Human Factors Evaluation Methods for Traffic Control Devices

Fred Hanscom
Transportation Research Corp.
Fhancsom@erols.com

Robert Dewar
Western Ergonomics
westerg@shaw.ca

INTRODUCTION

The U S Manual on Uniform Traffic Control Devices (MUTCD) outlines the steps in a process for requesting and conducting experimentations for new traffic control devices (TCDs). This process should contain a detailed research plan to include, if appropriate, both before and after studies as well as quantitative data describing the performance of the experimental device. However, no instruction or guidance is provided in the MUTCD with regard to specific evaluation techniques. Due to the critical role of human factors in TCD design and effectiveness, it is essential that evaluation techniques be properly conducted and be sensitive to road user capabilities and limitations.

Many currently used TCDs were developed several decades ago with little or no scientific data to support their design or effectiveness. Shapiro, et al. (1) in 1987 indicate that previously many TCD standards were based on subjective opinion, often that of those on relevant committees responsible for the design of devices such as symbol signs. Their research study identified standards that lacked a research basis or were in conflict with research studies. This was the experience of one of the authors who was a voting member of the U S National Committee on Uniform Traffic Control Devices in the 1970s. Additional research was recommended by Shapiro, et al. for 17 TCD standards.

In recent decades much more effort has been spent on the scientific evaluation of TCDs, especially those new ones being introduced into a system. There are a number of criteria for an effective TCD. These include comprehension, legibility (both distance and glance legibility), conspicuity, response time, and learnability (how well the meaning of a TCD can be remembered). Their relative weightings have not been determined, but some are more important than others. A survey by Dewar (2) asked traffic engineers and sign experts in The US, Canada, Australia and New Zealand about the importance of several criteria for the design and evaluation of traffic sign symbols. The most important were comprehension, legibility and conspicuity. It is appropriate to include at least the most important ones (comprehension and legibility distance) when evaluating these devices.

If a new symbol sign message is required the first step is to determine if a current one is in use elsewhere (another country or state), and if so whether it has been shown to be effective. If an effective one is found, there may be no need to design anther version.

METHODOLOGY

Methods can be broadly categorized as either field (on-the-road or on a test track) or laboratory measures, including driving simulators, surveys and focus groups. Laboratory studies are usually done with artificial stimuli in a test laboratory or classroom, with one subject or small groups, depending on the nature of the test procedure. Computers may be used to present stimuli (e.g., signs, signal configurations). Some laboratories use driving simulators to mimic more effectively the real world of driving. Simulators can be expensive and fail to replicate the road world of driving unless they are motion-based.

Planning and evaluation of TCDs should include the following steps.

1. Problem Identification - observation of road user movements, vehicular studies, crash records, etc.
2. Development of a Research Question - a hypothesis that can be tested in the study (e.g., will a new device elicit the specified behavior; is design A better than design B?).
3. Design the Evaluation - identification of a sequence of before-after studies, conducted over time or across different locations, a laboratory study or survey to test for comprehension and legibility. In the case of the former approach consideration should be given to a "control" study, at a similar location or locations where no treatment has been applied.
4. Finalize the Evaluation Plan – e.g., consider methodological trade-offs as constrained by time and budget
5. The conduct of the evaluation must be based on the following:

- Data Collection Plan - observation of appropriate events to ensure valid measurement of effectiveness
- Determination of Sample Size and composition - a sufficient number of observations is required to ensure statistical suitability of results and an adequate sample of road users - novices, seniors, both genders.
- Determine Data Collection Periods - periods of data collection (time, day of week, weather) must be consistent between before and after studies for field measures.
- Statistical Analyses - testing for statistical significance to determine the likelihood that any observed change was caused by the treatment
- Assess Practical Difference - a calculated "statistically significant" result may be too small to represent a practical effect
- Evaluate the Results – e.g., the cost-benefit analysis is useful to determine whether a TCD is cost-effective in terms of collision reduction or changes in road user behavior.

A variety of methods are available for the evaluation of TCDs (3,4). Dewar and Ells compared three methods of sign evaluation – driving toward the signs on a rural road at 30 or 50 mph, a modification of this procedure where subjects drove at 17 mph toward signs 1/3 the standard size, and a laboratory measure of reaction time to the same signs. The measure was the distance, or time, for drivers to classify and to identify the sign messages. Results showed the three methods were closely related.

