STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

ADA Notice
For individuals with sensory disabilities, this document is available in alternate formats. For information call (916) 654-6410 or TDD (916) 654-3880 or write Records and Forms Management, 1120 N Street, MS-89, Sacramento, CA 95814.

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA21-3105		

4. TITLE AND SUBTITLE	5. REPORT DATE
Information Display Board for Corridor Management in California	10/30/2020
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Dr. Ching Yao Chang Dr. Peggy Wang	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
RICHMOND FIELD STATION 1357 S.46th Street, Bldg. 452 MC 3580 Richmond, CA 94804-4648	11. CONTRACT OR GRANT NUMBER Contract 65A0637 Task ID 3105
12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Division of Research, Innovation and System Information California Department of Transportation 1727 30th Street, 3rd Floor, MS-83 Sacramento, CA 94273-0001	07/01/2018 to 06/30/2020 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report documents the efforts conducted under contract 65A0637, task ID 3105. Under this project the contractor (University of California Berkeley, PATH Program) conducted research to gain Federal Highway Administration (FHWA) approval for experimental use of the text and graphical messages designs on Information Display Boards (IDB) in the field. Some of the suggested concepts of displays and IDB displays do not comply with the standards and guidelines of the current Manual on Uniform Traffic Control Devices (MUTCD). This research was structured to conduct a systematic and well-grounded human-factor study for the proposed IDB display designs that convey information in an instructive and helpful manner to the traveling public.

17. KEY WORDS	18. DISTRIBUTION STATEMENT	
Manual on Uniform Traffic Control Devices (MUTCD), Information Display Board, Changeable message Signs,		
19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	79	

DISCLAIMER STATEMENT

Information Display Board for Corridor Management in California
FINAL REPORT

Peggy Wang & Ching Yao Chan

CALIFORNIA PATH, UC BERKELEY

Table of Contents

List of Figures

List of Tables

Executive Summary

California Department of Transportation (Caltrans) has proposed to display graphical messages and text messages on Information Display Board (IDB) along Interstate 80 between the Carquinez Bridge and the Bay Bridge. The purpose is to convey additional traveler information, such as estimated travel time to popular destinations and expected delays, in order to enhance drivers' knowledge of the downstream traffic conditions and to provide information about alternative routes and alternative transportation modes. California PATH, on behalf of Caltrans, conducted an independent research project using human-factors approaches, aiming at generating a recommendation report in order to be used by Caltrans to obtain approval from FHWA. Besides, this project also aimed to conduct fundamental research for the evaluation of IDBs, with knowledge and insights that can be potentially generalized on other corridors in California and across the country.

As summarized in Part I of this report, a laboratory test was conducted to evaluate the understandability of different designs. Five categories of messages were proposed to be displayed: travel time messages, transit travel time messages, special messages, up-to-six-lines messages, and Graphic Route Information Panels (GRIPs). Varying factors such as with/without symbols, sizes and designs of symbols, and color combinations were considered for the sign design. The effect of test subjects' native language was also investigated. In total, fifty-two local commuters of Intestate 80 Smart Corridor took part in the study. The results of the lab testing showed that symbolic signs were superior to pure-text signs in terms of understandability. Large symbols with simple representation were better understood than small symbols. The viewing time of GRIPs, especially dual-link GRIPs, was significantly longer than other signs.

In the second phase of the project, a simulator test was conducted to further evaluate three categories of messages, which is summarized in Part II of this report. The three categories of messages included: (1) up-to-six lines information in a single phase to display travel-time messages; (2) transit travel-time messages; and (3) GRIPs presenting color-coded congestion levels as well as travel time. Twenty-four local commuters participated in the study. Results showed that: (1) perceived easiness for the five-line and six-line travel time message were significantly lower than the three-line travel time message; (2) perceived easiness and symbol helpfulness for the BART (Bay Area Rapid Transit) logo was significantly higher than the generic light-rail symbol; (3) for single-link GRIP, displays without a legend and with the bottom-up orientation were perceived as easier to understand; (4) understanding accuracy of single-link GRIP with 3 destinations was significantly higher than single-link GRIP with 4 destinations; (5) single-link GRIPs were perceived as easier to understand in comparison with dual-link GRIPs; and (6) participants took longer time to perceive the dual-link GRIPs.

For the third and fourth phase of this project, the plan was to do a field testing followed with focus group discussions in order to evaluate the legibility distance of the IDB signs, which were proposed to be displayed on the I-80 Corridor. The PATH team did a literature review about the legibility distance of CMS and proposed the experimental design, including the testing routes, testing time, participants, testing procedure, as well as measurements. The PATH team also reviewed the testing plan with engineers from Caltrans D4 team and revised the experimental set-up based on their feedback. The test plan for the field testing is summarized in Part III of this report.

In order to conduct the experiment with the suggested displays on the IDBs, the PATH team drafted a report that supported Caltrans' request of experiment permission from FHWA. In the report, we first summarized the experiment set-up and findings of both the lab testing and the simulator testing. We also explained the potential benefits and importance of further testing the IDB signs in the field. However, after review, the FHWA denied the request due to various considerations. Subsequently, the PATH team drafted and provided a rebuttal response for Caltrans' consideration to address issues raised by FHWA. In Part IV of this report, we summarize the request for permission to experiment and also the rebuttal letter to FHWA.

The field testing could not be implemented as a result of FHWA's denial of the request for permission to carry out the on-road experiment. As an alternative, the PATH team suggested doing another lab testing in order to address FHWA's comments on extended viewing time. The lab testing was proposed to use a protocol of fixed display time (e.g., 6 seconds, or 8 seconds) controlled by the testing program rather than free viewing time controlled by the participants. This will help evaluate the effect of learning and familiarity after the subjects have the opportunities to view the displays multiple times. Besides the revised lab testing, the PATH team took a further look at other potential issues and concerns that were not fully answered in the original project scope and suggested other topics for future studies beyond this project, which is summarized in Part V of this report.

In summary, this project shed light on various aspects of message design for CMSs using advanced display technologies. Based on the results of this project, we recommend that advanced CMS with higher resolution LED displays, like the IDBs, should maintain the information on one sign with less than five lines of message texts. When the objective of the CMS is to provide traveler information for local commuters, drivers' familiarity with the objects (e.g., bus, BART) should be considered for selecting or designing the symbols. The well-known symbols (e.g., BART logo) should be used over other standards symbols (e.g., light-rail). This finding should apply to other locally well-known brands of light-rail transit services as well. Single-link GRIPs with 3 destinations have similar understandability as the three-line message. As single-link GRIPs also provide more information (congestion level) than the three-line message, they are recommended for further evaluation and implementation. However, adding more information onto the sign, such as the road-work legend or incident legend, will increase the complexity of the sign, which is not recommended. For the same reason, dual-link GRIPs with three interchanges or destinations are not recommended either. Going forward, on-road experiments to collect field data will provide additional insights regarding the overall performance of the IDBs, which will allow researchers to evaluate the legibility distance of the IDB displays as well their potential impacts on the traffic operations.

Part I: Lab Testing

Summary

Six advanced changeable message signs have been installed by the California Department of Transportation along a segment of Interstate 80 in the San Francisco Bay area to convey traveler information. The advanced changeable message signs are also referred to as Information Display Boards (IDBs), which use LED technology and provide full-color display aiming for better legibility and enhanced recognition. Five categories of messages are proposed to be displayed: travel time messages, transit travel time messages, special messages, up-to-six-lines messages, and Graphic Route Information Panels (GRIPs). Varying factors such as with/without symbols, sizes and designs of symbols, and color combinations, are considered for the sign design. A lab testing is conducted to evaluate the understandability of different designs. The effect of native language is also investigated. In total, fifty-two local commuters of the I-80 Corridor took part in the study. The results of the lab testing show that symbolic signs are superior to pure-text signs in terms of understandability. Large symbols with simple representation are better understood than small symbols. The viewing time of GRIPs, especially dual-link GRIPs, is significantly longer than other signs. The consistency of message content, symbol design, and color combinations increase the understandability of signs. Participants' education and language have a significant influence on understanding accuracy.

Section 1: Introduction

California Department of Transportation (Caltrans) has proposed to display graphical messages and text messages on Information Display Board (IDB) along Interstate 80 between the Carquinez Bridge and the Bay Bridge. The purpose is to convey additional traveler information, such as estimated travel time to popular destinations and expected delays, in order to enhance drivers' knowledge of the downstream traffic conditions and to provide information about alternative routes and alternative transportation modes. With the existing Model 500 Changeable Message Signs (CMS), three primary types of incident information are displayed to provide drivers with details of incidents and freeway conditions. The incident information categories include (1) what happened (e.g., ACCIDENT AT ASHBY), (2) where (e.g., LEFT LANE BLOCKED), and (3) driver action (e.g., EXPECT DELAYS). With the IDBs, in addition to incident information mentioned above, Caltrans also plans to display public safety messages such as adverse weather conditions in addition to a more diversified set of traveler information mentioned above. The IDBs have a form factor that differs from the typical Model 500 CMS. The visible optical area is 157 1/2" x 196 7/8", with a resolution of 216 x 270 pixels full matrix. It includes technological features to provide a full-color display in order to provide better legibility and enhanced sign recognition.

Selective preliminary concepts of graphical images are shown in Figure 1[1]. The preliminary design of such displays has been reviewed by the California Traffic Control Device Committee (CTCDC). Caltrans has engaged Federal Highway Administration (FHWA) and is seeking consensus and approval on the contents of the IDBs that can be deployed on the I-80 Smart Corridor[2]. Real-world experimentation of up to six lines message display concepts has been tentatively approved by FHWA. Caltrans is seeking approval from

---

[1] Supporting documents provided by Caltrans D4: *(1) Information Display Board White Paper Previous Final Revisions – May 15, 2012* and (2) *Request for Permission to Experiment with the Messages and Graphics on Dynamic Message Signs on Freeway.*

[2] Discussions with Caltrans D4 and DRI on October 31, 2016.

FHWA on the design of additional IDB displays that can be deployed for experimental purposes on the I-80 Corridor. [3]

a b c

Figure 1. Selective Preliminary Concepts of Graphical Images

California PATH, on behalf of Caltrans, is conducting an independent research project using human-factors approaches, aiming at generating a recommendation report in order to be used by Caltrans to obtain approval from FHWA. Besides, this project aims to conduct fundamental research for the evaluation of IDBs, which has the potential to be generalized on other corridors in California and across the country.

From a survey of traffic sign experts in Australia, New Zealand, and the United States, Dewar (1988) examined several criteria for traffic sign symbols and the importance of the criteria that should be assigned to each symbol in the design and evaluation process. Understandability was the factor rated most important, with conspicuity second. Reaction time, legibility distance, and glance legibility were rated equally important. Learnability was considered the least important. Using a driving simulator, Charlton (2006) compared multiple measures of sign processing, including attentional conspicuity (participants were instructed to say aloud the name of anything that attracted their attention), search conspicuity (participants were instructed to name aloud hazards or hazard warning signs whenever they noticed them), implicit and explicit recognition, dynamic comprehension (correct identification while driving), static comprehension (correct identification when presented as a still image at the end of the session), and sign priming. It was found that attentional and search conspicuity and static comprehension were the most reliable indicators of a sign's overall performance. Hence, from a systematic point of view, a four-stage plan has been proposed for this project, which includes lab testing, simulator testing, field testing, as well as focus group discussions based on the field testing.

The first phase of this project is the lab testing, which aimed to evaluate the understandability of various messages and different graphical designs. More specifically, the objectives of the lab testing were threefold:

- To measure comprehension accuracy of each message;
- To measure the time that drivers need to interpret each message; and
- To identify design elements that cause confusion.

[3] Discussions with Caltrans D4, DRI and FHWA representative in the project kick-off meeting on June 19, 2017.

The lab testing was completed in April 2018. In this part, we summarize the experimental design and the implementation as well as present results and findings of the lab testing. In addition to this introductory section, this part contains the following four sections: (1) Description of all messages and alternative design concepts; (2) Experimental design and testing procedures of the lab testing; (3) Data analysis and results; and (4) Discussion and recommended principles for IDB design.

Section 2: Five Categories of Messages and Alternative Designs

Caltrans District 4 provided five categories of messages, which have been proposed to be displayed on the IDBs. For each category of message, alternative designs were proposed and evaluated in the lab testing. In this section, we will first describe the five categories of messages. Then we will explain the design factors considered for each message category and show all the alternative designs.

Five Categories of Messages

1. Travel Time Messages

The purpose of travel time messages is to convey estimated travel time to reach popular and noticeable destinations, including conjunctive highways and airports. Examples of travel time messages are shown in Figure 2. Figure 2a and Figure 2b show travel time to conjunctive highways. Figure 2c shows travel time to three important airports near the I-80 corridor.

a b c

Figure 2. Examples of Travel Time Messages

2. Transit Agency Logos for Transit Travel Time Messages

Transit travel time messages are trying to convey the information of travel time from certain starting points to various destinations by taking public transportations, such as BART and bus (AC Transit and WESTCAT). Transit agency logos are located on the top of the sign.

a b c

Figure 3. BART, AC Transit and WESTCAT Travel Time Messages

3. Special Messages

The purpose of special messages is to convey information about incidents, roadwork, adverse road conditions, as well as the abnormal status of train, BART, and bus services. In total, there are six special messages, which are ACCIDENT AHEAD, ROADWORK, SLIPPERY, NO BART SERVICE, DELAYED BUS SERVICE, and DELAYED CAPITOL CORRIDOR SERVICE. Figure 4 shows two examples of the special messages, with Figure 4a alerting drivers of an accident ahead, and Figure 4b indicating that BART service is not available.

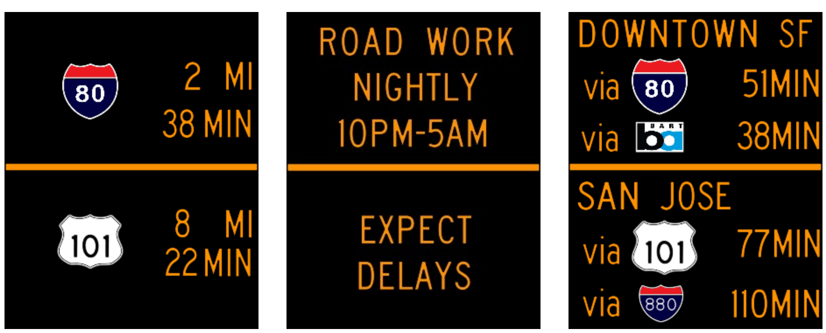
a b

Figure 4. Examples of Special Messages

4. Up to Six Lines Messages

As shown in Figure 5, up to six lines of messages display both graphical messages and text messages containing up to six lines of information with a clear horizontal line to break the information into two virtual phases. Figure 5a conveys a message of distance and travel time to two conjunctive highways. Figure 5b conveys a message of roadwork alert, time, and driver action. Figure 5c conveys a message of travel time to two cities, with choices of transportation modes and alternative routes.

According to the discussion on the California MUTCD[4], it is suggested to use up to 3 lines of 20 alphanumeric characters and no more than two phases in a display, which is for the text-based messages only. The text messages on the IDBs will follow the guidance on CA MUTCD. However, more than 3 lines of text will be displayed on the IDBs.



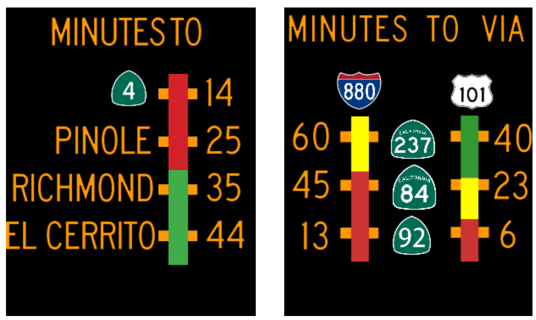
a b c

Figure 5. Examples of Up to Six Lines Messages

[4] Section 2L.04 of California MUTCD 2014 edition.

5. Graphic Route Information Panel (GRIP)

GRIPs include single-link and dual-link GRIP. Caltrans would like to explore if link-based GRIPs are more informative than text-based displays. Figure 6a is an example of a single-link based GRIP, presenting travel time to four different destinations, including one highway and three cities. Figure 6b shows an example of a dual-link based GRIP, with two links presenting travel time to different destinations along two different routes. The colors used to represent congestion levels on the GRIP are green (normal operating speed), yellow (slow traffic conditions), and red (stop-and-go conditions). The three colors have been proved well interpreted by drivers and are more helpful for drivers to make decisions of selecting alternative routes compared with text-based messages, according to a study conducted by the Texas Transportation Institute (Ullman et al., 2009). The GRIP will be limited to displaying up to 4 interchanges and/or destinations in a linear format. Caltrans would like to validate that the simpler link-based GRIP is beneficial.



a b

Figure 6. Examples of Single-link GRIP and Dual-link GRIP

Alternative Designs for Each Category of Messages

The design of the five categories of messages provided by Caltrans serves as a baseline, from which alternative designs have been derived. Different factors are considered in order to compare the effects of different design elements and explore a better design for each message. Alternative design factors such as with/without symbols, various symbol designs, different color combinations for each category of messages are summarized in Table 1.

Table 1. Design Factors for Each Message Category in the Lab Testing

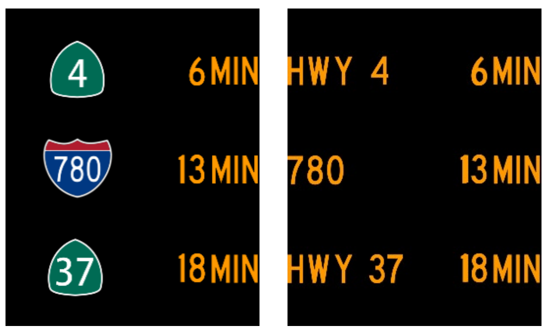
Message Categories	Design Factors
1. Travel time messages	• Symbol & Text vs. Text only
2. Transit agency logos for transit travel time messages	• Color combinations for logo • Symbol & Text vs. Text only
3. Special messages	• Symbol & Text vs. Text only • Design of symbols
4. Up to six lines of messages	• Color of the horizontal line
5. GRIP	• The complexity of GRIPs (Units of information)
5.1 Single-link GRIP	• Orientation: top-bottom vs. bottom-top • With or without accident/roadwork legends
5.2 Dual-link GRIP	• Positions of destinations (left vs. middle)

- Positions of route shields (top vs. bottom)

Due to the constraint of the dimensional size of the IDB board, font size and character spacing cannot be arbitrarily changed. The default font size height is 24 pixels, the default font size width is 14 pixels, and the default character spacing is 4 pixels. Colors for legends and background are applied by following the guideline on CA MUTCD. A detailed explanation of the design factors and the graphical images of alternative designs for each message category are listed below.

