals have found critical uses for the manual, including as input to environmental analyses.

Courses will have to focus not only on practitioners but on the policymakers and administrators who must act on the basis of capacity analyses and related information. It is at this level that the revisions in LOS criteria must be most clearly transmitted to avoid the misuse of new analysis output based on old criteria.

Level of Service F

The 1965 HCM uses level of service F to describe conditions in a queue of vehicles that forms behind a breakdown point or constriction. Such breakdowns or constrictions exist at points where the arrival volume exceeds the departure volume, leading to queue formation.

Because analysts are generally more interested in the point of the breakdown and its cause than in the operating conditions in the queue, many analysts have used the LOS F designation to describe the point of the breakdown. This is technically incorrect, because operations at the point of the breakdown are virtually always at or near capacity. Operations within the queue are what is properly described as LOS F. Nevertheless, LOS F designation of the breakdown point is useful to highlight such critical points.

Circular 212 and other new procedures do not clarify this point to a substantial degree, although sample problems indicate that the use of LOS F to designate points where demand exceeds capacity is accepted.

SUMMARY AND CONCLUSION

Level of service criteria and measures used in Circular 212 and other new capacity procedures indicate that the concept of LOS defined in the 1965 HCM is being more closely adhered to than is the 1965 HCM itself. Table 2, which gives criteria for LOS used in the 1965 HCM and in more recent procedures, illustrates this point.

Volume and volume-based measures were not intended to be among the defining criteria for LOS. Measures describing the basic service quality perceived by the user--travel time, freedom to maneuver, comfort and convenience, safety, and so forth--are the terms used in the definition. Volumes were

TABLE 2 Level of Service Criteria Compared: 1965 HCM and Recent Techniques

	Measures of Effectiveness Used to Define LOS	
Type of Facility	1965 HCM	Recent Methods
Freeways	Operating speed	Density
	V/C ratio	Avg running speed
Multilane highways	Operating speed	Density
	V/C ratio	Avg running speed
Two-lane highways	Operating speed	Avg running speed
	V/C ratio	Percent time delayed
Ramps	Volumes	Volumes
Weaving areas	Volumes	Avg running speed
	Operating speed	
Signalized intersections	Load factor	Delay
Unsignalized intersections		Unused capacity
Arterials	Operating speed	Avg overall travel speed
Transit		Load factor
Pedestrians		Space

to be specified (based on empirical observation) for various types of facilities and levels of service. This procedure has been generally followed in the development of new capacity analysis methodologies. However, the use and interpretation of LOS have been substantially altered, causing a problem that will have to be addressed by re-education of users and policymakers alike.

REFERENCES

1. Highway Capacity Manual. Bureau of Public Roads, Washington, D.C., 1950.
2. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
3. R.P. Roess, W.B. McShane, and L.J. Pignataro. Freeway Level of Service: A Revised Approach. In Transportation Research Record 699, TRB, National Research Council, Washington, D.C., 1979, pp. 7-16.
4. Interim Materials on Highway Capacity. Transportation Research Circular 212. TRB, National Research Council, Washington, D.C., 1980, 276 pp.
5. JHK & Associates. Development of an Improved Highway Capacity Manual. Final Report on Phase I, NCHRP Project 3-28. TRB, National Research Council, Washington, D.C., 1979.
6. R.P. Roess, W.R. McShane, E.M. Linzer, and L.J. Pignataro. Freeway Capacity Analysis Procedures. DOT-FH-11-9336. Polytechnic Institute of New York, Brooklyn, N.Y.; FHWA, U.S. Department of Transportation, 1978.
7. JHK & Associates. Urban Signalized Intersection Capacity. Procedural Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., Feb. 1983.
8. C.J. Messer. Two-Way, Two-Lane Rural Highway Capacity. Procedural Report, NCHRP Project 3-28(A). Texas Transportation Institute, Texas A&M University, College Station; TRB, National Research Council, Washington, D.C., Feb. 1983.
9. D.S. Berry and P.K. Gandi. Headway Approach to Intersection Capacity. In Highway Research Record 453, HRB, National Research Council, Washington, D.C., 1973, pp. 56-60.
10. C.J. Messer and D.B. Fambro. Critical Lane Analysis for Intersection Design. In Transportation Research Record 644, TRB, National Research Council, Washington, D.C., 1977, pp. 26-37.
11. X. Miller. Capacity of Signalized Intersections in Australia. Bulletin 271. Australian Road Research Board, Numawading, Victoria, March 1968.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.