The use of computers and driving simulators allows the introduction of additional variables such as loading tasks (e.g., count backward from 100 in threes, respond to a target shown at random on the dashboard), vehicle handling characteristics, and environmental conditions (e.g., darkness, rain). However, simulators can induce motion sickness, especially for older drivers. Laboratory measures include drivers viewing images of TCDs to establish comprehension or legibility, as well as "paper-and-pencil" tests to measure comprehension, preferences, etc.

<u>Comprehension</u>

A number of methods are available for determining how well users understand TCDs. The procedures include writing the meaning of a TCD on an answer sheet, selecting the most appropriate answer in a multiple choice (MC) format, rating the clarity of the device's meaning, or indicating the action to be taken in response to the TCD (e.g., which way a driver may turn when seeing a configuration of signals or a pavement marking). Having subjects write out the meanings of TCDs is the most time-consuming laboratory technique, but it is the preferred one, as it provides the richest data, allowing, for example, evaluation of the nature of the errors and confusions among symbols within the same signing system. The MC test can also be used (e.g., which of 4 signs means added lane? which of 4 answers is correct for this symbol? or which of 4 actions is allowed when this marking is seen?). It is essential with MC tests that reasonable “wrong” answers are used in order to avoid correct guessing. One mistake often made by those using this method is failure to correct for guessing. If there are four choices in the MC test, then a subject can get ¼ correct just be guessing, as chances of being correct and one in four. So a correction needs to be made. The formula for this is:

FS = R - W/(N - 1)
FS= "corrected" score
R= number of items answered right
W= number of items answered wrong
N = number of options (alternatives)

For example: assume a 40-item sign test where the subject gets 28 correct and 12 wrong. The corrected score would
be 28 – 12/3 = 24 correct

Another method to gauge comprehension is the confusion matrix. This involves recording the number of responses that were correct and those wrong responses that were given another meaning. For example, an added lane symbol may be thought to be added lane, merge, yield, one way, or stop. The table below, with imaginary data, provides an example of how this method might be used. It can be seen that the added lane sign was only confused with a merge sign. The frequency of these confusions can provide insight into why errors are made and can lead to sign redesign if needed.

		Response			
	Added lane	merge	yield	one way	stop
Sign					
Added lane.	78	15	7	0	0
merge	2	90	6	1	1
yield	1	2	94	1	2
one way	0	3	0	97	0
Stop	0	0	2	0	98

In a comprehension study of all the symbols in the 1988 US MUTCD (5) the most common wrong answer was MERGE.

Comprehension of roundabout traffic control devices included in the 2009 MUTCD, (e.g., Figure 2D-8 destination signs and Section 3C.01 pavement markings) was tested in a laboratory study (6) to ensure their understanding by road users. The tested devices included guide signs, regulatory/warning signs, central island treatments, and pavement markings

Participants were told to follow a route to a specific destination and were shown sequential pictures of a roundabout approach and entry scenes. They made choices (and indicated their confidence in each choice) as to the correct-inner circle lane for their intended destination. Figure 1 below depicts an example scene in which drivers would designate their choice of lanes (A or B). The applied TCD measure of effectiveness was the device configuration that produced highest proportion of correct lane choices.