1. Travel Time Messages

In order to compare the IDB displays using both symbols and texts versus the existent displays (e.g., Model 500 CMS) using only texts, the travel time messages are also displayed in pure text format. An example is shown in Figure 7. The purpose is to explore whether the IDB displays with symbols provide more helpful information to drivers.



a b

Figure 7. Alternative Designs of Travel Time Messages

2. Transit Agency Logos for Transit Travel Time Messages

There are four alternative graphical designs for each message. In one design, the transit agency logos have their original color combinations, for example, the black and blue color combination for the BART logo, as shown in Figure 8a. In another two designs, different combinations for the color of logo and color of background are proposed. Taking the BART logo as an example, the two combinations are a black logo with a white background and a white logo with a black background, as shown in Figure 8b and Figure 8c, respectively. In the fourth design, the transit travel time message is in a pure text format, as shown in Figure 8d, in order to compare the IDB display with existent displays. The transit travel time information uses amber color for the text and black color for the background. Similar design considerations are also applied to AC Transit and WESTCAT travel time messages.

a b c d

Figure 8. Alternative Designs of BART Logo with BART Travel Time Messages

3. Special Messages

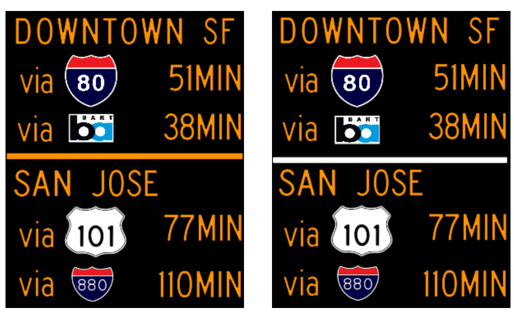
Two design factors are considered for special messages. One design factor is whether having symbols in the message is easier to understand. For messages of abnormal transit services (i.e., no BART service, delayed bus service, delayed train service), the other design factor is the format of symbols, including shape, size, and color. Two pairs of exemplar designs for special messages are shown in Figure 9. Figure 9a and Figure 9b are ACCIDENT AHEAD messages with a symbol and without a symbol. Figure 9c and Figure 9d are NO BART SERVICE messages with a large symbol, and with a small symbol and a red circle on top of it.

a b c d

Figure 9. Alternative Designs of Special Messages

4. Up to Six Lines Messages

In order to break the messages into two virtual phases, the color of the horizontal line is considered as a design factor for up to six lines messages, which is hypothesized to have different visual effects. Yellow and white are considered, as shown in Figure 10**.**

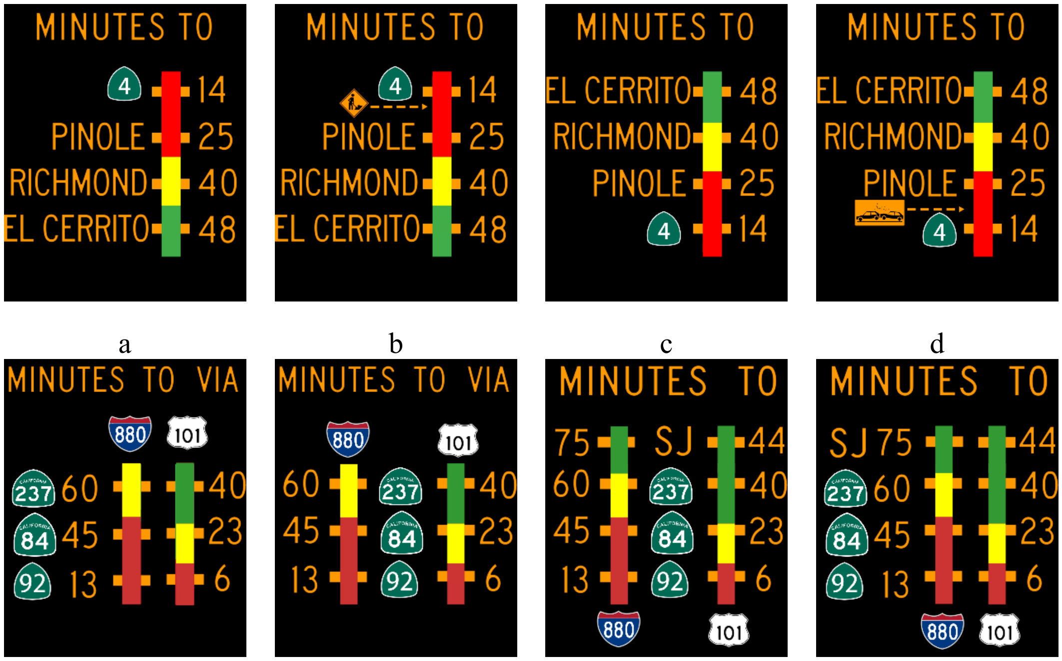


a b

Figure 10. Alternative Designs of Up to Six Lines of Messages

5. GRIPs

Five design factors are considered for GRIPs. The first factor is the complexity of the messages, which is presented by the number of links (single-link or dual-link) and the number of destinations (3 or 4 destinations). For single-link GRIPs, another two factors are considered: (1) orientation of approaching destinations (top-bottom or bottom-top), and (2) with and without roadwork/accident legends in conjunction with color-coded congestion levels. For dual-link GRIPs, another two design factors are considered: (1) position of the destinations (on the left side or in the middle of the two links) and (2) position of the route shields (at the top or bottom of the links). Examples of GRIPs are shown in Figure 11. Figure 11a-d are single-link GRIPs, presenting color-coded congestion levels as well as travel time through the current route. Figure 11e-h are dual-link GRIPs, presenting color-coded congestion levels and travel time through two different routes.



a b c d

e f g h

Figure 11. Alternative Designs of GRIPs

After determining the design factors for each category of messages, the Telegra traffic control software made available by Caltrans District 4 office was used to generate the graphical images for the travel time messages and up to six lines messages. For the other message categories, some design elements were not available in the database of the Telegra software, for instance, the proposed new symbols for special messages, the transit agency logos for transit travel time messages, and the GRIPs. These signs could only be edited by other means. Adobe Photoshop cc was used to generate those graphical images. All graphical images were edited as PNG files with a resolution of 3600 x 4500 pixels, which matched the proportional relationship between the width and the height of the IDB.

Section 3: Methodology of the Lab Testing

Experimental Design

Several factors were considered in the experimental design. The main factor was the design factors for each message category described in Table 1. Blocking factors[5] were also considered for the lab testing, including participants' gender (male and female), age (elder driver group and non-elder driver group), and native language (English and Spanish), which have been suggested to potentially account for comprehension variability of traffic signs (Shinar, 2002). A minimum of six participants was expected for each participant group.

[5] In the statistical theory of the design of experiments, blocking is the arranging of experimental units in groups (blocks) that are similar to one another, https://en.wikipedia.org/wiki/Blocking_(statistics)

Testing Procedures

The lab testing was conducted in a quiet conference room at the Richmond Field Station of the University of California at Berkeley. Upon arrival at the Richmond Field Station, participants were given a brief introduction of the study, including the purpose of the study and what they needed to do throughout the testing procedure. After the introduction, participants were asked to read and sign an informed consent form as well as a demographic information form including information about age, gender, education level, times of driving along the I-80 corridor, et al. Then they were seated at a table, facing a computer monitor which was about 8 feet away standing at the other end of the table, as shown in Figure 12. The computer monitor was an ASUS PB328Q model with 32 inches WQHD resolution display.

Participants viewed signs displayed on an HTML webpage, one sign at a time. Participants had complete control over how long to view each sign. When feeling confident of understanding the meaning of the message, participants used a mouse to click the sign. Then the sign was hidden. The HTML webpage recorded the start and the end viewing times for each sign. After that, experimenters asked questions about the specific sign in order to collect participants' comprehension and recommendation. The questions included:

- What information does the sign provide? Please list some details that you saw.
- Is it easy or difficult for you to understand the sign? On a scale of 1 to 5, 1 is very hard, and 5 is very easy?
- (For displays with symbols) Did you see the symbol on the sign? Do you understand the meaning of the symbol? How helpful is the symbol for your understanding of the sign? On a scale of 1 to 5, 1 is not helpful at all, and 5 is very helpful?
- What recommendations do you have in order to improve this sign?

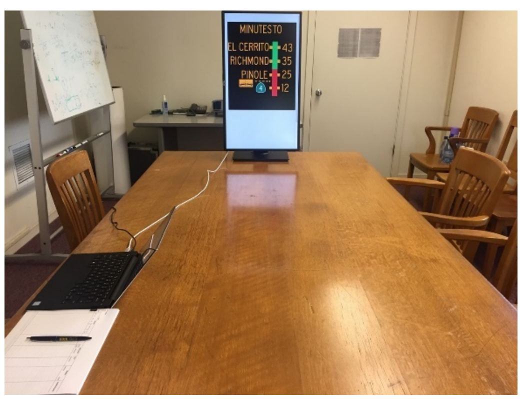


Figure 12. Lab Testing: Experiment Set-up

In order to manage the duration of the lab testing within 1.5 hours, 34 signs were selected with a number of messages from each category and a number of graphical designs of each design factor. After evaluating all 34 displays, participants were asked to select the most preferred design for each message. Twenty messages with a total of 55 signs were included, as shown in Appendix A.

Measurements

With the procedure described above, participants' understanding of each display was collected. More specifically, the measurements include:

- Comprehension accuracy: Whether each participant understood the message correctly?
- Easiness to understand: Is it easy or difficult to understand each message?
- Viewing time: How much time did it take each participant to process the message?
- Helpfulness of symbols: Whether each participant thought symbols on the display helped them to understand the message?
- Recommendation: What design elements caused drivers' confusion, and how did each participant suggest to improve?
- Comparison and preference: please select one of your most preferred design for each message.

Participants

In total, 52 participants took part in the lab testing, including 27 native English speakers and 25 native Spanish speakers. The number of non-elder drivers and elderly drivers is 28 and 24, respectively. The number of male and female drivers are 27 and 25, respectively. The number of participants from sub-groups is shown in Table 2, with a minimum of 6 participants in each. During the process of recruiting elder female native Spanish speakers, it was found that female native Spanish speakers who live along the I-80 corridor seldom drive after 65 years old. They most likely travel dependently, either relying on their family members to drive them around or use other transportation modes (e.g., shuttle bus service). Considering this fact, we lowered the threshold of elder female native Spanish speakers from 65 to 61. As a result, the non-elder female native Spanish speakers group age from 18 to 60 years old. The elder female native Spanish speakers group age 61 years old and above.

Table 2. Lab Testing: Number of Participants from Each Demographic Group

Language, gender, age			Number of participants in each sub-group	Sub-total	
Native English speakers	Male	Non-elder (18-64 years old)	9	15	27
		Elder (65 years old and above)	6		
	Female	Non-elder (18-64 years old)	6	12	
		Elder (65 years old and above)	6		
Native Spanish speakers	Male	Non-elder (18-64 years old)	6	12	25
		Elder (65 years old and above)	6		
	Female	Non-elder (18-60 years old)	7	13	
		Elder (61 years old and above)	6		

All participants hold valid California driver licenses. As shown in Table 3, all participants have high school and above education. Years of driving experience and times of driving on the I-80 corridor in each month indicate participants' driving experience is evenly distributed, and most of them travel more than 5 times each month along the I-80 corridor.

Table 3. Lab Testing: Participants' Demographic Information

Education level	**Percentage**	**Driving experience (years)**	**Percentage**	**Times of driving on I-80 corridor (per month)**	**Percentage**
Middle school or under	0	less than 10	3 (6%)	less than 5	21 (40%)
High school	15 (29%)	10 to 30	15 (29%)	6 to 15	9 (17%)
College graduate	23 (44%)	31 to 50	27 (52%)	16 to 30	14 (27%)
Post graduate	14 (27%)	51 and above	7 (13%)	31 and above	8 (15%)

Section 4: Results

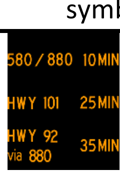
In this section, the understandability of each message category is analyzed, including accuracy, viewing time, perceived easiness, and symbol helpfulness. Some participants have misunderstood some specific design elements (i.e., certain wording, symbol, or abbreviations), which contributed to the variance of overall understanding accuracy, viewing time, and perceived easiness. After analysis of understandability, comprehension of specific elements that caused confusion and improvement recommendations are summarized. Finally, new designs for each problematic sign are proposed.

1. Travel Time Messages

Understandability

Table 4 shows the results of understandability for the travel time messages, including accuracy, viewing time, perceived easiness, as well as perceived helpfulness of the symbols, if there is any. For the highway travel time message, 77.3% of participants understood the display with highway symbols, compared with 82.0% for the display without symbols. Paired t-test was used to analyze the statistical difference of viewing time and perceived easiness between the display with a symbol and the display without a symbol. However, no significant difference was found. Most participants perceived that the symbols were helpful, with an average of 4.82 for the highway symbols and 4.74 for the airport symbol.

Table 4. Understandability of Travel Time Messages

Graphical Images		Accuracy		Time (Second)		Easiness		Symbol helpfulness
With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	
		77.3%	82.0%	15.64 (6.69)	14.22 (6.92)	3.26 (3.92)	3.30 (3.44)	4.82 (0.49)

 	82.0%	83.0%	13.79 (6.39)	13.22 (5.82)	3.53 (3.38)	3.43 (3.07)	4.74 (0.59)

Comprehension of specific elements

Some specific elements on the display to participants' misunderstanding of the message. For the above travel time messages, four design elements caused confusion to the participants, which are summarized as follows.

– Symbol of 580/880 

vs. 

Thirteen participants misunderstood this part of the message as from I-580 to I-880. One example of participants verbatim was *"drives from 580 to 880. But does not tell where the destination is."* Comparing this line of information between the display with symbols and the display with pure text, the "/" on the pure-text display is missing on the with-symbol display. It could be the reason that lowered the understandability of the display with symbols. However, this confusion could result from a lack of driving context information in the lab testing. So, it needs to be further tested in the next phase of the project, which is the simulator testing.

– 92 via 880 

There were 9 participants who commented that this element was confusing. They either misunderstood it as from 92 to 880 or didn't know what it meant. The word "via" was hard to understand for those participants.

– Airplane symbol 

The airplane symbol was on top of the three lines of travel time information. Ten participants misunderstood this sign as it went to one airport only. One participant's verbatim is *"from 880 to SFO".* When this participant was further asked how many airports? He replied as "*Only one, SFO.*" Another participant misunderstood it as "from 3 places to the airport, for example from San Francisco to airport".

– SJC via 880 

Sixteen participants didn't understand this line of the message. After viewing the sign, they right away asked what did "SJC" mean. Most of them tried to guess what did "C" stand for. One example of verbatim is "*does SJC means San Jose city?".* Also, five participants misunderstood this line as going to all airports via I-880. One example of verbatim was "*it is from 880 to SFO and OAK*.

Recommendations

– Show I-580 and I-880 on separate lines in order to make the sign as simple as possible.

– Avoid using "via" (92 via 880, SJC via 880).

– Airports & airplane symbol: put the text of "AIRPORTS" on the top of the message and put the airplane symbol in front of each line.

– SJC: put SJC only (without via 880), or remove this line because San Jose airport is far away. Hence this piece of information is less likely needed for people who travel on the I-80 corridor.

Proposed new designs

Based on the above recommendations, proposed new designs for the two travel time messages have been created and are shown in Figure 13.

Figure 13. Proposed Design for the Two Travel Time Messages

2. Transit Agency Logos for Transit Travel Time Messages

Understandability

As shown in Table 5, 83.3% of participants understood the AC transit travel time message with the AC transit agency logo, compared with 72.9% without the transit logo. Paired t-test was used to analyze the statistical difference of viewing time and perceived easiness between the display with a symbol and the display without a symbol. The perceived easiness for the with-symbol display (mean=3.32, STD=3.27) was significantly higher than the perceived easiness of the without-symbol display (mean=2.78, STD=3.82).

For the BART travel time message, 91.8% of participants understood the display with the BART logo, compared with 90.5% without the BART logo. The viewing time for the with-logo display (mean=14.12, STD=4.88) was significantly shorter than the viewing time (mean=18.98, STD=9.66) for the without-logo display.

Most participants perceived the symbols as helpful, with an average of 4.58 for the AC transit travel time message, 4.65 for the BART travel time message, and 4.40 for the WESTCAT travel time message.

Table 5. Understandability of Transit Agency Logos for Transit Travel Time Messages

Graphical Images		Accuracy		Time (Second)		Easiness		Symbol helpfulness
With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	
		83.3%	72.9%	17.76 (6.79)	18.11 (7.18)	3.32* (3.27)	2.78* (3.82)	4.58 0.84

		91.8%	90.5%	14.12* (4.88)	18.98* (9.66)	3.70 (2.36)	3.56 (2.63)	4.65 (0.86)
		89.1%	87.8%	14.50 (4.89)	14.67 (6.18)	3.58 (2.39)	3.22 (2.90)	4.40 (1.03)

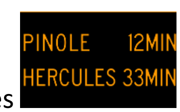
** Indicate significant difference, p-value < 0.05*

Comprehension of specific elements

For the transit agency logos of transit travel time messages, five design elements caused confusions to the participants, which are summarized as follows:

– The first phase of the messages 

Twelve participants didn't understand whether these messages were about public transportation or about driving. One participant misunderstood the BART travel time message as "driving from BART station to the other two places." One reason for this confusion could be that participants were not familiar with the public transportation agencies. Twenty-four participants mentioned that they didn't know WESTCAT, and some of them also mentioned they didn't know AC transit. One other reason could be that the message didn't have a clear format or wording to indicate travel time from which station to which station. Four participants mentioned that they did not understand where the starting point was.

– The second phase of the messages 

For the second phase of the message, 5 participants mentioned that they were not sure whether the number was time for the bus to arrive at the station or travel time to certain destinations. It was not clearly stated for those participants.

– Impact of the horizontal line

Four participants commented that the horizontal line made the two phases look unrelated.

– Abbreviation of DN (del Norte)

For 20 participants, the meaning of DN was not clear. They could not interpret it as del Norte and suggested that it should be spelled out.

– Overall messages

Overall, ten participants commented that these transit travel time messages were lacking information for decision making. Further information was needed, such as parking availability in bus or BART stations, how to go from bus or BART station to their final destinations, travel time when compared with driving. A few participants strongly recommended not to put these signs on the freeway as they could be a distraction to drivers.

