

## ***EVALUATING DESIGN OPTIONS FOR A DYNAMIC TRAFFIC SIGN***

Janet Creaser<sup>1</sup>, Michael Rakauskas<sup>1</sup>, Michael Manser<sup>1</sup>, Nicholas Ward<sup>2</sup>

<sup>1</sup>HumanFIRST Program, ITS Institute, University of Minnesota  
Minneapolis, MN, USA

<sup>2</sup>Western Transportation Institute  
Bozeman, MT, USA

Email: janetc@me.umn.edu; mickr@me.umn.edu;  
mikem@me.umn.edu; nward@ie.montana.edu

**Summary:** This study describes two usability methods that were used to determine the final design of a prototype dynamic traffic sign. The Cooperative Intersection Collision Avoidance System-Stop Sign Assist (CICAS-SSA) is an infrastructure-based driver support system to improve gap acceptance at rural stop-controlled intersections. This study evaluated drivers' comprehension of recommended design changes made to the SSA message set using paper-and-pencil and computerized testing. The goal was to choose the final interface design that would later be tested using driving simulation. Overall, comprehension was highest for sign messages that showed prohibitive information and was lowest for signs indicating no traffic was detected near the intersection. The results for the design options were similar between studies, allowing for the selection of a final set of design features for the interface. Results also suggest that the two methodologies provided a low-cost alternative to simulation for down-selecting the design options.

### **INTRODUCTION**

The Cooperative Intersection Collision Avoidance System-Stop Sign Assist (CICAS-SSA) is an infrastructure-based driver support system to improve gap acceptance at rural stop-controlled intersections (Laberge, Creaser, Rakauskas & Ward, 2006; Creaser, Rakauskas, Ward, Laberge & Donath, 2007). The SSA tracks vehicle locations on the major road and displays information about approaching traffic to the driver on the minor road. Based on the results of the initial study and how best to meet MUTCD guidelines, one of the three prototype concepts received a number of design change recommendations (see Table 1). The Countdown sign uses a timer to show how far away approaching vehicles are in the nearest set of lanes, while the icon on top provides a judgment about the safety of the available gap for the near and far lanes. For this paper, only the sign messages displayed for drivers trying to cross or turn from the stop sign are discussed. Because of the multiple design recommendations, it was necessary to determine which were best before this prototype concept was to be used in the final phase of simulator testing.

A number of usability methods exist for testing traffic sign comprehension. Ideally, new sign designs would be tested in simulation or on the road. Chrysler et al. (2004) noted that, conceptually, a driver executing the correct maneuver or behavior at the appropriate time after viewing a traffic sign would indicate true comprehension of the sign. However, running an additional simulation study was not cost effective or feasible to down-select the design options

for the Countdown prototype. Therefore, two methods frequently used to determine icon and sign comprehension were employed to determine the best design options for this sign’s states.

The first method was a paper-and-pencil test based on work by Campbell et al. (2004a; 2004b) that was developed to test in-vehicle icons for ITS systems. Participants first view the sign and then write down what they think it means. After this, they are presented with the correct meaning of the sign and asked to rank their preference for each design option. An option with high comprehension rates and high preference rankings is likely a good design.

**Table 1. Countdown Sign design changes**

Message State	Prototype Concept	Prototype Description	Proposed Design Options to be Tested	Change Descriptions
Do Not Enter; Traffic Too Close in Near Lanes		Red timer background; red “do not enter” icon.	  	Red “do not enter” icon changed to yellow diamond. Icon options: 1. crash icon. 2. “wait” with red hand icon. 3. “do not enter” text.
Do Not Cross/ Turn Left; Traffic in Far Lanes		Black timer background; red circle/slash “no left turn/cross” icon.	 	Red circle & slash changed to yellow diamond or yellow circle. Tested options: 1. Yellow diamond with black lines. 2. Yellow circle with black arrows.
Proceed with Caution; No Traffic Detected		Black timer background; yellow “caution” diamond.	 	“Caution” text changed to “Look for Traffic”. Tested options: 1. Yellow rectangle. 2. Yellow diamond.