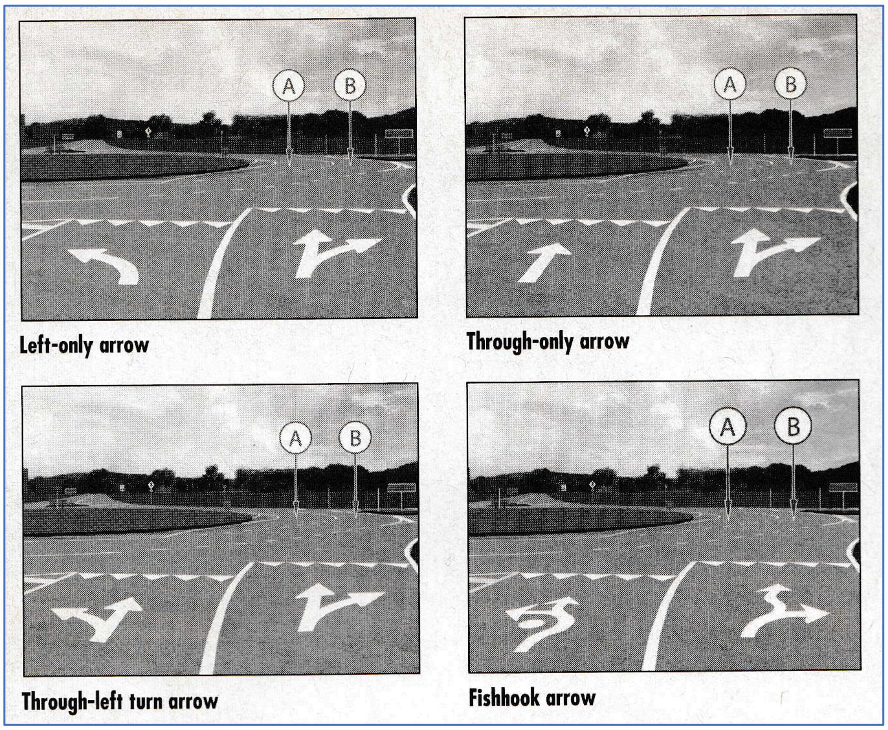


Figure 1. Example approach scene stimuli: participants asked to choose lane A or B

In addition to determining the proportion of correct line choices associated with areas markings, a structured interview procedure was also applied. Participant responses to the question, "What does this device tell you?" were insightful regarding driver interpretations of tested devices.

The value of appraising traffic sign symbols by experts and applying three ergonomic design principles to traffic sign comprehension has been demonstrated by Ben-Bassat, et al. (7,8) who had 27 human factors and ergonomics experts from 10 countries evaluate 31 conventional signs and 1-3 alternative symbol designs for each of these signs for their compliance of the signs with the three ergonomic guidelines of message-symbol compatibility, standardization, and familiarity. The experts assigned higher ratings to the alternative designs for 19 of the 31 signs.

Legibility

Legibility measures to determine the distance at which a sign message is legible for signs may have the driver drive or be moved toward the sign (3) or the sign moved toward the driver (9). One can also gradually increasing the size of the image on a computer screen until the subject can describe its contents or identify the message (5). Legibility of traffic signs was studied by Khavanin and Schwab (11) who had press a button when they could read signs while driving toward them.

Glance legibility can be gauged by presenting the TCD for a very brief duration to find the proportion of drivers able to identify the message at a specified brief interval, or by increasing the duration of the presentation until the driver can identify the message (5).

Perception-response time

The speed with which drivers can interpret a TCD, especially signs, can be measured by displaying a photo, computer image or slide of the TCD and recording how quickly the driver identifies it by naming the message (being recorded by a voice key) or pressing a button to indicate - yes or no - whether the sign shown did or did not correspond with a name of the sign given in advance. Using this method Ells and Dewar (10) found that symbol signs were recognized faster than word signs.

Conspicuity

Conspicuity, how noticeable a device is in the road environment, can be determined for example by noting how many drivers notice the sign on the road (12) or timing how long it takes to detect it when displayed among a number of other signs, or a visually cluttered roadway scene, presented in a photo, on a computer screen (5).

Learnability

To determine how readily a TCD can be remembered subjects can be told the meaning of the device (e.g., a sign or pavement marking), then tested weeks or months later to see if they remember the meaning.