Recommendations

- Put the bus symbol to indicate bus services (AC Transit and WESTCAT)
- Revise wording into "**From** XX station **to** XX."
- Spell out DN as del Norte.
- Remove the horizontal line
- Further, decide upon the needs of putting transit travel time messages on highway signs.

Proposed new designs

Based on the above recommendations, proposed new designs for each transit travel time message are shown in Figure 14.

Figure 14. Proposed Design for Transit Agency Logos for Transit Travel Time Messages

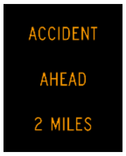
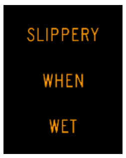
3. Special Messages

Understandability

As shown in Table 6, the accuracy for the special messages of ACCIDENT AHEAD, ROADWORK, SLIPPERY was 100% for displays both with symbols and without symbols. It indicates that these displays are fairly simple and easy for everybody to understand. Paired t-test was used to analyze the statistical difference of viewing time and perceived easiness between the displays with a symbol and the displays without a symbol. For the ACCIDENT AHEAD message, the viewing time of the with-symbol display (mean=7.06, STD=2.99) was significantly longer than the viewing time of the without-symbol display (mean=5.88, STD=2.61), which could be caused by the pink color of the symbol. For the ROADWORK message (mean=4.64 vs. mean=4.26) and the SLIPPERY special message (mean=4.72 vs. mean=4.03), the perceived easiness of the with-symbol display was significantly higher than the perceived easiness of the without-symbol display.

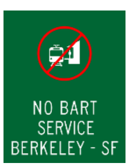
Table 6. Understandability of Special Messages (1)

Graphical Images		Accuracy		Time (Second)		Easiness		Symbol helpfulness
With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	With symbol	Without symbol	

		100.0%	100.0%	7.06* (2.99)	5.88* (2.61)	4.51 (1.02)	4.59 (0.81)	4.37 (1.13)
		100.0%	100.0%	7.50 (2.36)	7.49 (2.76)	4.64* (0.66)	4.26* (1.08)	4.79 (0.50)
		100.0%	100.0%	5.33 (1.69)	5.74 (2.50)	4.72* (0.54)	4.03* (1.74)	4.93 (0.25)

As shown in Table 7, for the special messages about the abnormal status of transit services, the displays with large symbols had higher understanding accuracy comparing with the displays with a small symbol. The large-symbol displays needed significantly shorter viewing time for both the DELAYED BUS message (mean=8.18 vs. mean=9.89) and the DELAYED CAPITOL CORRIDOR message (mean=9.36 vs. mean=11.65).

Table 7. Understandability of Special Messages (2)

Graphical Images		Accuracy		Time		Easiness		Symbol helpfulness	
Large symbol	Small symbol	Large symbol	Small symbol	Large symbol	Small symbol	Large symbol	Small symbol	Large symbol	Small symbol
		100.0%	97.8%	9.26 (3.56)	9.58 (3.98)	4.36 (0.97)	4.31 (1.39)	4.31 (1.12)	4.29 (1.28)
		100.0%	95.8%	8.18* (3.15)	9.89* (4.55)	4.48 (1.01)	4.22 (1.61)	4.41 (1.28)	4.26 (1.03)
		91.5%	87.8%	9.36* (4.17)	11.65* (5.74)	3.98 (2.30)	3.89 (2.15)	4.39 (0.93)	4.11 (1.06)

** Indicate significant difference, p-value < 0.05*

Comprehension of specific elements

For the special messages, four design elements caused confusions to the participants, which are summarized as follows:

– The pink color of the accident symbol 

Twenty-one participants commented on the pink color of the symbol for the message of "ACCIDENT AHEAD". Exemplar verbatim was "not seen pink color before", "why pink?", "makes me wonder" "pink means relax". This could be one reason why this display needed longer time to understand.

– The background color of "No service or delayed service"

Fifteen participants expressed concern regarding the green background color for the special messages of no service or delayed service, which was perceived as confusing. Exemplar verbatim was "green means go" "green is OK for information" "color green means good rather than a problem".

– BART & Capitol Corridor symbols  

Sixteen participants commented that the symbol for BART and symbol for the Capitol Corridor looked similar. Participants were not sure it was a BART or a train according to the symbol. Exemplar verbatim was "symbol of BART looks similar as AMTRAK" and "I don't know Capitol Corridor, thinks it's a BART". Another confusion was that the symbol with a person waiting on the platform was easily interpreted as "service is on time", which was mentioned by 5 participants. Exemplar verbatim was "the symbol with a person waiting is kind of confusing, telling me there is service".

– No estimated delay time of bus/train

Sixteen participants commented that there was no estimated delay time for the messages of delayed services for bus or train. One exemplar verbatim was "without how long of a delay, still don't know what to do".

Recommendations

– For the ACCIDENT AHEAD message, change the color of the symbol from pink to yellow, which is the standard cautionary color.

– Change the green background of "NO SERVICE" or "DELAYED SERVICE" to other cautionary colors in order to draw attention.

– For BART service-related messages, use the BART logo rather than the light-rail symbol.

– For the Capitol Corridor, use its own logo (with embedded train symbol on logo).

– Add estimated delay time for bus/train if possible.

Proposed new designs

Based on the above recommendations, proposed new designs for the four special messages (ACCIDENT AHEAD, NO BART SERVICE, DELAYED BUS SERVICE, and DELAYED CAPITOL CORRIDOR SERVICE) are shown in Figure 15.

Figure 15. Proposed Design for the Four Special Messages

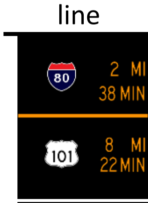
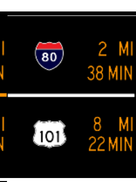
4. Up to Six Lines Messages

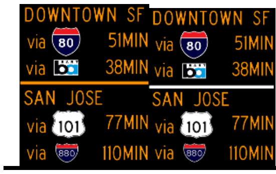
Understandability

Results of understandability for the two pairs of up to six lines messages are shown in Table 8. For the four-line message, the accuracy of the white-line display was much lower than the yellow-line one. One reason could be that the viewing sequence of the yellow-line display and the white-line display was unbalanced, which meant that for most of the participants, the white-line display was viewed before the yellow-line display. Hence the yellow-line display was more likely to be understood because of the learning effect. The unbalanced viewing sequence resulted from the pseudo-random number generator of HTML. On the other hand, the content of this message was not well thought of, where revising I-80 into I-580 would make more sense.

For the six-line message (travel time to DOWNTOWN SF and SAN JOSE), the perceived easiness for the display with a white line (mean=3.74, STD= 3.06) was significantly higher than the display with yellow line (mean=3.26, STD=3.76).

Table 8. Understandability of Up to Six Lines Messages

Graphical Images		Accuracy		Time (Second)		Easiness		Symbol helpfulness	
Yellow line	White line	Yellow line	White line	Yellow line	White line	Yellow line	White line	Yellow line	White line
		82.2%	65.2%	14.08 (4.24)	14.15 (6.74)	3.07 (3.01)	2.55 (4.25)	4.54 (0.69)	4.74 (0.62)

	79.2%	87.5%	19.42 (9.54)	16.54 (5.61)	3.26* (3.76)	3.74* (3.06)	4.70 (0.69)	4.69 (0.63)

** Indicate significant difference, p-value < 0.05*

Comprehension of specific elements

For the up to six lines messages, three design elements caused confusions to the participants, which are summarized as follows:

– MI as Miles & MIN as Minutes 

Sixteen participants thought both the two numbers were minutes. Hence it made them question why there were two different numbers. Exemplar verbatim was "I don't understand. I think there are two different times" "not sure why there are two-time slots"

– Consistency between the two phases of the messages

For the message of travel time to DOWNTOWN SF, it showed both travel time by driving and by BART. Five participants said that they were expecting a similar set of information for San Jose. Exemplar verbatim was "I was confused San Francisco has BART information, but San Jose has no BART information" "not consistent, information for San Francisco is different with information for San Jose, one has BART, the other one does not".

– Comparing with dual-link GRIP

Compared with the display of dual-link GRIP, the six-line message with two destinations was well understood. Participants commented that this was a better way to show alliterative routes or transportation modes when compared with the format of GRIP.

Recommendations

– For the four-line message, spell the MI out and avoid using the abbreviation;
    - OR remove MI;
    - OR add more space between miles and minutes;
    - OR change the layout of miles and minutes and display the two numbers horizontally;
– For the six-line message, put the two phases of the message in the same format in order to make it as easy to understand as possible.

Proposed new designs

Based on the above recommendations, proposed new designs for the four-line message and six-line message are shown in Figure 16.

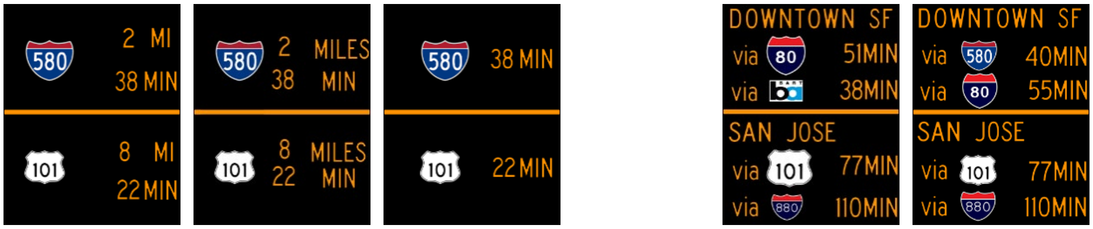


Figure 16. Proposed Design for Up to Six Lines Messages

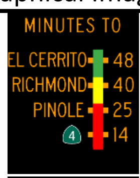
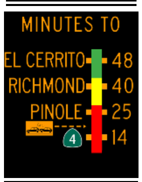
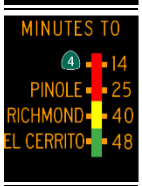
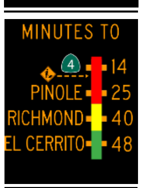
5. GRIPs

Understandability

The results of understandability for single-link GRIP are shown in Table 9. The accuracy for displays with legends (91.8% and 90.9%) was higher than accuracy for displays without legend (87.5% and 83.0%). The viewing time for every display was higher than 15 seconds.

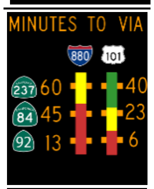
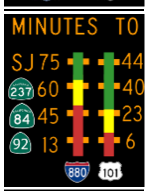
ANOVA was used to analyze the difference in viewing time and perceived easiness among the four displays. The GRIP presentation was perceived as more helpful when there was a crash legend (4.31 vs. 3.97) or roadwork legend (4.36 vs. 4.09). No significant difference was found.

Table 9. Understandability of Single-link GRIPs

Graphical Images	Accuracy	Time (Second)	Easiness	Symbol helpfulness
	87.5%	17.84 (10.18)	3.26 (2.68)	3.97 (1.40)
	91.8%	17.05 (6.14)	3.67 (2.59)	GRIP: 4.31 (1.18) Crash: 4.33 (1.19)
	83.0%	15.88 (5.10)	3.38 (3.33)	4.09 (1.39)
	90.9%	17.30 (7.25)	3.84 (2.23)	GRIP: 4.36 (1.22) Roadwork: 4.52 (0.75)

Results of understandability of dual-link GRIPs are shown in Table 10. For every display, the accuracy was below 90%, the viewing time was above 23 seconds, the perceived easiness was lower than 3, and the symbol helpfulness was much lower when compared with other message categories. No significant difference was found among the four displays regarding viewing time and perceived easiness.

Table 10. Understandability of Dual-link GRIPs

Graphical Images	Accuracy	Time (Second)	Easiness	Symbol helpfulness
	67.4%	23.59 (13.30)	2.29 (3.76)	3.85 (1.37)
	72.9%	24.07 (10.81)	2.55 (4.03)	4.03 (1.35)
	81.3%	23.29 (7.93)	2.49 (3.18)	3.77 (1.31)
	75.6%	27.71 (12.80)	2.79 (3.47)	3.90 (1.45)

Comprehension of specific elements

Three design elements caused difficulties for participants to understand the GRIPs, which are summarized as follows:

– Units of information

Many participants commented that there was too much information shown on one display. Especially for the dual-link GRIP, 21 users thought it was too complicated and needed time to learn and try to understand it. Exemplar verbatim was “too much information, probably will slow down the traffic” “too much information, I have to study it” and “not enough time to take all the information while driving”.

– Color of GRIP

Fifteen participants didn't understand the meaning of the color on the GRIP. Misunderstandings of the color included “I don't know what it (the color) means, red looks like a warning”, “red is to stop, yellow is about caution, green is to go”. Combined with the orientation of the display, the color coding made participants confused and led them to think logically it was not right that red-colored destinations took fewer minutes and green colored destinations took more minutes.

However, the legend of roadwork and accident helped some participants to figure out what the color meant, which also explained the higher understanding accuracy for displays of single-link GRIP with a legend. An example of verbatim was “It (legend of an accident) is helpful. It explains why it is red color.”

– Learning effect

For fourteen participants, they felt it was easier to understand the GRIPs after viewing a few of them. Exemplar verbatim was “it is easier as have seen for several times” “getting easier and easier now”.

Recommendations

- Only display single-link GRIPs with less number of destinations (2 or 3). Give time for the public to learn the format of GRIP; gradually displaying more destinations or displaying dual-link GRIP.
- Educate the public about the meanings of the three colors on GRIP;
- Add crash or roadwork legends if applicable;
- Orientation: using the arrow to indicate the orientation;
- It is sufficient to show color only or minutes only;

Proposed new designs

Based on the above recommendations, proposed new designs for the single-line GRIP and dual-link GRIP are shown in Figure 17.

Figure 17. Proposed Design for GRIPs

6. Comparison and Preference

After completing the evaluation of the 34 displays, participants were asked to select one of the most preferred design for each message. The results are shown in Table 11, with the most preferred display for each message, ID of the display (correspondent with Appendix A), as well as the percentage of being selected by all participants.

- For travel time messages and some special messages (i.e., ACCIDENT AHEAD, ROADWORK, SLIPPERY) the display designs with symbols were preferred by most participants (more than 90%).
- For the transit agency logos, the original colors were preferred by most participants.
- For the public transportation special messages (i.e., NO BART, DELAYED BUS, DELAYED CAPITOL CORRIDOR), the display designs with a more meaningful symbol (IDB 35) as well as large symbols (IDB 36 & IDB 39) were preferred.
- For the up to six lines messages, the white horizontal line was slightly more preferred than the yellow horizontal line, especially for IDB 47, which has completely separated two phases of the information.
- For the single-link GRIP, 60.0% of the participants preferred the design with top-bottom orientation and with the legend, comparing with the other three designs.

– Similarly, for the dual-link GRIP, 38.0% of the participants preferred the design with destinations on the left and route shields on the top.

Table 11. Most Preferred Design and Percentage of Selection for Each Message

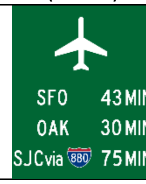
IDB 01 (98.0%)

IDB 03 (98.0%)

IDB 05 (96.1%)

IDB 07 (94.1%)

IDB 11 (96.1%)

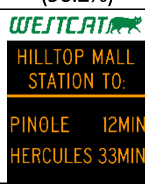


IDB 13 (100.0%)

IDB 16 (86.3%)

IDB 20 (92.2%)

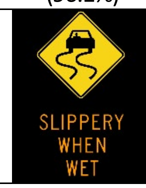
IDB 24 (90.2%)



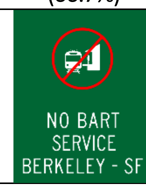
IDB 27 (86.3%)

IDB 29 (96.1%)

IDB 31 (96.1%)



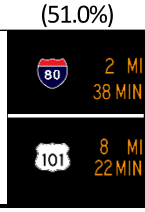
IDB 35 (66.7%)



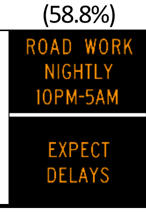
IDB 36 (56.9%)

IDB 39 (60.8%)

IDB 43 (51.0%)



IDB 45 (58.8%)



IDB 47 (60.8%)

IDB 51 (60.0%)

IDB 53 (38.0%)

As a summary, the following conclusions are derived regarding users' preferences.

– Signs with symbols are always preferred comparing with pure-text signs.

– Large symbols are preferred comparing with small symbols.

– The original color combination of transit agency logos is preferred.

– The white horizontal line is slightly more preferred than the yellow horizontal line, especially for signs with two independent phases of information.

– GRIP with top-to-bottom orientation is preferred comparing with bottom-to-top orientation.

– GRIP with events (e.g., accident, roadwork) legends is preferred.

– Dual-link GRIP with destinations on the left and highway routes on the top is preferred.

7. Effects of Language, Age, Gender, and Education

The effects of language, age, gender, and education were analyzed using ANOVA. The language effect on each participant's understanding accuracy, average viewing time for all displays, as well as the average rating of easiness for all displays are shown in Table 12. There was significant difference in the understanding accuracy between the native English speakers (Mean=90.95%, STD=13.0%) and the native

Spanish speakers (Mean=83.4%, STD=13.0%). There was no significant difference in viewing time and rating of easiness between the native English participants and the native Spanish participants.

Table 12. Lab Testing: Language Effect

Language	Understanding Accuracy		Average viewing time (Second)		Average rating of easiness	
	Mean	STD	Mean	STD	Mean	STD
Native English speakers	90.95%*	13.0%	14.51	4.63	3.56	0.87
Native Spanish speakers	83.40%*	13.0%	14.67	4.47	3.58	0.58

** Indicate significant difference, p-value < 0.05*

Results of the age effect are shown in Table 13. There was no significant difference regarding understanding accuracy, average viewing time, or average rating of easiness between non-elder and elder participants.

Table 13. Lab Testing: Age Effect

Age	Understanding Accuracy		Average viewing time (Second)		Average rating of easiness	
	Mean	STD	Mean	STD	Mean	STD
Non-elder drivers	89.84%	12.17%	14.58	4.90	3.65	0.75
Elder drivers	84.34%	14.43%	14.59	4.11	3.48	0.73

Results of the gender effect are shown in Table 14. There was no significant difference regarding understanding accuracy, average viewing time, or average rating of easiness between female and male participants.