The second method used timed presentation of the sign design options to determine comprehension. Images of the sign were shown on a screen for a limited time and then participants selected a response from a multiple-choice list. Chrysler et al. (2004) found that a timed presentation method using a 3 s exposure produced similar comprehension rates of static signs when compared to comprehension rates obtained through the use of an interactive driving simulator for the same signs. The limited time presentation may produce a cognitive load similar to that observed in a simulator or on the road. In the current study, participants viewed each design option three times using presentation times of 1 s, 2 s and 3 s. Because dynamic traffic signs are rare, these methods have not been previously employed to examine the comprehension of dynamic sign messages. This paper will also examine some of the issues with adapting these methods to a dynamic traffic sign.

## METHODS

### Experiment1

*Participants.* Sixty participants were recruited for Experiment 1 in three age groups, each comprised of 20 participants (10 male; 10 female). The age groups were Young (18-25;

$M=22.9$ ), Middle (30-55;  $M=42.6$ ) and Older (60+;  $M=65.6$ ). All participants held a valid driver's license and reported driving at least a few days every week. Participants were recruited through a local recruiting agency and paid \$40 cash for their participation at the end of the study.

*Materials.* Two test booklets were created for the paper-and-pencil comprehension and ranking tests. The test booklets were identical, other than the presentation order of the signs. Signs were randomly ordered for Version 1 and were presented in reverse order for Version 2.

*Procedures.* Participants completed the informed consent process and were provided an oral introduction to the study. The Study Introduction described the context in which the signs would be used and showed a diagram of a stop-controlled intersection. A video of a simulated intersection was shown to demonstrate how traffic flowed at the intersection and what maneuvers were available to drivers trying to enter from the stop sign (i.e., turn left, turn right, cross over). The written context for the introduction provided for this study was developed using the guidelines in Campbell et al. (2004a). Too little or inappropriate context may result in unrealistically low comprehension while too much context may result in unrealistically high comprehension of the candidate signs. Pilot testing was used to finalize the descriptions and ensure that participants understood the signs and messages they would see were dynamic. Participants then completed the booklet at their own pace.

## Experiment 2

*Participants.* Sixty participants were recruited for Experiment 2 in three age groups, each comprised of 20 participants (10 male; 10 female). The age groups were Young (18-29;  $M=24.3$ ) Middle (30-55;  $M=42$ ) and Older (60+;  $M=63.9$ ). All participants held a valid driver's license and reported driving at least a few days every week. Participants were recruited through a local recruiting agency and paid \$40 cash for their participation at the end of the study.

*Apparatus.* Participants were seated inside the HumanFIRST driving simulator and the sign images were presented on the forward screen using E-prime software (v1.1; Psychology Software Tools, 2003) running on an IBM-compatible PC running Windows XP. Participant responses were logged using buttons on the Psychology Software Tools' Serial Response Box. The response box buttons were labeled to match the response options shown on the screen (A-D).

*Procedures.* Participants completed the informed consent process and were provided with the same contextual introduction that was used in Experiment 1. Screens were presented in the following order: instructions (unlimited viewing time), fixation screen (3 s presentation), image presentation (1 s, 2 s, or 3 s), response screen with the multiple choice options listed. Once participants finished the sequence, they pressed the "start" button and began the sequence again until all trials were complete. Presentation was blocked by time and counterbalanced across participants, such that participants saw all of the images at one time presentation (e.g., 1 s) before moving onto the next time block (e.g., 2 s or 3 s). Images were randomized within each time block by the software. Participants were asked to respond as accurately and as quickly as possible once the image had been presented. Only one response was correct for each of the sign images. The options were: A. Do not enter the intersection; B. Can enter the intersection to turn

right only; C. Can enter the intersection to cross over, turn right or turn left; D. I do not know what the sign means. The timed response was measured from when the response options screen appeared until participants entered a response on the response box.

## RESULTS

Data collected in Experiment 1 was analyzed by two researchers using the scale described in Table 2 (Campbell et al., 2004b). A high percentage of high comprehension scores (1 or 2) should be balanced with a low rate of critical confusions for a design to be both well understood and safe. The level of inter-rater reliability was determined using a consensus estimate approach. Consensus estimates of inter-rater reliability assume that reasonable observers can come to exact agreement on how to apply the various categories of a scoring system to the responses (Stemler, 2004). The percent agreement for this study was 83.3% between the two raters. Inter-rater reliability is considered good when percent agreement is 70% or greater (Stemler, 2004).