On-road measures

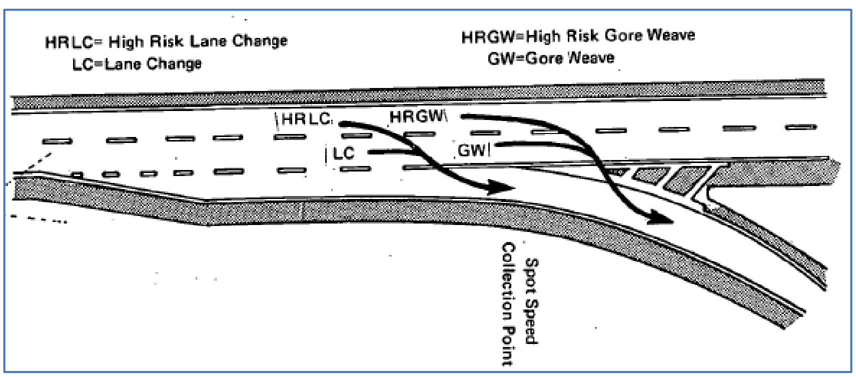
Performance of a TCD on the road can be assessed by observing driver, cyclist and pedestrian behavior. The TCD at a control site or sites with similar geometry and traffic volume is needed for comparison with the device of interest, or a before-and-after study can be done to determine the effectiveness of a new or changed TCD. On-road measures may involve distance needed to detect and identify target signs or signals or measuring time taken to changing lanes in response to a pavement marking. Instrumented vehicles are sometimes used to record speed, steering, braking, placement on the road, etc. Driver behavior (e.g., illegal turns, running red lights) or pedestrian behavior (e.g., late entry to cross an intersection, jaywalking) can be observed at specific sites using video recording or personal observations. The advantage of video is that it can be viewed a number of times, if necessary, and scored by two recorders. Any discrepancy between these two can be resolved by having a third person score the data.

With on-road measures care must be taken to observe behavior not only at the location of interest, but at control locations with similar road configuration, traffic volume, etc. in order to know whether any new or changed TCD changed road user behavior. In addition, factors such as time of day, day of the week and season need to be controlled.

Eye-movement measures can also be used on the road or in the lab to find out how effectively or quickly drivers detect a TCD as they search the road environment or a display shown on a screen. This can provide information on how sign design or placement attracts a driver's attention and what part of the sign is viewed first. The latter may have implications on where to place specific information on the sign panel.

Field observational studies of road user behaviors provide detailed and unique insights regarding the effectiveness of TCDs. Both the measures (observed data variables) and the methods (how relevant variables are observed) are critical to the conduct of any field study. The most commonly applied measures, vehicle speeds and speed-profiles, are gathered using manual-stopwatch timing, video burst sampling, or pavement instrumentation. Vehicle lateral roadway position and vehicle erratic maneuvers are obtainable via video. The following are examples of TCD evaluation methods apply to operational measures to determine the effectiveness of designated TCDs:

Lane-change behavior approaching interchange with diagrammatic guide signs. The objective is to establish the validity of selected Measures of Effectiveness (MOEs) for guide signs that comply with the MUTCD. These MOEs comprise an implementable tool by which traffic engineers can judge the effectiveness of new guide signs. For example, in one study (13) the specific lane-change behaviors illustrated in the figure below were videotaped, and drivers making these maneuvers were stopped by State Police and requested to complete a 4-page questionnaire.



Driver behaviors indicative of potential guide signing issues

The level of hazard associated with each of the above noted lane change behaviors was identified utilizing over 300 driver surveys, each associated with its respective lane-change behavior. The questionnaire analysis addressed the following driver human-factors issues:

- Driver information processing - including survey items that confirmed the related nature of the vehicle action (e.g., sign detection, interpretation, information retention).

- Driver predisposition – including sources of variance (such as biographical, risk-taking, driver experience, and comfort factors) capable of influencing a driver's response to guide signs.

The evaluation tool is applied by practitioners to estimate hazardous operational conditions resulting from potential guide sign deficiencies as follows.

Gore Weave (and High-Risk Gore Weave) Guide sign issues may be associated with driver information processing:

- Greater sign information processing difficulty with all guide signs on interchange approach.
- Less certain of action response to all guide signs on approach.
- Less time available to read and respond to intermediate exit direction sign.
- Lower preference rating for intermediate exit direction sign.
- Less likely to detect at least one guide sign.

Driving Slowly Guide sign issues may be associated with driver information processing as follows:

- Greater information processing difficulty with at least one guide sign.
- Lower preference rating of gore-located exit direction sign.

Late Lane Change Guide sign issues may be associated with driver information processing as follows:

- Greater information processing difficulty with at least two guide signs.
- Less certain of action taken to gore-located exit directions signs and one advance sign

Speed profiles in advance of warning signs can be assessed with pavement instrumentation capable of recording vehicle speed distribution data when installed at critical distances in advance of warning signs to be evaluated. Necessary considerations are AASHTO-defined driver stopping sight distances, e.g., requirements for driver detection, recognition, and response to a traffic control device. Speed data gathered at this advanced distance are useful to define a baseline condition against which speeds influenced by tested TCDs can be compared. Speed profiles between the advance placement and the location of the targeted roadway condition provide insight regarding the effectiveness of the tested device.