Table 14. Lab Testing: Gender Effect

Gender	Understanding Accuracy		Average viewing time (Second)		Average rating of easiness	
	Mean	STD	Mean	STD	Mean	STD
Female drivers	85.2%	15.9%	15.53	5.16	3.61	0.76
Male drivers	89.2%	10.4%	13.71	3.69	3.53	0.72

The effect of education level on each participant's understanding accuracy, average viewing time for all displays, as well as the average rating of easiness for all displays are shown in Table 15. Education had a significant effect on understanding accuracy. Post-hoc analysis (Turkey) was used to analyze the difference between each two education levels. It was found that there was a significant difference in understanding accuracy between participants with high school education (Mean=79.7%, STD=17.6%) and participants with post-graduate education (Mean=95.8%, STD=5.2%). There was no significant difference in viewing time or rating of easiness among participants with different education levels.

Table 15. Lab Testing: Effect of Education

Education	Understanding Accuracy		Average viewing time (Second)		Average rating of easiness	
	Mean	STD	Mean	STD	Mean	STD
Highschool	79.7%*	17.6%	16.69	4.69	3.43	0.77
College	87.3%	10.8%	13.76	4.66	3.49	0.76
Post graduate	95.8%*	5.2%	13.67	3.48	3.88	0.60

** Indicate significant difference, p-value < 0.05*

Section 5: Discussion

Traffic sign design should be guided by established ergonomics principles to enhance comprehension (Ben-Bassat, 2006). Based on the analysis, we come out with the following principles for IDB design, including (1) use of symbols, (2) consistency, (3) simplicity, and (4) provision of sufficient information for decision making. The above four principles are further discussed in this section.

Use of Symbols

According to the results of understandability for travel time messages, transit travel time messages, and special messages, most displays with both symbols and texts have higher understanding accuracy, require shorter viewing time, and have higher perceived easiness comparing with their pure-text counterparts. From the results of the comparison and preference, the most preferred displays are always the ones with symbols. This confirms the benefits of displaying both symbols and texts on IDBs.

About symbol size, large symbols require shorter viewing time compared with small symbols according to the statistical results of the public transportation related special messages. For the consideration of alternative designs for the public transportation related special messages, one intention was to reflect the abnormal status (i.e., NO BART, DELAYED BUS, DELAYED CAPITOL CORRIDOR) on symbols. Therefore, additional elements such as NO SERVICE, DELAYED were added, and new symbols were designed. However, those additional elements may have made the symbols too complicated, which took participants longer time to comprehend. Large symbols with simple presentation help drivers to understand traffic signs quicker. Large symbols were also preferred in the comparison testing.

Consistency

For the special messages, many participants commented on the pink color of the ACCIDENT AHEAD symbol, which was unusual and consumed extra mental effort to process. For the public transportation related special messages, many participants expressed concerns about using green background color for abnormal public transportation service messages. This reflects the fact that the green color is usually used to indicate good working status rather than abnormal status. Colors such as yellow or amber should be used for warning and to draw special attention, which is actually well-aligned with the CA MUTCD[6].

---

[6] CA MUTCD Section 2L.04, Design Characteristics, colors of CMS.

A few participants also commented on the consistency problem with the six-line message, which was used to show travel options to downtown San Francisco and San Jose. Although the two phases of the message were independent of each other, participants still expected to see similar patterns of information across the two phases. With consistent patterns across the two phases, it would require less mental workload to process the information on the sign and also make it easier to remember the information after viewing the sign.

Simplicity

For the dual-link GRIPs, almost half of the participants commented that it was way too complicated to put so much information on one sign. Comparing the dual-link GRIP with the six-line message, the dual-link GRIP led to lower understanding accuracy. Some participants explicitly expressed the preference of the six-line message over dual-link GRIP.

One reason is that the format of the six-line message is simpler than the format of dual-link GRIP. Both of the two types of messages have different routes and multiple destinations. However, the dual-link GRIP has 3 or 4 destinations, compared with 2 destinations on the six-line message. Besides the travel time (minutes), it also adds the element of different colors, which almost one-third of the participants don't understand. The above effects add up and make the dual-link too hard to understand.

For some Spanish speakers, it was hard to understand the word "via". For example, the airport message, they were more inclined to misunderstand "SJC VIA 880" as going to all three airports via I-880. For the elder drivers, although there was no significant difference in their understanding comparing with the non-elder drivers, many of them conveyed that they would prefer to get the message at one glance. If it is too complicated, they just give up studying it. Hence, although the IDB is capable of providing more information on one sign, it should display simple messages in order to make it easy to understand for all drivers.

Sufficiency

Transit travel time messages and public transportation related special messages were designed to be displayed on the IDB in order to provide information on alternative transportation modes besides driving. Many participants had the question of why these messages were needed if they were driving on the highway. Besides, in order to make the decision of taking public transportation rather than driving, extra information such as parking availability, bus or BART arrival time and travel time, how to get from the transit station to their final destinations, would further be needed. As a result, presenting all the above necessary information on one sign led to a challenging question.

References

[1] Ben-Bassat, T., and Shinar, D. (2006), Ergonomic guidelines for traffic signs design increase signs comprehension. Human Factors, 48, 182-195.

[2] Charlton, S.G. (2006), Conspicuity, memorability, comprehension, and priming in road hazard warning signs, Accident Analysis and Prevention, 38 (3), 496-506.

[3] Dewar, G. (1988), Criteria for the design and evaluation of traffic sign symbols, Transportation Research Record, 1160, 1-6.

[4] Shinar, D., Dewar, R. E., Summala, H., and Zakowska, L. (2002), Traffic sign symbol comprehension: A cross-cultural study. Ergonomics, 46, 1549–1565.

[5] Ullman, B. R., Trout, N. D. and Dudek, C. L. (2009), Use of graphics and symbols on dynamic message signs: Technical report. 0-5256-1. Texas Transportation Institute, Texas A&M University System, College Station.

Appendix A: All Designs from the Lab Testing

**Indicate the 34 signs which were evaluated in the first part of the testing*

Part II: Simulator Testing

Summary

Based on the findings and proposed new designs for some signs in the lab testing, in this phase of the project, we further evaluated three categories of messages which were proposed to be displayed on the IDBs, including (1) up-to-six lines of information in a single phase to display travel-time messages; (2) transit travel-time messages; and (3) GRIPs presenting color-coded congestion levels as well as travel time. To evaluate the understandability of these three categories of messages and different graphical designs, a simulator study was carried out. Twenty-four local commuters of Intestate 80 Smart Corridor participated in the study. Results showed that: (1) perceived easiness for the five-line and six-line travel time message were significantly lower than the three-line travel time message; (2) perceived easiness and symbol helpfulness for the BART (Bay Area Rapid Transit) logo was significantly higher than the standard light-rail symbol; (3) for single-link GRIP, displays without legends and with the bottom-up orientation were perceived as easier to understand; (4) understanding accuracy of single-link GRIP with 3 destinations was significantly higher than single-link GRIP with 4 destinations; (5) single-link GRIPs were perceived as easier to understand in comparison with dual-link GRIPs; and (6) participants took longer time to perceive the dual-link GRIPs.

Section 1: Introduction

This part summarizes the work on a driving simulator study conducted in the second phase of the I-80 Information Display Board project, with which we were aiming to further evaluate the understandability of various messages and different graphical designs in the driving context. Specifically, the objectives of the simulator testing are to evaluate: (1) how easy it is for drivers to understand each display; (2) how helpful it is for drivers to make decisions regarding specific destinations; and (3) the effectiveness of different displays and specific design elements (e.g., symbols, words).

Compared with the lab testing, the advantages of a simulator testing are two-fold. On the one hand, the simulation program provides a realistic representation of the road geometry and traffic conditions where the signs are displayed. On the other hand, a simulator testing requires the study participants to spend both physical and mental efforts on the driving task while looking at the signs. The above two reasons allow the driving simulator testing to offer higher validity for investigating the understandability and effectiveness of the IDB signs.

There have been a number of previous studies that aimed to understand drivers' perception of road signs. Various methods have been applied in those researches, including eye movement methods, methods of using verbal reports while driving, methods of asking drivers to recall a sign after passing it, and methods of recording drivers' behavior. As analyzed by Martens (2000), all methods have their specific advantages and limitations. For the eye movement method, the main advantage is that under most circumstances, it is relatively free from bias despite the instructions. Eye fixations do not necessarily mean the object is actually perceived, but they serve as a good indicator to provide some insights regarding the complicated process of object perception. The main disadvantage with the eye movement method is that despite fixation on a certain object, attention can be directed to another location (Gao et al., 2006; Hills, 1980). The method of using verbal reports while driving can provide some insights into the internal process of the driver, which can complement information from eye movement studies by making drivers verbalize what attracts their attention (Charlton, 2006). However, this method is mostly used to evaluate the conspicuity of a sign rather than the understandability. In the method of recall after passing a sign, drivers are asked whether they perceive certain sign information. With this method, the experiment should be set up in a way that participants have a clear picture of what to report. In addition, the reports should be requested immediately after the trial in order to keep the interfering factor of "forgetting" to a minimum level (Charlton, 2006). However, some degree of memory decay is inevitable.

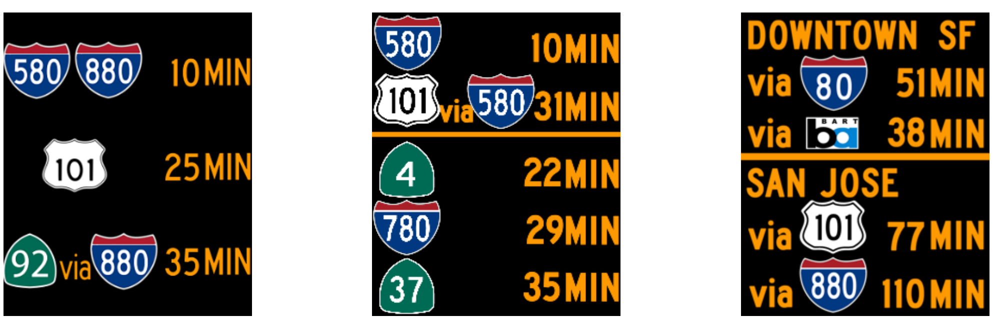
Fisher (1992) argued that the true measure of a sign's effectiveness was not recall, recognition, or naming, but the extent to which the sign content would affect drivers' preparedness for and subsequent responsiveness to events. A combination of several research methods was recommended. Hence, in this simulator testing, we used a combination of methods, including eye movements, recall after passing the sign, as well as the drivers' preparedness for arriving at the destination of each trip.

Section 2: Three Categories of Messages and Alternative Designs

Some of the displays from the lab testing were carried over into the simulator testing, such as the transit travel time messages. Some of the displays were well understood by most of the drivers in the lab testing, such as the special messages. Hence, they were not included in the simulator testing. For some other displays, new designs were proposed after the lab testing and included in the simulator testing. Additionally, one display was newly proposed by Caltrans. In total, a collection of 26 displays was evaluated in the simulator testing. The 26 displays were divided into three categories of messages. In this section, we describe all three categories of messages and explain the design factors considered for each category.

1. Travel Time and Up to Six Lines Messages

The travel time messages and up-to six lines messages have been evaluated in the lab testing. In the simulator testing, only one travel time message was included, as shown in Figure 18a. Two other up-to six lines messages were also included in the simulator testing. One was a newly proposed five-line message, as shown in Figure 18b, which had two phases of messages with one horizontal line in between. The other one was a six-line message, as shown in Figure 18c, which also had two phases with traveling information about one city in each phase. Caltrans would like to use the three-line travel time message as the baseline and compare the easiness and helpfulness of the other two messages.



a b c

Figure 18. Travel Time and Up to Six Lines Messages

2. Transit Travel Time Messages

Four types of transit travel time messages are evaluated in the simulator testing. These signs provide information about travel time from certain starting points to various destinations (1) by AC Transit, (2) by WESTCAT, (3) by BART, and (4) special messages about BART service.

AC Transit travel time message

The AC Transit travel time messages are shown in Figure 19. There are three alternative graphical designs. In the first design, it has the AC Transit logo on top of the display. After the logo, it has texts of the starting point and the destinations, with a horizontal line in between, as shown in Figure 19a. In the second design, there is an additional little bus symbol on the left side of the AC Transit logo, as shown in Figure 19b, which is a new design proposed after the lab testing. With the bus logo, it is intended to express that the message is about a bus service, hoping it would help drivers who are not familiar with AC Transit. In the third design, as shown in Figure 19c, there are only texts of "AC TRANSIT" without the logo.

a b c

Figure 19. Transit Travel Time Messages: AC Transit

WESTCAT travel time message

The WESTCAT travel time messages are shown in Figure 20. Similarly, there are also three alternative graphical designs. In the first design, it has the WESTCAT logo on top of the display, as shown in Figure 20a. In the second design, there is an additional little bus symbol on the left side of the WESTCAT logo, as shown in Figure 20b. The bus logo is intended to be helpful for drivers who are not familiar with WESTCAT. In the third design, as shown in Figure 20c, there are only texts of "WESTCAT" without the logo.

a b c

Figure 20. Transit Travel Time Messages: WESTCAT

BART travel time message

The BART travel time messages are shown in Figure 21. There are two alternative graphical designs. In the first design, it has the BART logo on top of the display. After the logo, it has texts of the starting point and two destinations, with a horizontal line in between, as shown in Figure 21a. In the other design, as shown in Figure 21b, there are only texts of "BART" without the logo.

a b

Figure 21. Transit Travel Time Messages: BART

BART service special message

The special messages of BART service are shown in Figure 22. The message is about abnormal BART service from Berkeley to San Francisco. There are also two alternative graphical designs. In the first design, it uses the light-rail symbol for BART, with texts blow the symbol expressing no service, as shown in Figure 22a. In the other design, it uses the BART logo, with similar texts below the logo, as shown in Figure 22b.

a. b.

Figure 22. Transit Travel Time Messages: No BART Service

3. Graphic Route Information Panels (GRIPs)

Both single-link GRIPs and dual-link GRIPs are included in the simulator testing.

Single-link GRIPs

Three design factors are considered for single-link GRIPs: (1) orientation of approaching destinations (top-bottom or bottom-top); (2) with or without the roadwork legend in conjunction with color-coded congestion levels; and (3) number of destinations (3 or 4 destinations). There are 7 single-link GRIPs in the simulator testing, as shown in Figure 23. Based on suggestions by the Caltrans D4 team, arrows are added to all GRIPs as an indicator of the orientation. Hence, all of the single-link GRIPs have the arrows, with either up-arrow (e.g., Figure 23a) or down-arrow (e.g., Figure 23b). The roadwork legend is used to explain the traffic conditions on some of the single-link GRIPs (e.g., Figure 23c, Figure 23d). As shown in Figure 23e and Figure 23f, the displays show travel time information of four destinations. A new design with only the traffic conditions but not the travel time is proposed after the lab testing, as shown in Figure 23g, which is also included in the simulator testing.

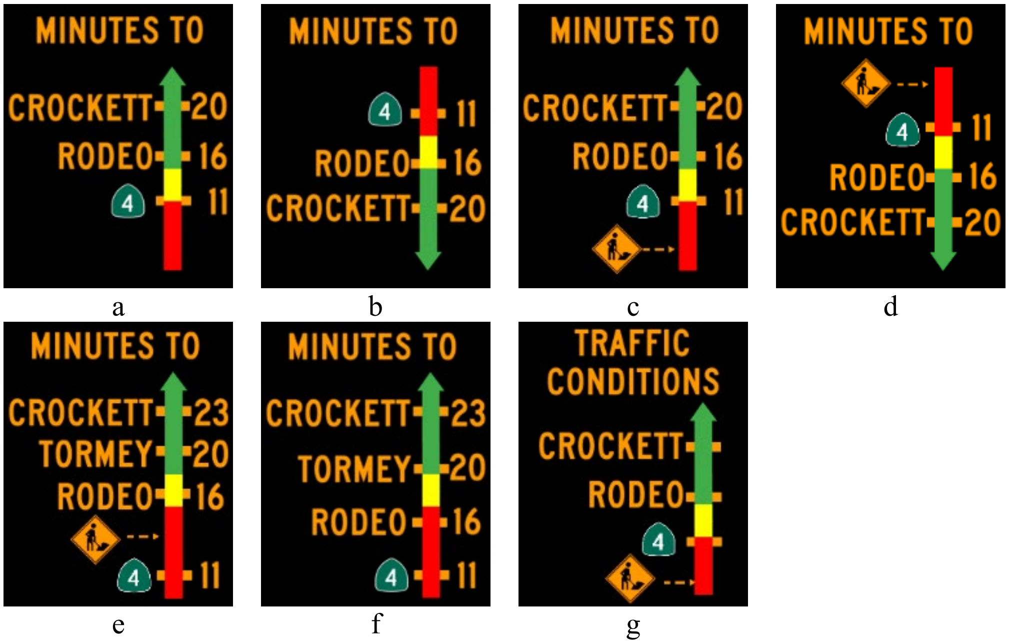


Figure 23. Single-link GRIPs

Dual-link GRIPs

Two design factors are considered for dual-link GRIPs: (1) position of the destinations (on the left side or in the middle of the two links) and (2) position of the route shields (at the top or bottom of the links). All arrows of the dual-link GRIPs are up-arrows. For example, Figure 24a has the destinations in the middle of the two links and has the route shields at the top. Figure 24d has the destinations on the left side and has the route shields at the bottom of the links. Similarly, as shown in Figure 24e and Figure 24f, the two displays only have traffic conditions without travel time information, which are proposed after the lab testing.

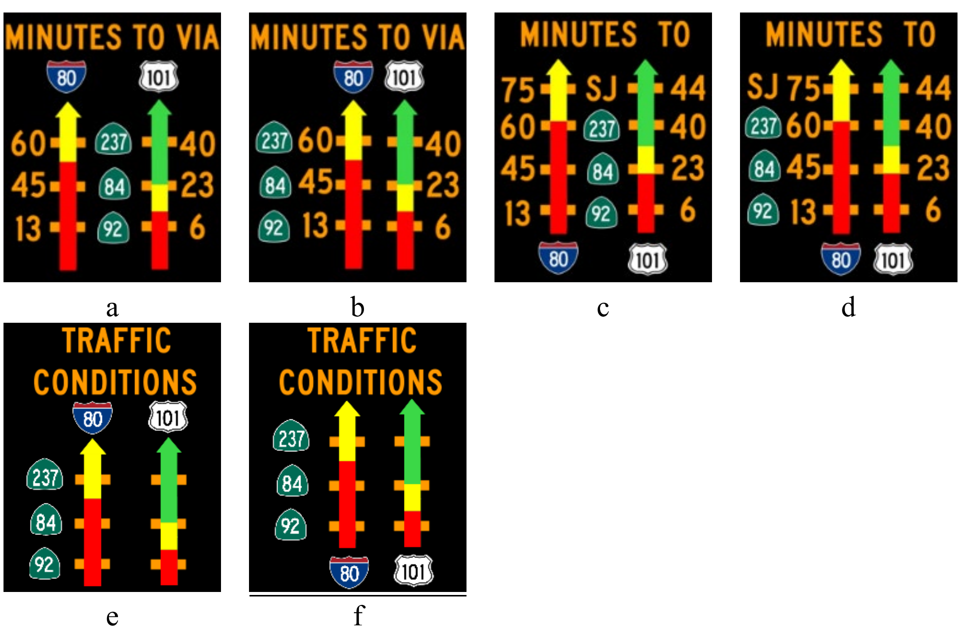


Figure 24. Dual-link GRIPs

As a summary, the design factors for each message category are shown in Table 16. For the travel time messages, the design factor is the number of lines of messages (i.e., three, five, and six). The goal is to compare the five-line and six-line messages with the three-line message, respectively. For the transit travel time messages, the goal is to compare different symbols (logos). For the GRIPs, the purpose is to explore the difference regarding orientations, legends, positions of destinations, and positions of route shields. It is also of interest to compare the difference between the single-link GRIPs and the dual-link GRIPs.