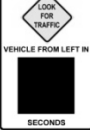
**Table 2. Rating scale for scoring subject responses to the signs (from Campbell et al., 2004b)**

Score	Description	Implications for Comprehension
1	The response matches the intended meaning of the icon exactly.	Scores of 1 or 2 indicate HIGH comprehension of the sign's meaning.
2	The response captures all major informational elements of the intended meaning of the icon, but is missing one or more minor information elements.	
3	The response captures some of the intended meaning of the icon, but it is missing one or more major informational elements.	Scores of 3 or 4 indicate a partial, or LOW, understanding.
4	The response does not match the intended meaning of the icon, but it captures some major or minor informational elements.	
5	The response does not match the intended meaning of the icon, but it is somewhat relevant.	Scores from 5-8 indicate responses that show no comprehension of the sign's meaning (NONE).
6	Participant's response is in no way relevant to the intended meaning of the icon.	
7	Participant indicated he/she did not understand the icon.	
8	No answer.	
9	For safety-critical icons, identify the number and percentage of critical confusions or errors. Critical confusions or errors reflect responses that indicate that the subject perceived the message to convey a potentially unsafe action.	A score of 9 means a critical confusion occurred (CC),

Data collected in Experiment 2 was used to calculate the total percentage correct for each presentation (by time period) for each age group and the average response time to an image.

Table 3 shows the results for both Experiments.

**Table 3. Countdown Sign Comprehension Rates**

Message	Signs	Experiment 1 Comprehension				Experiment 2 Results			
		High 1-2	Low 3-4	None 5-8	CC* 9	Rank	% Correct	Mean RT	Age effect
Do Not Enter; Traffic Too Close	1 	25%	42%	20%	13%	20%	85%	2.80 s	No
Do Not Enter; Traffic Too Close	2 	58%	18%	5%	18%	33%	83.9%	2.88 s	No
Do Not Enter; Traffic Too Close	3 	43%	35%	3%	18%	45%	80%	2.98 s	No
Do Not Cross/Turn Left	4 	7%	32%	33%	28%	45%	59.4%	5.65 s	Yes Y:85% M:53% O:43%
Do Not Cross/Turn Left	5 	10%	35%	20%	35%	55%	60.6%	4.85 s	Yes Y:75% M:57% O:47%
Proceed with Caution	6 	27%	48%	17%	8%	77%	65.6%	4.46 s	No
Proceed with Caution	7 	57%	22%	18%	3%	23%	66.1%	3.84 s	No

\* CC = Critical confusions.

## DISCUSSION

This study used two common methods for assessing the comprehension of traffic signs and applied them to the dynamic design options for the Countdown SSA message states. The two Do Not Enter options that performed best both used text (“do not enter”, “wait”). Because the goal of the SSA for this message state is to get drivers to wait for a less risky gap, both of these options likely work because the words clearly indicate what to do. In contrast, the crash icon (1) was too abstract in Experiment 1, with participants indicating they thought it meant the intersection was a high crash zone or that trucks were present. Although the “wait” icon (3) was most preferred, the “do not enter” (2) option had better comprehension rates in both studies.

Therefore, it was recommended as the best design option for this message on the SSA Countdown sign.

It was more difficult to determine which Do Not Cross/Turn Left option should be selected. Both options had very poor comprehension in Experiment 1 combined with high rates of critical confusions. It is likely this message state itself is problematic because participants must interpret the bottom portion (timer) of the sign as indicating the near lanes are clear while using the top portion (icon) to understand that traffic is detected in the far lanes. When multiple choice options were provided in Experiment 2, comprehension increased significantly. This suggests that drivers were better able to discern what this sign state meant when provided with clear behavioral options. Overall, it was decided that the small improvements in performance seen with the yellow circle/black arrows option (5) would have to suffice for moving onto the next phase of testing in the driving simulator. However, these basic usability tests encouraged the researchers to examine this sign state carefully during the simulator experiment.

For the Proceed with Caution option, the diamond (7) performed better than the yellow rectangle (6) in Experiment 1, but comprehension and response times were similar for the designs in Experiment 2. The high preference for the rectangle was attributed to its larger size in the images based on participant comments. This result was an artifact of the study materials. All text and icons would meet approved MUTCD standards for size and legibility when deployed, but this was not easily conveyed using a paper presentation of the sign messages. Although the diamond performed better overall, there were concerns that using a diamond with text for two of the sign's message states (Do Not Enter & Proceed with Caution) could result in drivers missing sign changes as the sign flipped between messages in its active state. For this reason, the rectangle was chosen instead of the diamond to prevent possible confusion between the Do Not Enter and Proceed with Caution states. An error in identifying differences between the two states could be catastrophic in the real world given their two very different meanings.