The deployment of roadway instrumentation and timing of data collection intervals are based on specific conditions (e.g., weather, traffic volume) where the warning signs are intended to address. Separate speed profiles should be developed for targeted drivers (e.g., higher speed drivers). Supplementary driver questionnaires are useful to address critical issues such as driver observations and perceived credibility of the warning signs.

An example speed profile in advance of icy bridge warning signs is the study by Hanscom (14). The critical human factors issue addressed by this study was that bridges typically freeze before roadways; therefore, icy conditions may be unexpected. Designated speed-collection points permitted an assessment of the following:

- baseline driver speed absent influence of the bridge,
- driver speed as the bridge was clearly in view,
- speed at the critical bridge entry point
- maintained speed on the bridge.

Measures of driver behavior at roundabout signs and markings have included observation of erratic maneuvers and conflicts from overhead video recordings, an in-vehicle eye tracker to investigate driver gaze patterns (number and duration glances) and gaze direction while traversing multilane roundabouts.

Instrumented vehicles, driven in either test-track or open-road environments, are capable of gathering detailed driver human factors and operational data. Vehicle instrumentation typically includes recording equipment to document driver movements and verbal responses that are precisely associated with vehicle performance characteristics. More sophisticated installations include eye-marker equipment, whereby driver eye movements and dwell time provide a definitive measure of driver detection. It is noteworthy that drive eye movements are subconscious occurrences, thereby producing results that are uncontaminated due to their collection in a research setting.

A good example of the variety of measures than can be found from an instrumented vehicle study was that of Stout, et al. (15) who developed prototype work zone devices (barriers, delineators and signs) and tested these on an unused airport runway to avoid safety and liability problems of using experimental traffic control devices on public highways. TCD effectiveness measures gathered via the instrumented vehicle were:

- <u>Device Recognition Time</u> - The elapsed time from the moment when a test device comes into the subject's field of view until the subject recognizes it as a device that may affect driving.
- <u>Device Interpretation Time</u> - The amount of time required for the subject to interpret an appropriate response to the device.
- <u>Interpretation Correctness</u> - Whether the subject responded correctly to the device.
- <u>Interpretation Issue-</u> Whether the subject misunderstood the message of the device.
- <u>Helpfulness Rating</u> - Categorical scale: Very helpful, Helpful, Not very helpful, Not at all helpful.
- <u>Safety Rating</u> - Categorical scale: Much safer, Somewhat safer, A little safer, No safer at all
- <u>Approach Speed</u> - Speed of the vehicle at the time that the device first came into the driver's view.
- <u>Device Arrival Speed</u>- Speed when the vehicle arrived at the device location.
- <u>Approach Speed Profile</u> - Difference between the above two speeds.
- <u>Device Approach Time</u> - Elapsed time between the approach and arrival speed measurements.
- <u>Speed Variance</u> - The mathematical variance function based upon a set of speed measurements taken between the approach and arrival speed points.

Sampling

An important issue in any evaluation is adequate sampling. The appropriate number and type of road users should be selected to participate in the study. It may be difficult to get a "representative" sample of road users, but the sample should include an appropriate mix of age and gender, especially including older divers or pedestrians and novice drivers. Good sources of test subjects include driver licensing offices, church groups, senior citizens centers and service clubs. However, many TCD studies have involved young University students, hardly a representative sample. As an incentive, participants may have to be paid an honorarium or a donation made to the relevant organization. In the study by Dewar, et al. (5) older drivers performed worse on all measures than did middle-aged or young drivers, confirming the importance of including seniors in the sample. And the driving population is ageing, so there will be more older drivers on the road.