Table 16. Design Factors for Each Message Category in the Simulator Testing

Message Categories	Design Factors
– Travel time and up to six lines messages	• Number of lines of messages
– Transit travel time messages	• Transit logo vs. Text only • Different symbols
– GRIP	
• Single-link GRIP	• Orientation: top-bottom vs. bottom-top • With or without legend • Number of destinations
• Dual-link GRIP	• Positions of destinations (left vs. middle) • Positions of route shields (top vs. bottom)

Section 3: Methodology of the Simulator Testing

Experiment Set-up

The driving simulator is located in McLaughlin Hall at the main campus of the University of California, Berkeley. The model of the simulator is Force Dynamic 401CR, as shown in Figure 25a, mounted on a floor, allowing for continuous rotation, with triple-monitor display and surround sound. Participants can control both the steering wheel and the pedals (throttle and brake), which require them to make both physical and attentional efforts onto the driving tasks and make the simulation more representative of driving a car in reality.

a b

Figure 25. Simulator Testing: Experiment Set-up

A driving simulation program was developed to display the designed signs in real-world videos. The on-road video and vehicle motion data were collected using an instrumented car. A two-minute-long driving data were recorded for each trial. Data of the same location were recorded multiple times in order to test different signs. During data collection runs, the instrumented car was always driven on the far-right lane while approaching the signs so that the viewing angle for each sign would be the same. The driving speed was kept at about 60 mph for the travel time messages and the transit travel time messages. It was about 30 mph for GRIPs. Three Logitech C922X cameras were used to record the views of front, left-front, right-front viewing angles with a total of a 150-degree field of view. The resolution of the video was 1920*1080. The frequency was 30 frames per second. The vehicle motion data were recorded at 100Hz with a 3DM-GX4-45 IMU sensor.

Image processing was used to integrate the designed signs onto the recorded videos. Firstly, the edge of the IDB sign was detected in order to recognize the position of the sign inside of each video frame. Then a filtering process based on time series was conducted to maintain the consistency. Secondly, the designed sign image was mapped onto each frame using the projective transformation. To make the visual effect of the signs as realistic as that of seeing them on the road, two techniques were applied: (1) The pixel information from the background and the border of real-world signs was extracted and fused into the designed signs. (2) The video images were zoomed in at the ratio of 2:1. One exemplar frame of the processed image with the designed sign is shown in Figure 25b.

After the image processing, the videos were played on three screens of the simulator in the same timeline. The motion data was also played in the same timeline. A filter of motion data was added to adjust the vertical and horizontal motion to make the experience close to driving a car in reality. The speed for playing the video and the motion was controlled by the input driving speed from the participants using the simulator. A speedometer was displayed on the screen to indicate the current driving speed. A driving speed of 55-75 mph was recommended for videos displaying travel time messages and transit travel time messages. A speed of 25-45 mph was recommended for videos displaying GRIPs. In order to control the viewing time for each sign, a point from which people with normal vision can easily see the sign clearly was decided based on the results of empirical tests. From this point, the simulation program started to display the designed signs. Once the program started to display the designed sign, the speed of the simulator was maintained at a constant value so that the sign could be displayed in a fixed duration. For travel time messages and transit travel time messages, the time duration was 15 seconds. For GRIPs, the time duration was 20 seconds. In total, 27 trials were implemented, with one practice run and 26 trials for the 26 signs. The driving time for each trial was between 1 and 1.5 minutes.

Testing Procedures

The testing procedures for the human-subject simulator testing included the screening process, orientation, practice session, and actual driving experiment session. If the participants didn't take part in the earlier lab testing, they were asked to schedule a pre-test at least one week before the simulator testing. In the pre-test, they were shown eight signs, including two travel time messages, one transit agency logo on transit travel time message, two special messages, one six lines message, one single-link GRIP, and one dual-link GRIP. The experimenters asked the participants several questions for each sign to evaluate their comprehension and preference. The duration of the pre-test for each participant was approximately 20 minutes. During the pre-test, the experimenters also collected participants' demographic information, which included age, gender, education level, times of driving along the I-80 corridor, native language (English or Spanish), et al. After the pre-test, we went through a screening process to determine if the individuals could safely participate in the simulator testing. Individuals were asked to complete a screening form. To be qualified for participating in this study, each individual must hold a valid California driver's license, frequently travel on the I-80 corridor, and meet other physical requirements regarding their height, weight, history of bone-injury et al.

After the screening process, participants were asked to read and sign an informed consent form. Before getting onto the driving simulator, participants were given a safety orientation in order to let them know how to safely operate the simulator. After the participants were comfortably seated, the experimenters calibrated the eye tracker. Then a Bay Area map was used to explain the locations of the IDB signs and the cities in the Bay Area that they would use as destinations for each trip.

The next step was the practice session, the purpose of which was to let participants learn how to operate the simulator and get familiar with the evaluation procedure. Participants were asked to press the throttle and start playing the video. They were instructed to maintain the driving speed within a specific range (e.g., between 55 mph and 75 mph) and change lanes only when there was a green arrow shown on the screen, indicating changing lanes to the right or left. The video sequence of the practice session was about 2 minutes long. Right after viewing the sign, participants were asked the following questions:

- General information of the sign;
- Specific information associated with the destination, which varied for different categories of messages. For example, specific information for GRIP included orientation, the meaning of colors, meaning of the legend, highway options, and location of destinations;
- Symbol helpfulness for signs with either new proposed symbols or for comparison of different symbols, on a scale of 1 to 5.
- Easiness of each sign, on a scale of 1 to 5.

After the practice session, it was the main testing, in which participants evaluated all 26 signs. The evaluation procedure was the same as the practice session. On average, it took about 2 minutes to complete the evaluation of each sign. The viewing sequence for all signs was randomized for each participant. In order to avoid the deterioration in the accuracy of the eye tracker, a re-calibration procedure was implemented after the evaluation of every five signs. After every 30 minutes, participants were asked to get off the simulator and take a break. In total, the experiment took about 90 minutes for each participant.

Measurement

With the procedure described above, participants' eye movement and subjective evaluation of each display were collected. With the above data, we aimed to obtain the following measurements:

- Visual attention on each display and on different elements, such as how long participants looked at the entire display and what elements they were mostly focusing on.
- Easiness to understand: Is it easy or difficult to understand each message?
- Comprehension accuracy of general information of the sign: Did participants understand the general information each display trying to convey?
- Comprehension accuracy of detailed information: Did participants find the information related to the destination of each trip?

Participants

For the simulator testing, the main attributes to evaluate are the design factors described in Table 16. Similar to the lab testing, blocking factors[7] include participants' gender (male and female) and native language (English and Spanish), which have been suggested to potentially account for comprehension variability of traffic signs. Based on the results of the lab testing, age has no significant influence on the understandability. Hence, age is not considered in the simulator testing. A minimum of six participants is included for each participant group.

In total, twenty-four participants are recruited, including 12 native English speakers and 12 native Spanish speakers. The demographic information of all participants is shown in Table 17 and Table 18.

Table 17. Simulator testing: Demographic Information of Participants (1)

Item	Mean	STD
Age (years)	47.71	10.35
Driving Experience (years)	23.48	10.62
Times driving on I-80 corridor (per month)	24.31	18.36

Table 18. Simulator Testing: Demographic Information of Participants (2)

Items	Number	Percentage
Gender		
Male	12	50%
Female	12	50%
Language		
English	12	50%
Spanish	12	50%
Education level		
Middle school or under	0	0
High School	5	20.83%
College Graduate	13	54.17%
Post Graduate	6	25.00%

Participants' years of driving experience and times of driving on the I-80 Corridor per month are shown in Table 19, indicating that most participants are experienced drivers and daily commuters of the I-80 corridor.

Table 19. Simulator Testing: Driving Experience and Times Driving on I-80 per Month

Driving experience (years)	Number	Percentage	Times driving on I-80 per month	Number	Percentage
0-9	3	12.50%	0-9	7	29.17%
10-19	6	25.00%	10-19	1	4.16%
20-29	7	29.17%	20-29	7	29.17%

[7] In the statistical theory of the design of experiments, blocking is the arranging of experimental units in groups (blocks) that are similar to one another, https://en.wikipedia.org/wiki/Blocking_(statistics)

30-39	7	29.17%	30-39	1	4.16%
40-49	1	4.16%	40-49	6	25.00%
50+	0	0.00	50+	2	8.33%

Section 4: Results and Findings for Each Category of Messages

In this section, the perceived easiness, understanding accuracy of general information, understanding accuracy of detailed information for each sign are analyzed in order to determine the effects of different design factors. Participants' comments regarding some design factors (e.g., preference of orientation) and specific design elements (i.e., bus symbol) are summarized. The eye-tracking data are analyzed in order to see the total fixation time and distribution of visual attention onto specific design elements for each display. The effects of gender, native language, and education levels on the understanding of all displays are also analyzed.

1. Travel Time and Up to Six Lines Messages

Understandability

Table 20 shows the results of understandability for the travel time and up to six lines messages, which include mean and standard deviation (STD) for easiness and understanding accuracy of general information, as well as results for understanding accuracy about destinations.

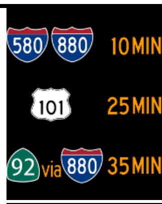
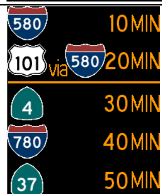
Using the three-line travel time message as the baseline, a paired t-test[8] was used to compare the statistical difference with the five-line message and the six-line message, respectively. For the five-line message, the perceived easiness (mean=4.04, STD=1.08) was significantly lower (p=0.001) than the perceived easiness of the three-line travel time message (mean=4.63, STD=0.71). For the six-line message, the perceived easiness (mean=4.22, STD=1.09) was also significantly lower (p=0.03) than the perceived easiness of the three-line travel time message. The understanding accuracy of the general information (mean=90.28%, STD=23.00%) was also significantly lower (p=0.02) than the understanding accuracy of the general information for the three-line travel time message (mean=100%, STD=0).

For the three-line message, 15 out of 24 participants (62.50%) correctly found the information of the destination (580 & 880 junctions, 10 minutes). For the five-line message, 12 out of 24 participants (50%) correctly found the information of the destination (101 via 580, 20 minutes). For the six-line message, 14 out of 24 participants (58.33%) found the information of the destination (via 101, 77 minutes).

Table 20. Understandability of Travel Time and Up To Six Lines Messages

Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about destinations

[8] A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample. http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf

1		4.63 (0.71) *	100.00% (0) *	15 (62.50%)
2		4.04 (1.08) *	100.00% (0)	12 (50.00%)
3		4.22 (1.09) *	90.28% (23.00%) *	14 (58.33%)

** Indicate significant difference, p-value < 0.05*

Comments on detailed design elements

After completing evaluating each sign, participants commented on the specific elements on the display. For the travel time and up to six lines messages, participants had several comments regarding the five-line message and the six-line message. The comments are highlighted and summarized as follows.

– Three-line message

There were not many comments for the three-line message. One participant misunderstood the first line from 580 to 880. Another comment was suggesting to use city names rather than highway shields on the sign.

– Five-line message

Six participants commented that there was too much information on one sign. Examples of participant's verbatim were "*way too much information on one sign*" and "*pick the information from a lot*". Because it displayed five lines of messages on one sign, two of the participants also commented on the spacing and layout "*101 via 580 are squeezed in*" and "*spacing between 101 and 580 is too crowded. It looks like 3 minutes (actually it is 20 minutes).*"

Four participants also commented on the horizontal line between the first and the second virtual phase. Examples of their verbatim were "*The first two options are easy… But I am not sure about the other three*", "*why the line is there*", "*I didn't get the (meaning) of the lower part (of the message)*" or "*need to have header to give you clue of what the sign is about*".

– Six-line message

Three participants commented that there was too much information on this display, such as "a lot of information".

Visual attention

The raw data from the eye tracker included the timestamp and position (x and y position) of each gaze point. Two participants' number of gaze points were significantly smaller than others, which might be caused by the deterioration in the accuracy of the eye tracker calibration. Both these two participants' visual attention data were excluded from the analysis.

The heatmap of the raw gaze-point data from the 22 remaining participants for each display is generated. The following gradient of colors is used: (1) red, (2) yellow, (3) green, (4) cyan, and (5) blue. These colors indicate attention from highest to lowest.

The heatmaps for the travel time and up to six lines messages are shown in Figure 26. For the three-line message, as shown in Figure 26a, participants focused their attention on each of the three lines. Besides, the word "via" on the third line drew the greatest attention in comparison with other elements on the display. Either the graphical design (e.g., smaller font) or the meaning of this word made the participants wonder. For the five-line message, as shown in Figure 26b, participants didn't pay even attention to all five lines. In contrast, they spent a lot of attention on the middle portion of the sign, where there was no much information. To a certain extent, this heatmap indicated that participants looked at the second line of the message, which was related to their destination. Afterward, they spent most of their attention trying to figure out the meaning of the information below the horizontal line. As shown in Figure 26c, the heatmap of the six-line message indicated that the words "San Jose" (destination of this trip) took a reasonable amount of attention. Other than San Jose, the word "via" took participants' special attention when compared with other elements on the display.

a b c

Figure 26. Heatmap of Visual Attention for Travel Time and Up To Six Lines Messages

The fixation duration time was detected from the raw gaze-point data and was calculated for each display for each participant. A paired sample t-test was used to compare the average fixation time between the three-line message and the five-line message, as well as the six-line message, respectively. The fixation time for the three-line message (mean=11.74, STD=2.17) was higher than the five-line message (mean=10.14, STD=3.31) with statistical significance (p=0.05). No significant difference was found between the three-line message and the six-line message (mean=11.22, STD=3.03). The reason why the five-line message took less fixation time could be that participants looked but could not find any meaningful information from the contents below the horizontal line on the five-line message. Hence, they didn't spend more attention to the details, which were reflected in the heatmap and consistent with participants' comments. On the other hand, it was worth noting that for the same display, there was a big variance of fixation time among participants. Therefore, the measurement and result of fixation duration time should be used as a reference only.

2. Transit Travel Time Messages

AC Transit travel time message

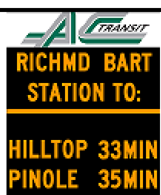
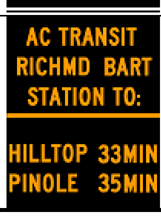
Understandability

Table 21 shows the results of understandability for each design of the AC Transit travel time message, which includes mean and standard deviation (STD) for easiness and understanding accuracy of general information, as well as results for understanding accuracy about destinations.

Single-factor ANOVA [9] was used to analyze the statistical difference of perceived easiness and understanding accuracy of general information for the three displays. The perceived easiness of the display with the AC Transit logo only (mean=4.17, STD=0.87) was higher than the other two displays (mean=3.98, STD=1.24; mean=4.06, STD=1.24). However, there was no significant difference. The understanding accuracy of the general information for the display with the AC Transit logo only (mean=90.63%, STD=14.39%) was also higher than the other two displays (mean=83.33%, STD=19.03%; mean=89.58%, STD=16.34%). There was no significant difference, either.

As shown in Table 21, the understanding accuracy about the destinations for the three displays was similar to one another.

Table 21. Understandability of AC Transit Travel Time Message

<table>
<tr><th></th><th>Graphical Images</th><th>Perceived easiness</th><th>Understanding accuracy of general information</th><th>Understanding accuracy about destinations</th></tr>
<tr><td>1</td><td>
</td><td>4.17 (0.87)</td><td>90.63% (14.39%)</td><td>20 (83.33%)</td></tr>
<tr><td>2</td><td>
</td><td>3.98 (1.24)</td><td>83.33% (19.03%)</td><td>19 (79.17%)</td></tr>
<tr><td>3</td><td>
</td><td>4.06 (1.24)</td><td>89.58% (16.34%)</td><td>20 (83.33%)</td></tr>
</table>

[9] ANOVA -short for "analysis of variance"- is a statistical technique for testing if 3(+) population means are all equal. https://www.spss-tutorials.com/anova-what-is-it/

Comments on detailed design elements

– Confusion of whether the sign is about AC Transit

Among the three displays, there were seven times that participants commented that they were confused and didn't know whether this was bus or BART service. Examples of participants verbatim were "*When you added Bart station to the sign, it makes it hard to understand it is Bus or Bart Service*." "*The symbol confuses me, I was not sure if it is bus or train although I know AC is a bus service*", "*couldn't make sense of the sign*" and "*Saw it all. But don't understand"*. The reason is due to the content of the sign, which happened to have the BART station as the origin of the AC transit bus.