Age is a known factor in the comprehension of traffic signs and symbols (e.g., Shinar et al., 2004; Dewar et al., 1994). In the previous IDS study older drivers (age 55+) had lower comprehension for all the prototype sign concepts when compared to the younger age group (18-40) (Creaser et al., 2007). In the current study, a clear age effect emerged for older drivers in the Do Not Cross/Turn Left state in Experiment 2. This sign state may be problematic because it requires the integration of information about the near and far lanes using two distinct pieces of information (bottom & top of sign), which may be more difficult for older drivers because of normal declines in information processing capacity.

Overall, the two usability studies highlighted problems with potential design options, but neither was optimal by itself for identifying a final design. The timed multiple-choice test uses objective performance measures, which is an attractive feature of the method. This is because it does not require inter-rater reliability to be calibrated like the paper-and-pencil test does. However, in this study, the timed presentation produced similar comprehension rates across design options within a set. Because of this similarity in performance, this method would not have been sufficient by itself to identify the best design options. For example, the crash icon (1) performed equally well compared to the other two options, but performed poorly in the paper-and-pencil test. Without the results and comments from the paper-and-pencil test, this test would not have identified the

problems with this icon. It also did not appear that this method sufficiently taxed cognitive load to produce differences between the designs, even with short presentation times. This may have been a problem with the wording of the multiple choice options or may simply be a function of having the options present. In future, this method might benefit from having participants rank design options at the end, as in Experiment 1. Finally, learning effects were noticed using this method. Regardless of presentation time order, subsequent viewings resulted in a higher percentage correct. To prevent learning effects a single time presentation should be used to avoid showing a single design more than once.

In comparison, the paper-and-pencil test produced varying comprehension rates across design options. The main benefit of this method is the identification of critical confusions, where drivers assume the sign means something that may lead them into a safety critical situation when interacting with the signs. In this study, the critical confusions identified for the Do Not Cross/Turn Left sign messages helped the researchers understand problems with this message state.

## CONCLUSIONS

Overall, these two methods were useful for down-selecting design options for dynamic traffic signs when the appropriate context was provided to participants to help them understand the dynamic nature of the sign. However, a realistic simulation of a dynamic sign's functions is likely to produce the best assessment of comprehension, particularly when drivers are required to interact with the sign, such as in a driving simulator.

## ACKNOWLEDGEMENTS

We wish to acknowledge the support by the US DOT / FHWA (Cooperative Agreement Number DTFH61-07-H-00003) and Minnesota DOT (Contract No. 89031).

## REFERENCES

- Campbell, J.L., Richman, J.B., Carney, C., & Lee, J.D. (2004a). *In-vehicle display icons and other information elements volume 1: Guidelines* (Rep. No. FHWA-RD-03-065). McLean, VA: Federal Highway Administration.
- Campbell, J.L., Hoffmeister, D.H., Keifer, R.J., Selke, D.J., Green, P.A., & Richman, J.B. (2004b). Comprehension Testing of Active Safety Symbols (2004-01-0450). *2004 SAE World Congress and Exhibition Technical Papers*, Detroit, MI, March 8-11, 2004.
- Chrysler, S.T., Wright, J., & Williams, A. (2004). *3D Visualization as a Tool to Evaluate Sign Comprehension* (Rep. No. SWUTC/04/167721-1). College Station, TX: Southwest University Transportation Center.
- Creaser, J.I., Rakauskas, M.E., Ward, N.J., Laberge, J.C., & Donath, M. (2007). Concept evaluation of intersection decision support (IDS) system interfaces to support drivers' gap acceptance decisions at rural stop-controlled intersections. *Transportation Research Part F*, *10*, 208-228.

- Dewar, R.E., Kline, D.W., & Swanson, H.A. (1994). Age differences in comprehension of traffic sign symbols. *Transportation Research Record 1456*, 1-10.
- Laberge, J.C., Creaser, J.I., Rakauskas, M.E., & Ward, N.J. (2006). Design of an intersection decision support (IDS) interface to reduce crashes at rural stop-controlled intersections. *Transportation Research Part C, 14*, 39-56.
- Shinar, D., Dewar, R.E., Summala, H., & Zakowska, L. (2003). Traffic sign symbol comprehension: a cross-cultural study. *Ergonomics, 46*(15), 1549-1565.
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*, Retrieved May 24, 2007 from <http://PAREonline.net/>.