Context

One's ability to detect and recognize a TCD is influenced in part by the context in which it is seen. The road environment frequently provides a clue as to the message on the TCD. If a sign or signal is not understood when first seen, its meaning may become evident to the driver after it is encountered a few times, or even once (e.g., a curve seen ahead on the road beyond a curve warning sign helps the driver understand the sign). Hence, it is wise to place a TCD in its proper context in a laboratory setting when testing how well drivers can detect, read and understand it. Context can be provided to the subject in different ways. It can be shown in a picture of the sign in a road scene, described to the subject (e.g., this sign appears on a road in advance of a steep hill, intersection, or steep mountain).

Criteria for performance

It is necessary to decide what proportion of road users must understand and respond correctly in a timely manner to a sign or signal configuration, for example, in order for it to be considered effective. Criterion levels of 65 percent correct comprehension have often been used, but even at 65 percent comprehension, more than one-third of people would not understand the meaning intended. The cut-off used to determine TCD effectiveness depends upon the importance of the message and the consequences of failing to detect or understand the message. One would use a more strict (higher level) cut-off for a NO LEFT TURN or DO NOT ENTER sign than for a CAMPGROUND or NO LITTERING sign, for example.

REFERENCES

1.Shapiro, P. S., Upchurch, J. E., Loewen, J. and Siaurusaitis, V. (1987). Identification of needed traffic control devices research. Transportation Research Record #1114, 11-20.

2. Dewar, R. E. Criteria for the design and evaluation of traffic sign symbols. Transportation Research Record, 1988, #1160, 1-6.

3. Dewar, R. E. and Ells, J. G. 1974b A comparison of three methods for evaluating traffic signs. Transportation Research Record, 503, 38-47.

4. Dewar, R. E. and Ells, J. G. (1984). Methods of evaluation of traffic signs. In R. Easterby and H. Zwaga (Eds.) Information Design: The Design and Evaluation of Signs and Printed Material. Chichester: John Wiley, 77-90.

5. Dewar, R. E., Kline, D. W., Schieber, F. and Swanson, H. A. (1994). Symbol signing design for older drivers. Federal Highway Administration Report No. FHWA-RD-94-069.

6. Hanscom, F. R. (2010) .Drivers' Understanding of Innovative Roundabout Traffic Control Devices, ITE Journal, Volume 80, Number 7, Institute of Transportation Engineers, Washington, DC.

7. Ben-Bassat, T., Shinar, D., Almqvist, R., Caird, J. K., Dewar, R. E., Lehtonen, E., Salmon, P., Sinclair, M., Summala, H., Zakowska, L., and Liberman, G. (2019). Expert evaluation of traffic signs: conventional vs. alternative designs. *Ergonomics*, *62*(6), 734-747, 1-3.

8. Ben-Basset, T., Shinar, D., Caird, J. K., Dewar, R. E., Lehtonen, E., Sinclair, M., Zakowska, L., Simmons, S. M., Liberman, G. and Pronin, M. (2021). Ergonomic Design Improves Cross-cultural Road Sign Comprehension. Transportation Research, Part F, 78, 267-279.

9. Mackett-Stout, J. and Dewar, R. E. Evaluation of symbol public information signs. Human Factor, 23, 139-151, 1981.

10. Ells, J. G. and Dewar, R. E. Rapid comprehension of verbal and symbolic traffic sign messages. Human Factors, 1979, 21, 161-168.

11. Khavanin, M. R., and Schwab, R.N. (1991). Traffic sign legibility and conspicuity for the older drivers. In Compendium of Technical Papers from the 61st Annual Meeting of the Institute of Transportation Engineers. Milwaukee.

12. Cole, B. L. and Hughes, P. (1984). A field trial of attention and search conspicuity. *Human Factors*, 26, 299-313.

13. Hanscom, F.R. and Berger, W.G. (1977). *Motorist response to highway guide signing. Volume I. Field evaluation of measures. Final Report.* NCHRP Project 3-21). NCHRP Research

Results Digest, Digest 91, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC.

14. Hanscom, F.R. (1975). An Evaluation of Icy Bridge Warning Signs, Past Presidents Award Recipient Paper, 45th Annual Meeting Compendium, Institute of Traffic Engineers, Washington DC.

15. Stout, D., Graham, J., Fish, J. and Hanscom, F. (1993). Maintenance Work Zone Safety Devices, Development and Evaluation. Final Report SHRP-H-371, Strategic Highway Safety Research Program, National Research Council, Washington, DC.