– Confusion of where the origin is

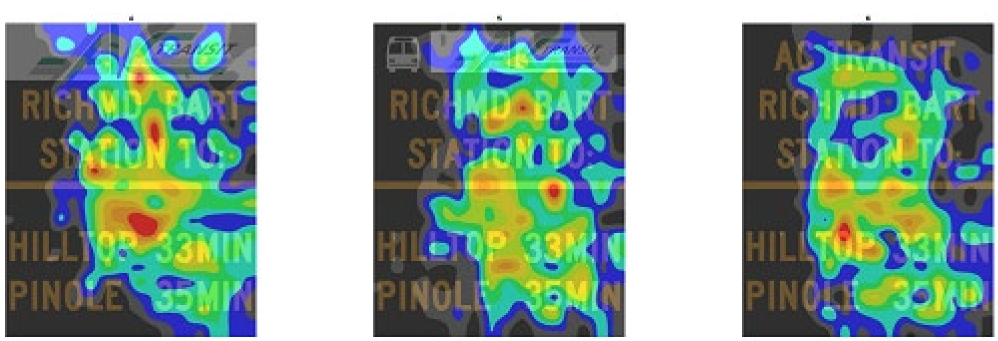
Partly due to the confusion of AC transit or BART, some participants could not tell the origin. Examples of participants' verbatim were "*(It is easy to understand) except the origin*" and *"(I) missed the starting point*". Another reason of why participants could not tell the origin was that they could not make sense of the layout of the display. It seemed not very clear to the participants by using the horizontal line to break the origin and the destinations. One participant commented on that "*The line seems to divide the time and location from top of the sign. [in fact], it should not do that.*"

– Preference of display with logo

Three participants mentioned the preference of the displays with the transit logo. Examples of their verbatim were *"(without logo), you have to read everything*" and "*better to have logo, which catches more of my attention.*"

Visual attention

The heatmaps for the AC Transit travel messages are shown in Figure 27. For all three displays, participants paid attention to both the first phase of the message (the starting point) and the second phase of the message (the destinations). As shown in Figure 27b, participants paid little attention to the bus symbol. It indicated that most participants didn't see the bus symbol, probably because it was too small or because participants were familiar with AC Transit, which was consistent with participants' comments.



a b c

Figure 27. Heatmap of Visual Attention for AC Transit Travel Time Message

WESTCAT travel time message

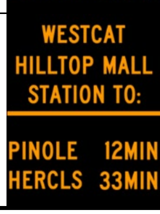
Understandability

Table 22 shows the results of understandability for each design of the AC Transit travel time message, which includes mean and standard deviation (STD) for easiness and understanding accuracy of general information, as well as results for understanding accuracy about destinations.

Single-factor ANOVA was used to analyze the statistical difference of perceived easiness and understanding accuracy of general information for the three displays. The perceived easiness of the display with the WESTCAT logo only (mean=4.61, STD=0.58) was higher than the other two displays (mean=4.30, STD=0.18; mean=4.22, STD=1.20). However, there was no significant difference. The understanding accuracy of the general information for the display with both bus symbol and the WESTCAT logo (mean=90.22%, STD=16.40%) was higher than the other two displays (mean=86.96%, STD=16.64%; mean=85.87%, STD=16.55%). However, the difference was not significant, either.

Understanding accuracy about the destination for the display with the WESTCAT logo only was slightly higher (66.67%) than the other two displays (50.00% and 58.33%).

Table 22. Understandability of WESTCAT Travel Time Message

	Graphical Images	Easiness	Understanding accuracy of general information	Understanding accuracy about destinations
1		4.61 (0.58)	86.96% (16.64%)	16 (66.67%)
2		4.30 (1.18)	90.22% (16.40%)	12 (50.00%)
3		4.22 (1.20)	85.87% (16.55%)	14 (58.33%)

Comments on detailed design elements

– Participants don’t know WESTCAT

Many of the participants didn’t know what WESTCAT was. For example, one participant mentioned that “I did not know about WESTCAT and that icon tells me that it's bus service”.

– Not enough information

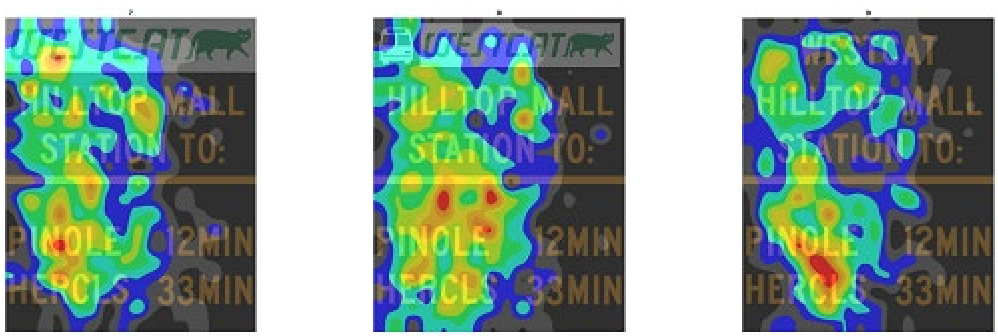
Four participants commented that the sign didn't provide enough information. Hence it was not useful. Examples of their reasons were "*which exit I should take to go to Hilltop mall station*" "*It is clear. But it is not useful*", and "*which exit I should take to go to the bus stop. How long does it take to go to the bus station?*"

– Preference of display with logo

Similar to the AC Transit message, participants showed a preference of having the WESTCAT logo on the display.

Visual attention

The heatmaps for the AC Transit travel messages are shown in Figure 28. For all three displays, participants paid attention to both the first phase of the message (the starting point) and the second phase of the message (the destinations). As shown in Figure 27b, participants paid some attention to the bus symbol. Relatively less attention was paid to the WESTCAT symbol. The reason could be that many participants were not familiar with WESTCAT. Hence, the bus symbol was needed for them to understand what WESTCAT was. Comparing Figure 27 and Figure 28, participants spent less attention on the first phase of the WESTCAT messages than on the first phase of the AC Transit messages, which indicated that the content (Richmond BART Station) of the first phase of the AC Transit messages was harder to understand.



a b c

Figure 28. Heatmap of Visual Attention for WESTCAT Travel Time Message

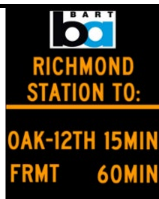
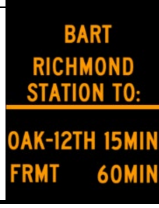
BART travel time message

Understandability

Table 23 shows the results of understandability for the BART travel time message, which include mean and standard deviation (STD) for easiness and understanding accuracy of general information, as well as results for understanding accuracy about destinations.

Paired t-test was used to compare the statistical difference between the display with the BART logo and the display with text only. The perceived easiness for the display with the BART logo (mean=4.59, STD=0.65) was higher than the display with text only (mean=4.52, STD=0.66). The understanding accuracy of the general information for the two displays was the same (mean=95.83%, STD=9.54%). However, no significant difference was found. For the display with the BART logo, the understanding accuracy about the destination was much higher (87.50%) than the display with text only (58.33%).

Table 23. Understandability of BART Travel Time Message

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about destinations
1		4.59 (0.65)	95.83% (9.54%)	21 (87.50%)
2		4.52 (0.66)	95.83% (9.54%)	14 (58.33%)

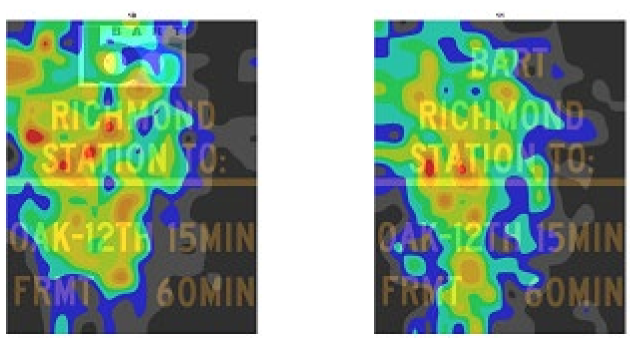
Comments on detailed design elements

– There were not many comments for the above two displays.

Three participants commented on the intention of having BART information on the highway, such as "*why show BART sign on highway*". Another participant had similar comment regarding the exit to the BART station. The verbatim was "*Which exit people should take for the BART station.*"

Visual attention

The heatmaps for the BART travel messages are shown in Figure 29. For the two displays, participants paid attention to both the first phase of the message (the starting point) and the second phase of the message (the destinations). In the two heatmaps, special attention was paid to the word "STATION", which indicated that "STATION" was the keyword for participants to understand the first phase (the starting point).



a b

Figure 29. Heatmap of Visual Attention for BART Travel Time Message

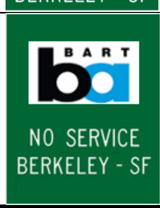
BART service special message

Understandability

Table 24 shows the results of understandability for the BART service special message, which includes mean and standard deviation (STD) for perceived easiness, understanding accuracy of general information, and perceived symbol helpfulness.

Paired t-test was used to compare the statistical difference between the display with the light rail symbol and the display with the BART logo. The perceived easiness for the display with the BART logo (mean=4.92, STD=0.28) was significantly higher (p=0.03) than the display with the light rail symbol (mean=4.46, STD=1.14). The understanding accuracy of general information for the display with the BART logo (mean=98.61%, STD=6.78%) was higher than the display with the light rail symbol (mean=97.22%, STD=9.43%). However, the difference was not significant. The rating of the symbol helpfulness for the BART logo (mean=4.96, STD=0.20) was significantly higher (p=0.001) than the light rail symbol (mean=4.35, STD=0.81).

Table 24. Understandability of BART Service Special Message

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Perceived symbol helpfulness
1		4.46 (1.14) *	97.22% (9.43%)	4.35 (0.81) *
2		4.92 (0.28) *	98.61% (6.78%)	4.96 (0.20) *

** Indicate significant difference, p-value < 0.05*

Comments on detailed design elements

– Background color

Five participants commented on the background color of the above two displays, in which they thought the background color should be yellow or red instead of green. Examples of their verbatim were "*if it is a problem, it should not be in green, you should use red or orange*", "*the background should be in red or orange*", and "*instead of green background, make it red to draw attention*".

– Preference of the BART logo in comparison with the light-rail symbol

Most participants preferred the BART logo over the light-rail symbol. Examples of participants comments regarding the light-rail symbol were "*The symbol is not Bart logo, it looks like Caltrain*", "*(It is) better to have the BART logo. This symbol looks like a train to me*", "*The symbol did not stand out. I mostly read the* *word rather than the symbol*". Examples of participants' comments regarding the BART logo were *"(It is) very easy to tell that the info is for BART*" and *"(It is) perfect*".

Visual attention

The heatmaps for the BART travel messages are shown in Figure 30. For the two displays, participants paid attention to the BART symbol and the word "NO", which was the keyword of the message. For the display with the light-rail symbol, as shown in Figure 30a, the light-rail symbol captured more attention than the keyword "NO", which indicated that the light-rail symbol took participants more attention to comprehend.

a b

Figure 30. Heatmap of Visual Attention for BART Service Special Message

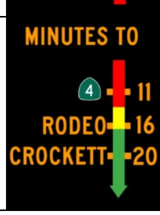
3. GRIPs

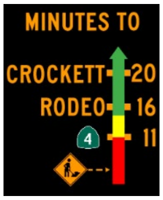
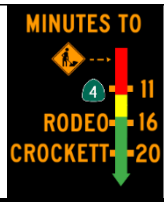
Single-link GRIP

Understandability

Seven displays of single-link GRIPs are tested. Table 25 shows the results of the first four single-link GRIPs, which include mean and standard deviation (STD) of perceived easiness, understanding accuracy of general information, as well as results for understanding accuracy of detailed information.

Table 25. Understandability of Single-link GRIP (1)

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about detailed information
1		4.38 (1.05)	100.00% (0) *	Orientation: 23 (95.83%)
2		4.38 (1.10)	100.00% (0) *	Orientation: 23 (95.83%)

3		4.71 (0.62)	98.96% (5.10%) *	Legend: 24 (100.00%)
4		4.15 (1.24)	96.88% (8.45%) *	NA

** Indicate significant difference, p-value < 0.05*

For these four displays of single link GRIPs, two of them have the bottom-top orientation. The other two have the top-bottom orientation. Two displays have the roadwork legend. The other two displays don't have the roadwork legend. Hence, the two-factor with replication ANOVA was used to analyze (1) the effect of orientation and (2) the effect of legend.

The perceived easiness for the display with bottom-top orientation (mean=4.38, STD=1.05; mean= 4.71, STD=0.62) was higher than the other three displays (mean=4.38, STD=1.10; mean=4.15, STD=1.24). However, no significant effect was found for the effect of orientation.

Regarding the understanding accuracy of general information, the effect of with and without legend was found significant (p=0.04). For the displays without legend, the understanding accuracy (mean=100.00%, STD=0) was significantly higher than the displays with legend (mean=98.96%, STD=5.10%; mean=96.88%, STD=8.45%).

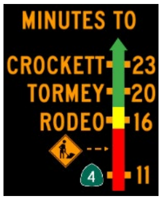
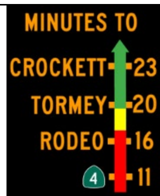
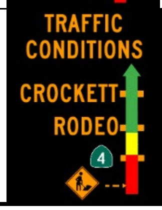
Regarding the understanding accuracy for the orientation, it was 23 out of 24 (95.83%) for both of the two displays, as shown in Table 25. The understanding accuracy for the legend was 100%.

Although there was no statistically significant difference, the results indicated that the (1) bottom-top orientation was easier to understand; (2) all participates understood the meaning of the roadwork legend.

Table 26 shows the results of the other three single-link GRIPs, which include mean and standard deviation (STD) of perceived easiness, understanding accuracy of general information, results for understanding accuracy of detailed information, as well as the helpfulness of using legend and colors.

Table 26. Understandability of Single-link GRIP (2)

Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about detailed information	Helpfulness

1		4.70 (0.47)	95.65% (9.70%) *	Legend: 23 (95.83%)	4.50 (1.09)
2		4.43 (1.16)	97.10% (9.59%) *	Color: 23 (100.00%)	4.43 (1.06)
3		4.67 (0.57)	95.83% (11.27%)	Color: 23 (100.00%)	4.83 (0.70)

** Indicate significant difference, p-value < 0.05*

The first two displays have 4 destinations, one with the roadwork legend and the other one without, as shown in Table 26. Together with the first display (3 destinations without legend) and the third display (3 destinations with legend) in Table 25, these four displays make another two-factor design. Hence, the two-factor ANOVA with replication was used to analyze (1) the effect of legend and (2) the effect of the number of destinations.

Regarding perceived easiness, no significant effect was found for neither factor.

Regarding understanding accuracy of general information, the effect of number of destinations was found significant (p=0.046). For the displays with 3 destinations, the understanding accuracy (mean=100.00%, STD=0; mean=98.96%, STD=5.10%) was significantly higher than the displays with 4 destinations (mean=95.65%, STD=9.70%; mean=97.10%, STD=9.59%).

As shown in Table 26, understanding accuracy of detailed information was 23 out of 24 (95.83%) for the legend on the first display. It was 23 out of 23 (100%) for the color on the second display. For both the legend and the color on the first and the second display, participants gave high ratings of the helpfulness (mean=4.50, STD=1.09; mean=4.43, STD=1.06).

For the third display in Table 26, it was compared with the third display in Table 25. The difference between these two displays lies in the travel time. Paired t-test was used for the analysis. There was no significant difference between these two displays regarding either the perceived easiness or the understanding accuracy of the general information. Regarding the color of this display, the understanding accuracy was 100%. The rating of the helpfulness of the color was also high (mean=4.83, STD=0.70).

Comments on detailed design elements

– Orientation

Many participants commented on the orientation of the single-link GRIPs. In 12 times, participants particularly mentioned that they liked the top-bottom orientation. The most frequently mentioned reason was that "this is the way I usually read". Examples of verbatim were "*This (bottom-top) orientation threw me off. I used to see closer (places) first (at the top)*" and *"The orientation is not I used to. I like the closest on top*".

In another 22 times, other participants particularly mentioned that they liked the bottom-top orientation. Examples of verbatim were "*Like arrow goes up*", "*The arrow is confusing. It should not point down*" and "*I should see the one I got first on the bottom*".

However, some participants also showed the learning effect of getting used to the orientation that they didn't like at first. One example of the participant's verbatim was "*I guess that I get used to it now*".

Overall, a higher number of participants preferred the orientation from bottom to top.

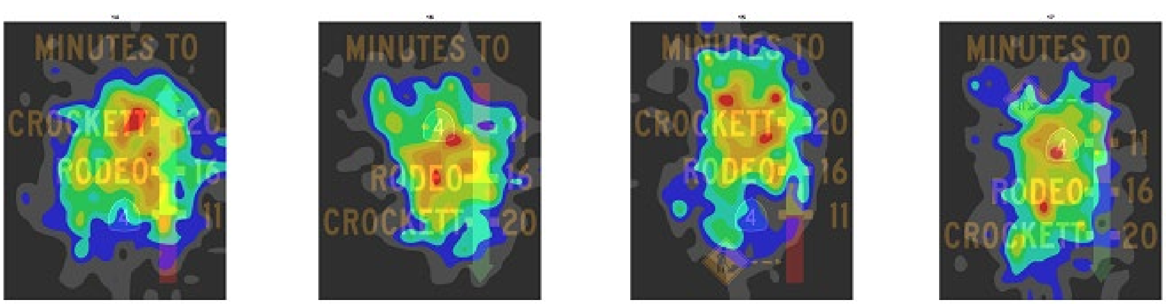
– Minutes and color

Several participants commented on having both minutes and colors. Examples of verbatim were "Time is more helpful", "How much of delay? The color is subjective", "I see the color after minutes. So, I am not sure they (the colors) are very helpful" and "Colors capture attention. Make me get prepared. But minutes (help me to) make decision".

As a summary, for those participants who commented on the function of minutes and color, both color and minutes are preferred. Providing the number of minutes is more helpful.

Visual attention

The heatmaps of the single-link GRIPs are shown in Figure 31. Participants paid most of the attention onto the link and the destinations. For example, as shown in Figure 31a, participants looked at all the three locations (Crockett, Rodeo, and Highway 4) and spent extra attention on the destination of the trip, which was Crockett. The patterns of the heatmaps for the other six displays were similar. For the display with only traffic conditions but not travel time, as shown in Figure 31g, participants also spent extra attention on the header of the sign, which was "TRAFFIC CONDITIONS". The reason was probably due to the uniqueness of this header comparing with the header ("MINUTES TO") of the other single-link GRIPs.



a b c d

e f g

Figure 31. Heatmap of Visual Attention for Single-link GRIP

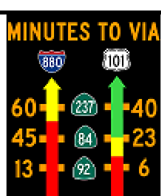
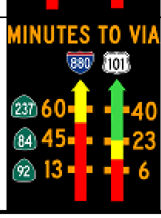
The fixation duration time was calculated for each single-link GRIP for each participant. ANOVA was used to compare the mean of fixation time for the six displays (from a to f, as shown in Figure 23), which had both travel time and traffic congestion information. However, no significant difference was found among those displays. A paired sample t-test was used to compare the mean of fixation time for the single-link display without travel time (as shown in Figure 23g) and its counterpart display with travel time (as shown in Figure 23c). It was found that the fixation duration time for the display with travel time (mean=14.09, STD=3.51) was significantly higher than the display without travel time (mean=12.17, STD=4.46), which indicated that having both travel time and traffic congestions information took significantly more of participants' visual attention.

Dual-link GRIP

Understandability

Six displays of dual-link GRIPs are tested. Table 27 shows the results of the first two dual-link GRIPs, which include mean and standard deviation (STD) of perceived easiness, understanding accuracy of general information, results for understanding accuracy about destinations, as well as understanding accuracy and perceived easiness of the positions of destinations.

Table 27. Understandability of Dual-link GRIP (1)

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about destinations	Position of destinations	
					Understanding accuracy	Easiness
1		4.21 (1.06)	98.61% 6.8%)	22 (95.65%)	22 (95.65%)	3.91 (1.44)
2		4.27 (0.64)	98.61% (6.8%)	19 (82.61%)	22 (95.65%)	4.39 (1.08)

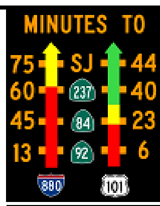
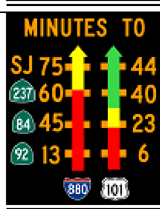
A paired t-test was used to compare the difference between the two displays, regarding the perceived easiness and the understanding accuracy of general information. However, there was no significant difference.

The understanding accuracy about the destination was 22 out of 23 (95.65%) for the first display and 19 out of 23 (82.61%) for the second display.

Regarding the position of destinations, the understanding accuracy for both positions was the same: 22 out of 23 (95.65%). Paired t-test was used to test the difference in the easiness of the two positions. No significant difference was found.

Shown in Table 28 are the results of another two dual-link GRIPs, which include mean and standard deviation (STD) of perceived easiness, understanding accuracy of general information, as well as results of understanding accuracy about destinations.

Table 28. Understandability of Dual-link GRIP (2)

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about destinations
1		4.11 (1.15)	91.67% (17.72%)	17 (70.83%)
2		4.20 (0.94)	95.83% (11.26%)	21 (87.50%)

Paired t-test was used to compare the difference in perceived easiness and understanding accuracy of the general information of these two displays. There was no significant difference.

Understanding accuracy about the destinations was 17 out of 24 (70.83%) for the first display and 21 out of 24 (87.50%) for the second display.

Paired t-test was used to compare the two displays with four destinations and the two displays with three destinations. Results are shown in Table 29. No significant difference was found for perceived easiness. The understanding accuracy of general information for the displays with three destinations (mean=98.61%, STD=4.71%) was significantly higher (p=0.035) than the displays with four destinations (mean=93.75%, STD=11.85%).

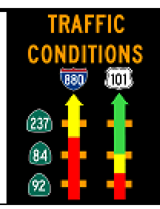
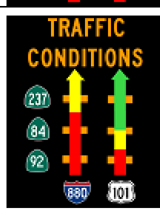
Table 29. Comparison of Understandability between 3-Destination and 4-Destination Dual-link GRIPs

Dual-link: Number of destinations	Perceived easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Dual-link with three destinations	4.24	0.71	98.61% *	4.71%
Dual-link with four destinations	4.10	0.76	93.75% *	11.85%

** Indicate significant difference, p-value < 0.05*

Table 30 shows the results of the last two dual-link GRIPs, which include mean and standard deviation (STD) of perceived easiness, understanding accuracy of general information, results of understanding accuracy about destinations, as well as understanding accuracy and perceived easiness of the positions of highway options.

Table 30. Understandability of Dual-link GRIP (3)

	Graphical Images	Perceived easiness	Understanding accuracy of general information	Understanding accuracy about destinations	Position of highway options	
					Understanding accuracy	Easiness
1		4.42 (0.71)	100.00% (0)	24 (100.00%)	20 (95.24%)	4.52 (1.17)
2		4.54 (0.66)	97.92% (10.21%)	23 (95.83%)	21 (100.00%)	4.71 (0.56)

Paired t-test was used to compare the difference between the two displays regarding the perceived easiness and the understanding accuracy of general information. However, there was no significant difference.

The understanding accuracy about the destination was 100.00% for the first display and 23 out of 24 (95.83%) for the second display.

Regarding the position of highway options, the understanding accuracy for the display having the highway options on top was 20 out of 21 (95.24%) and 100% for the display having the highway options at the bottom. Paired t-test was used to test the difference in easiness of the two displays. However, no significant difference was found.

Comments on detailed design elements

- Too much information

Many participants perceived the dual-link GRIP to have too much information on one sign. Examples of participants' verbatim were "It is a lot going on", "Too much information. They need to be separated into two signs", "I notice that I slowed down top read them", "Too much information. I look to colors, time, highways. Then I have already passed the sign." and "It's hard to take all information in."

- Positions of destinations

There were five participants who specifically pointed out that having the destinations in the middle was easier to read than having the destinations on the left. Examples of the verbatim were "*It is easier (with the destinations) in the middle*" and "*I prefer the SJ (one of the destinations) in the middle*". There was only one participant who specifically mentioned: "*Like the destinations on the left*".

- Position of highway shields

Three participants mentioned that they preferred to have the highway shields (880 or 101) on the top of the display. All the above three comments were made after they were viewing the dual-link GRIP with the header of "MINUTES TO", with four destinations on the left and with the two highway shields at the bottom of the display. The verbatim was "*It would be quicker if the highway shield were on top beside of SJ*", "*Better to have the 880 and 101 on top, close to the header*", and "*Put the highway shields on top*". On the other hand, there was one participant who mentioned that "*The highways were at the bottom. It helps.*"

- Not familiar with the highways (237, 92, and 84)

Due to unfamiliarity with the highways, six participants commented that they preferred to have city names on the display. Most participants of the simulator study were local commuters of I-80, who travelled less frequently on those highways on this sign. Examples of their verbatim were *"(I need to be) familiar with the area*", "*I like the SJ which means San Jose. I saw the other freeways but I did not pay attention to them*", and "*(It is) harder to catch number of highways (as beside of number of minutes)*".

Visual attention

The heatmaps of the dual-link GRIPs are shown in Figure 32. Participants paid most of the attention to the two links and the destinations. For example, as shown in Figure 32a, participants looked at the two links and the destinations located in the middle. The patterns of the heatmaps for all six dual-link GRIPs were similar. For the display with only traffic conditions but not travel time, as shown in Figure 32e and Figure 32f, participants also spent extra attention on the header of the sign, which was "TRAFFIC CONDITIONS". In comparison, less attention was paid to the header of "MINUTES TO" or "MINUTES TO VIA" on the other four displays.

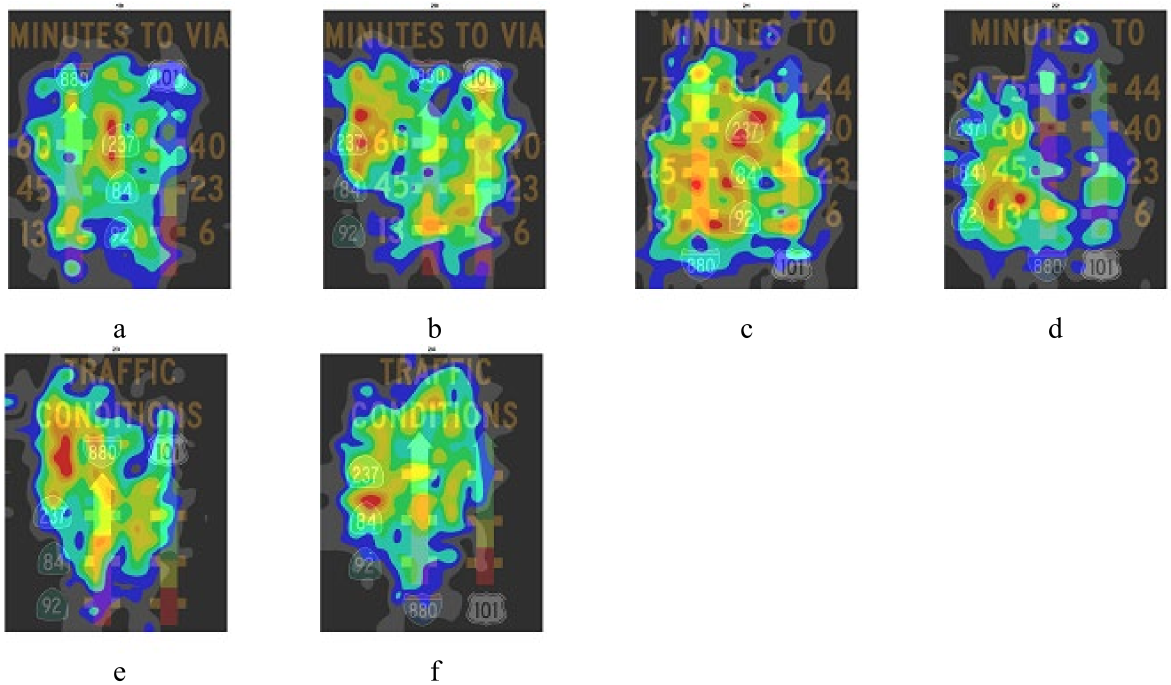


Figure 32. Heatmap of Visual Attention for Dual-link GRIP

The fixation duration time was calculated for each dual-link GRIP for each participant. A paired t-test was used to compare the mean of fixation time (Figure 24a vs. Figure 24b; Figure 24c vs. Figure 24d; Figure 24e vs. Figure 24f). It was found that the fixation duration time for the display shown in Figure 24c (mean=12.44, STD=3.73) was significantly shorter than the display shown in Figure 24d (mean=16.00, STD=2.26), which indicated that having the destinations in the middle was easier for participants to look comparing with having the destinations on the left, which was consistent with participants' comments. There was no significant difference for the other two pairs.

Comparison between single-link GRIP and dual-link GRIP

As shown in Table 31, a paired t-test was used to compare the difference between the single link GRIP and the dual-link GRIP, regarding both perceived easiness and understanding accuracy of general information. A marginal significant difference was found for the perceived easiness, with the p-value of 0.06. The perceived easiness for single-link GRIP (mean=4.45, STD=0.61) was higher than the dual-link GRIP (mean=4.28, STD=0.50).

Table 31. Comparison between Single-link and Dual-link GRIPs

GRIPs	Perceived easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Single link	4.45	0.61	98.07%	3.15%
Dual link	4.28	0.50	96.98%	4.50%

The average fixation duration time for single-link GRIPs and dual-link GRIPs was significantly different (p=0.0001). The fixation duration time for single-link GRIP (mean=13.88, STD=2.28) was lower than dual-link GRIP (mean=16.06, STD=1.50), which indicated that averagely participants took more time to look at the dual-link GRIPs.

Comparison between three-line travel time message and GRIP

In order to compare the single-link GRIP and dual-link GRIP with the three-line travel time message as the baseline, a paired t-test was used for the analysis of both perceived easiness, understanding accuracy of general information, as well as fixation duration time. As shown in Table 32, the understanding accuracy of general information for three-line message (mean=100.00%, STD=0) is significantly higher (p=0.004) than the single-link GRIP (mean=97.45%, STD=4.31%). As the total viewing time for GRIP was 20 seconds and 15 seconds for the three-line travel time messages, the percentage of fixation duration time was used for comparison purposes. It was found that the percentage of fixation time for the single-link GRIP (mean=69.38%, STD=11.38%) was significantly lower (p=0.001) than the three-line message (mean=79.26%, STD=14.04%), which implied that averagely single-link GRIP took less of participants' attention in comparison with the three-line message.

Table 32. Comparison between Three-line Travel Time Message and Single-link GRIPs

Messages	Perceived easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Three line	4.63	0.71	100.00% *	0
Single link	4.42	0.62	97.45% *	4.31%

** Indicate significant difference, p-value <0.05*

As shown in Table 33, there was a significant difference between the three-line message and dual-link GRIP regarding both the perceived easiness and understanding accuracy of general information. Dual-link GRIP (mean=4.27, STD=0.49) had lower perceived easiness than the three-line message. Its understanding accuracy of general information (mean=97.11%, STD=4.45%) was also significantly lower than the three-line message. Regarding the fixation duration time, no significant difference was found between the three-line message and the dual-link GRIP.

Table 33. Comparison between Three-line Travel Time Message and Dual-link GRIP

Messages	Perceived easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Three line	4.63 *	0.71	100.00% *	0
Dual link	4.27 *	0.49	97.11% *	4.45%

** Indicate significant difference, p-value <0.05*

Effects of Gender, Language, and Education

The effects of participants' gender, native language, and education levels on the understanding of all displays are analyzed. As shown in Table 34, understanding accuracy of female-driver participants (mean=96.07%, STD=3.01%) was higher than male-driver participants (mean=93.59%, STD=3.15%), with a marginal significant difference (p=0.07).

Table 34. Simulator Testing: Gender Effect

Gender	Average rating of easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Female	4.45	0.36	96.07%	3.01%
Male	4.24	0.60	93.59%	3.15%

As shown in Table 35, native English-speaking participants gave relatively lower ratings (mean=4.28, STD=0.50) of overall easiness of all displays than native Spanish-speaking participants (mean=4.49, STD=0.41). However, the difference was not significant. The understanding accuracy of general information for native English-speaking participants (mean=95.22%, STD=3.20%) was higher than native Spanish-speaking participants (mean=94.62%, STD=3.47%). The difference was not significant.

Table 35. Simulator Testing: Language Effect

Language	Average rating of easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
Native English speakers	4.28	0.50	95.22%	3.20%
Native Spanish speakers	4.49	0.41	94.62%	3.47%

As shown in Table 36, there was no significant effect of education levels on the understanding of all signs. Looking at the average rating of easiness from high-school participants (mean=3.88, STD=1.03), it was relatively lower than the rating of college graduate participants (mean=4.45, STD=0.42) and postgraduate participants (mean=4.35, STD=0.35). Similarly, the understanding accuracy of high-school participants was also relatively lower than college graduate participants and postgraduate participants. As reported in the Participants section, the education level was not a blocking factor. Hence, the number of participants from each education level was not balanced, which could be one reason for the non-significant effect of education.

Table 36. Simulator Testing: Effect of Education

Education	Average rating of easiness		Understanding accuracy of general information	
	Mean	STD	Mean	STD
High school	3.88	1.03	91.88%	2.04%
College	4.45	0.42	95.09%	3.59%
Post graduate	4.35	0.35	96.15%	2.12%

Section 5: Discussion and Conclusion

In this section, we discuss four design principles for all IDB signs based on the results from the simulator testing, including the amount of information on one sign, layout with two virtual phases, use of symbols and colors, and transit travel time messages.

Amount of Information on One Sign

For the signs with travel time and up to six lines messages, results indicated that most participants could understand the general information (e.g., travel time to certain highways) on the three-line message and five-line message. However, many participants commented that there was too much information on the five-line message. For the six-line message, most of the participants could figure out the connection between the two phases. However, because there was more information than they could process, many participants could not pick up information from the first phase of the message (DOWNTOWN SF). Hence, their general understanding accuracy for the six-line message was relatively low. For the purpose of improving the understandability of the signs as well as improving drivers' satisfaction, it is suggested maintaining the information on one sign less than five lines.

For the GRIPs, many participants were overwhelmed by the amount of information on the dual-link GRIPs. When compared with the three-line travel time message, both the perceived easiness and understanding accuracy of general information of the dual-link GRIP were significantly lower. In comparison with the single-link GRIP, the perceived easiness of dual-link GRIP was also significantly lower. The average fixation duration time as measured from the eye tracker showed that participants took significantly longer time to look at the dual-link GRIP than the time to look at the single-link GRIP. The most likely reason for the lower understandability and higher demand for visual attention of dual-link GRIP is due to the amount of information on the sign. On the other hand, there were either 3 or 4 destinations on single-link GRIPs. The comparison of the single-link GRIP and the three-line message indicated no significant difference regarding the perceived easiness. The average fixation duration time for single-link GRIP was even significantly shorter than the three-line message.

Layout with Two Virtual Phases

For the five-line travel time message, six-line travel time message, as well as transit travel time messages, there was a horizontal line to break the information on one sign into two virtual phases. Based on the test results of the five-line message, participants experienced confusion regarding the purpose of the horizontal line. They looked at the sign and tried to figure out the connection between the two phases. However, it didn't make much sense to most of the participants. Hence, they didn't see the connection between the first phases with the second phase and didn't go to the details of the second phase.

For the six-line travel time message, the horizontal line divided the message into two virtual phases, with the first phase about traffic information for one city and the second phase about information for another city. With this layout, participants could easily spot the city that was related to their destination and not necessarily needs to understand the information of the other city. For this type of information break-down, the layout with two virtual phases works well.

For the transit travel time messages, the horizontal line also divided the messages into two virtual phases. The first phase was the transit type (e.g., AC Transit or WESTCAT) and the starting point. The second phase was the travel time to different destinations. For the purpose of dividing the starting point and the destinations, the layout of having two virtual phases divided by the horizontal line also works well, although some participants took a while to figure it out.

Overall, the layout of having two virtual phases divided by the horizontal line should be applied only when there were some sorts of relevance between the two phases of information. If the two phases are not relevant to each other, it will be confusing with the use of the horizontal line to break the sign into two phases just in order to have a smaller number of lines of information in each phase.

Use of Symbols and Colors

Several symbols have been studied in the simulator testing, including the little bus symbol and the BART symbol. Compared with the design with pure text, the design with symbols was perceived to convey the message faster. With symbols, drivers could catch the meaning with a glance. Otherwise, they have to read through the texts. The preference of symbols over text is consistent with the findings from the lab testing.

After the lab testing, a new design with an additional bus symbol to indicate bus service was proposed for both the AC Transit and WESTCAT related messages. The intention of the design was to help drivers who are not familiar with either AC Transit or WESTCAT to recognize the intended contents. Based on the results of the simulator testing, on the one hand, the bus symbol did help drivers to know that WESTCAT is a bus service. On the other hand, the bus symbol was not perceived as very helpful by many participants. One reason is that two symbols together (bus symbol and the transit logo) make the sign look crowded. Another reason is that the bus symbol was too small and less likely to attract drivers' attention while they were driving.

For the BART symbol, almost all participants preferred to have the BART logo, rather than the light-rail symbol. One reason is that the light-rail symbol could mean other train services (e.g., Cal-train, Amtrak), which implies that it is not specific enough for BART. Hence it took longer time and more effort to process.

Regarding the use of colors on the displays, the green background color for the BART service special message was commented as problematic by many participants, which was consistent with the findings from the lab testing. Other background colors such as yellow or red are more conventionally used colors for special alert messages, which make more sense to drivers.

Transit Travel Time Messages

For the transit travel time messages, many participants didn't understand the purpose of displaying the transit travel information on highways. Participants were confused about why they would want to see the bus information while they were driving. They wanted to make the connection between seeing bus travel time information on the highway and their driving trip. One typical scenario that many participants thought of was to get off the highway and drive to the bus station. The other was to drive from the bus station to different destinations (e.g., Pinole, Hercules). Hence, they had difficulty figuring out the scenario that the sign was trying to convey. Additionally, they had difficulty in making sense of the travel time.

Less confusion was observed for the BART travel time message. One reason is that driving and taking BART makes more sense in comparison with driving and taking a bus. Based on participants' feedback to the transit travel time messages, the connection between driving and taking public transportation is perceived as critical for the understandability of the signs.

References

[1] Charlton, S. G. (2006). Conspicuity, memorability, comprehension, and priming in road hazard warning signs. Accident Analysis and Prevention, 11. https://doi.org/10/b2nphb

[2] Fisher, J. (1992). Testing the Effect of Road Traffic Signs' Informational Value on Driver Behavior. Human Factors, 34(2), 231–237. https://doi.org/10/ghnrhd

[3] Gao, X. W., Podladchikova, L., Shaposhnikov, D., Hong, K., & Shevtsova, N. (2006). Recognition of traffic signs based on their colour and shape features extracted using human vision models. J. Vis. Commun. Image R., 11. https://doi.org/10/btwn7t

[4] Hills, B. L. (1980). Vision, Visibility, and Perception in Driving. Perception, 9(2), 183–216. https://doi.org/10/d7wzsz

[5] Martens, M. H. (2000). Assessing Road Sign Perception: A Methodological Review. Transportation Human Factors, 2(4), 347–357. https://doi.org/10/fvnhmh

Part III: Plan for Field Testing

Introduction

As the third phase of this project, it was planned to do a field testing in order to evaluate the legibility distance of the IDB signs, which were proposed to be displayed on the I-80 Corridor. Therefore, the PATH team did a literature review about legibility distance of CMS and proposed the experimental design, including the testing routes, testing time, lane position, driving speed, participants, testing procedure, as well as measurements. The PATH team also reviewed the testing plan with engineers from Caltrans District 4 and revised the experimental set-up based on their feedback. The revised testing plan is summarized as follows:

Literature Review

Sign legibility is greatly affected by specific design characteristics that contribute to drivers' ability to perceive and understand a sign's message. Key design parameters determining the legibility of signs include legend color, font size, font style, and retroreflectivity (Campbell et al., 2012). Legibility distance is the distance from which a driver is able to read a sign (Upchurch et al., 1992). The legibility distance affects the maximum number of units of information that a CMS should display that will allow drivers to read and comprehend the message at prevailing operating speeds (B. R. Ullman et al., 2009). Generally, the larger the character, the farther away the message can be read by motorists, and so the longer the amount of time they have to read the sign (B. Ullman, Ullman, Dudek, & Ramirez, 2005).

Effects of different time of the day

Many studies have found that the legibility distance is different during the daytime and at night (e.g., Fitzpatrick et al., 2016). In general, legibility distance during the daytime is longer than at night. For example, the suggested legibility distance for the standard CMS messages (standard font, 18-inch character heights, approximately 13-inch character widths, and about 2.5-inch stroke widths) using LED technology is 800ft during the day and 600ft at night (B. R. Ullman et al., 2009).

Effects of the age difference

Past research has confirmed the substantial effects of age on visual performance. It was found that the age of the participants had an impact on sign detection time, both during the day and at night. Detection time for younger participants was shorter than the detection time for older participants (Fitzpatrick et al., 2016). Besides, older drivers are more sensitive to and take longer to recover from, glare (Upchurch et al., 1992).

Environment and task complexity:

It would be expected, in view of drivers' limited attentional capacity, that traffic signs are less likely to be reported in visually complex and attentional demanding environments, such as in the vicinity of major intersections, or, more generally, in urban areas compared with rural. (MacDonald & Hoffmann, 1991)

Methodology design

Participants

All participants should be local commuters of the I-80 Corridor. They are required to be licensed drivers and to pass a visual acuity test for 20/40 vision[10]. As the IDB displays colorful content, participants are

[10] The California DMV's visual acuity screening standard is 20/40 or better with both eyes together. (https://www.dmv.ca.gov/portal/dmv/detail/pubs/brochures/fast_facts/ffdl14)

also required to pass the color vision test. Based on the results of the lab testing and simulator testing, gender didn't have a significant effect on the understanding of the IDBs. Hence, the factor of gender will not be included in this phase. In total, we plan to have 24 participants considering the two independent variables (i.e., time of the day and age). The number of participants from each sub-group is shown in Table 37.

Table 37. Field Testing: Number of Participants from Each Group

All native English speaker	Time of the day		Sub-total
	Daytime	Nighttime	
Young driver	6	6	12
Elder driver	6	6	12
Sub-total	12	12	24

Time of the Day, Lane Position, and Driving Speed

The study will be conducted under both daytime and nighttime conditions. Nighttime studies will be conducted when it is completely dark.

Before approaching the IDB sign, participants will be instructed to drive on the right lanes of the highway and follow the speed of the traffic flow. They will be instructed to prompt once they can clearly see the content on the sign. Then, participants will keep driving. They will be instructed to prompt again at the point the vehicle passes the sign, and they could not see any content on the sign.

The average traffic flow of the I-80 corridor varies on different days of the week and at different times of the day. For the eastbound of the corridor (from the Bay Bridge to Carquinez Bridge), the peak traffic usually starts from 3:00 pm in the afternoon. For the westbound of the corridor (from the Carquinez Bridge to the Bay Bridge), the peak traffic usually starts in the very early morning till around 10 am. In order to avoid heavy traffic, we will arrange a daytime study between 10 am and 3 pm. For the nighttime study, we will start when it is completely dark, and without heavy traffic.

The IDB controllers collect the data from the photocell sensors and change the brightness based on the ambient light, so the brightness of the IDBs varies throughout the day. Usually, during the daytime, around 10 am in the morning or around 4 pm in the afternoon, the IDBs are always brighter than the night time (around 6 pm or 7 pm at night). Therefore, we also plan to get the luminance information of the IDB.

Measurements

Legibility distance and viewing time

The distance and time will be recorded when the participants first notice an IDB sign, when the participants can clearly read the message on the sign, and when the vehicle passed under the sign. Values for legibility distance and viewing time will be calculated from the measured distances. An example of legibility distance and viewing time is as follows: The lane closed ahead sign had an average legibility distance of 1,040ft at the speed of 65mph; the reading time was approximately 11s.

Viewing comfort

Besides legibility distance, we will also measure viewing comfort, which describes any discomfort caused by glare or harshness of the light. A brightly lit sign is sometimes so bright that it causes discomfort for a driver because the eye is slow to adapt to a bright light source in dark surroundings. After a driver passes the sign, the eye may adapt slowly to the dark surroundings, causing discomfort due to the inability to see well. These effects are more pronounced in locations where ambient light levels are very low (rural and semirural areas) and less pronounced where ambient light levels are medium to high (such as urban freeways). (Upchurch et al., 1992)

Participants will be asked to rate their discomfort with the glare from the IDBs as comfortable, irritating, or unbearable. According to (Fitzpatrick et al., 2016), the description of the three levels are as follows:

- Comfortable: The glare was not annoying, and the signal was easy to look at.
- Irritating: The glare was uncomfortable; however, participants could still look at it without the urge to look away.
- Unbearable: The glare was so intense that participants wanted to avoid looking at it.

Research Vehicle

The instrumented vehicle (Lincoln MKZ or MKS) for Automated Vehicle research in UC Berkeley will be leveraged and used as the testing vehicle for the IDB project. The instrumented vehicle is equipped with a GPS module, outward-facing camera, as well as IMU (inertial measurement unit). We will add an inward-facing camera onto the existing set-up of the instrumented vehicle in order to capture drivers' facial information (e.g., visual attention and head position) as well as the verbal responses from the driver (e.g., read aloud the message on the IDB). Participants will drive the instrumented vehicle during the testing. In post-processing, the GPS, video from the outward-facing camera, and the inward-facing camera will be synced based on the time.

Study Procedures

We will conduct a dynamic driving study on the I-80 corridor, with the study participants operating the instrumented vehicle.

Before approaching the IDB sign, participants will be instructed to drive on the right lanes of the highway and follow the speed of the traffic flow. Before approaching each IDB sign, they will be instructed to look at the sign and prompt the content of the message once they are able to clearly see words and symbols on the sign. Participants will keep driving. They will be instructed to prompt at the point the vehicle passes the sign, and they could not see any content on the sign.

In order to let participants get familiar with the study procedure before driving on the public road, there will be a practice session on the test track at Richmond Field Station. In the practice session, they will learn to operate the instrumented vehicle, learn to provide verbal response once they can clearly see a sign.

Participants will begin the study from the Richmond Field Station and drive along a pre-defined route, which will pass by the six IDBs installed on the corridor. An experimenter will be seated in the passenger seat to provide instructions to the participants and operate the data collection equipment (including recording verbal responses from the participants). Participants will be instructed to read the message aloud as soon as practicable while still maintaining the driving speed and staying within the required lane. The study will take about 1.5 hours from meeting the participant to the participant receiving their payment.

In order to counter the learning effects which may account for the variation of legibility distance for different IDB signs, three pre-defined routes will be used, as shown in Table 38. Participants will be randomly assigned to drive along with one of the three pre-defined routes.

Table 38. Field Testing: Pre-defined Routes

Routes	First IDB location	Waypoints				Last IDB location
Route 1	Gilman	Powell	Hilltop	Pinole	Willow	Cutting
Route 2	Hilltop	Pinole	Willow	Cutting	Gilman	Powell
Route 3	Cutting	Gilman	Powell	Hilltop	Pinole	Willow

References

[1] Campbell, J. L., Lichty, M. G., Brown, J. L., Richard, C. M., Graving, J. S., Graham, J., Harwood, D. (2012). NCHRP 600 - Human Factors Guidelines for Road Systems, 2nd Edition. Final Report.

[2] Fitzpatrick, K., Avelar, R., Pratt, M., Brewer, M., Robertson, J., Lindheimer, T., & Miles, J. (2016). Evaluation of Pedestrian Hybrid Beacons and Rapid Flashing Beacons. Retrieved from https://www.fhwa.dot.gov/publications/research/safety/16040/16040.pdf

[3] MacDonald, W. A., & Hoffmann, E. R. (1991). Drivers' awareness of traffic sign information. Ergonomics, 34(5), 585–612. https://doi.org/10.1080/00140139108967339

[4] Ullman, B. R., Trout, N. D., Dudek, C. L., Institute, T. T., Transportation, T. D. of, & Administration, F. H. (2009). Use of Graphics and Symbols on Dynamic Message Signs: Technical Report, 7(2), 192p. Retrieved from https://trid.trb.org/view/890543

[5] Ullman, B., Ullman, G., Dudek, C., & Ramirez, E. (2005). Legibility Distances of Smaller Letters in Changeable Message Signs with Light-Emitting Diodes. Transportation Research Record, 1918(1), 56–62. https://doi.org/10.3141/1918-07

[6] Upchurch, J., Armstrong, J. D., Baaj, M. H., Thomas, G. B., JD, A., MH, B., & GB, T. (1992). Evaluation of Variable Message Signs: Target Value, Legibility, and Viewing Comfort. Transportation Research Record, (1376), 35–44.

Part IV: Request for Experimentation

Introduction

After completing the lab testing and simulator testing, the PATH team drafted a report requesting permission for experimentation to FHWA. In the report, we first summarized the experiment set-up and findings of both the lab testing and the simulator testing. We also explained the potential benefits and importance of further testing the IDB signs in the field. Hence, we requested to display some IDB signs for experimentation purposes. However, after reviewing the report, FHWA denied the request due to various considerations. The PATH team drafted a rebuttal letter to further appeal to FHWA. In this part, we summarize the request for permission to experiment and also the rebuttal letter to FHWA.

Section 1: Discussion and Request for Permission to Experiment

The PATH team drafted a report based on the findings from both the lab testing and simulator testing to request permission to experiment from FHWA. In this section, we summarize the findings for each category of messages combining results from both lab testing and simulator testing. Then we elaborate on the need for on-road experiments in order to conduct a more comprehensive evaluation of the performance of the IDBs, including legibility distance and impact on traffic flow.

Summary of Findings for Each Category of Messages

- Three-line travel time message as a baseline

The three-line travel time messages in the lab testing could be used as a baseline for the evaluation of other categories of messages. In the lab testing, the average viewing time of the three-line travel time messages ranged from 13.22 seconds to 15.64 seconds. In the simulator testing, participants perceived the symbols on the three-line travel time messages as very helpful (the average rating was 4.63).

- Understandability of BART logo

In the lab testing, 91.8% of participants understood the message on the display with the BART logo. Besides, the display with BART logo took significantly less viewing time (14.12 seconds) compared with the display in the pure text format without BART logo (18.98 seconds). The average viewing time for the BART travel time message with the BART logo was similar to the average viewing time for the three-line travel time messages. Results in the simulator testing showed that the perceived easiness for the display with the BART logo was higher than without BART logo. The understanding accuracy of the general information was above 95.0%. For the display with the BART logo, the understanding accuracy about the destination was much higher than the display with text only. Combining results from both the lab testing and the simulator testing, the BART logo was easy to understand for all participants. It also significantly decreased the viewing time of the message. Therefore, the BART logo is suggested to be used as the symbol to present BART on IDBs.

- Up to six lines messages

The six-line messages have been approved for experiments in the field. We tested four-line and five-line messages in this study and summarized the results here.

For the four-line messages, the abbreviation of using MI for miles and MIN for minutes caused confusion to the participants. It was suggested to spell the MI out for improvement. Besides, the color of the horizontal line didn’t have a significant effect on the understandability.

In the simulator testing, results indicated that all participants could understand the general information on the five-line messages. Five-line message signs were not tested in lab testing.

In the lab testing, the six-line message with two destinations was well understood. Participants commented that this was a good way to show alternative routes or transportation modes. For the six-line messages, most of the participants (more than 90.0%) could figure out the connection between the two virtual phases and understand the general information of the message. The horizontal line divided the message into two virtual phases, with the first phase about traffic information for one city and the second phase about information for another city. With this layout, participants could easily spot the city that was relevant to their destinations and not necessarily needs to understand the information of the other city.

From the visual attention data of both five-line and six-line messages, participants didn't pay attention to all lines but rather mostly on two or three lines, among which it was relevant to their destinations. The fixation duration time for the five-line message was even shorter than the three-line message, which implies that more lines of information do not necessarily require more time to find the relevant traffic information. Caltrans will use the results of the human factors study to help guide and update their ongoing experimentation of six-line messages on the IDB.

– Single-link GRIP

For the single-link GRIP, the understanding accuracy in the lab testing was all above 83.0%. There was a significant learning effect. Many participants mentioned that it was easier to understand the single-link GRIP after viewing a few of them. In the simulator testing, the understanding accuracy was above 95.0% for single-link GRIPs. The rating of the helpfulness of the colors was also high (above 4.4, out of 5). Results indicated that the bottom-top orientation was easier to understand. All participants understood the meaning of the roadwork legend. In summary, the single-link GRIP was easy to understand and helped drivers to find the information they needed.

– Dual-link GRIP

For the dual-link GRIP, participants commented that it was complicated and took them a while to read and comprehend in the lab testing. Compared with the dual-link GRIP with the six-line message, the dual-link GRIP led to lower understanding accuracy.

Some participants explicitly expressed the preference of the six-line message over dual-link GRIP. One reason is that the format of the six-line message is simpler than the format of dual-link GRIP. Both of the two types of messages have different routes and multiple destinations. However, the dual-link GRIP has 3 or 4 destinations. It is relatively more complicated and harder to understand compared with having two destinations on the six-line message.

In the simulator testing, participants also commented on the amount of information on the dual-link GRIP. When compared with the three-line travel time message, both the perceived easiness and understanding accuracy of the dual-link GRIP were significantly lower. In comparison with the single-link GRIP, the perceived easiness of dual-link GRIP was also significantly lower. The average fixation duration time showed that participants took a significantly longer time to look at the dual-link GRIP than the time to look at the single-link GRIP. The number of destinations on the dual-link GRIP matters. In comparison with the dual-link GRIPs with four destinations and the dual-link GRIPs with three destinations, results showed that understanding accuracy of general information for the displays with three destinations was significantly higher than the displays with four destinations. Hence, displaying a fewer number of destinations (e.g., 2 or 3) on dual-link GRIPs would make it easier for drivers to understand this category of messages. Besides, having the destinations in the middle was easier for participants to read and comprehend comparing with having the destinations on the left.

Request for experimentation

Overall, the travel time message using the BART logo and single-link GRIP have demonstrated high understandability and improved effectiveness of presenting additional traveler information. Hence the research team recommends the experimental use of the BART logo and the single-link GRIP messages on IDBs installed along the project corridor. For the dual-link GRIP, it was recommended to start the experimentation with a fewer number of destinations, which will definitely be easier for drivers to learn and to understand. Subsequently, Caltrans can evaluate and make further decisions regarding whether a dual-link GRIP with more destinations should be displayed for on-road experiments.

Going forward, some on-road experiments to collect additional data may provide additional insights for Caltrans to help FHWA to expand options in the MUTCD. For example, driving experiments will allow Caltrans to evaluate the legibility distance of currently non-approved CAMUTCD IDB displays. Per the CA MUTCD, the recommended legibility distance for CMS is 600 feet for nighttime conditions and 800 feet for daylight conditions. The measurement of legibility distance will assess the adequacy of the size of graphics/symbols and letter heights of the IDBs. The on-road testing with subjects and instrumented vehicles can also allow the evaluation of potential effects to motorists viewing and comprehending the information displayed on the IDBs.

Additionally, with on-road experimental testing of various signs on the I-80 corridor, some observational studies of traffic speeds, freeway throughput, travel time reliability, and accidents rates can potentially be used to assess the before and after effects and to investigate whether the use of selective signs will have positive impacts on traffic operations.

Section 2: Rebuttal Letter

A response letter from FHWA was received in July 2019. Unfortunately, FHWA denied the request based on a number of considerations, including the general methodology and findings described in the referenced research report. As stated in the letter, "the methodology does not consider the limited amount of time available in which a driver would be able to view the sign while driving at highway speeds. The time durations stated for participant comprehension of the various graphical displays studied in this research are excessive. Human factors research studies on traffic control devices have determined that glances and attention away from the road longer than 2.0 seconds significantly increases the risk of a driver having a crash. It appears that the time required for an observer to process any of these graphical displays is far greater and would create an unacceptable safety risk to road users."

Excessive glance time and the related potential safety risks for traffic signs are valid concerns. However, the measured viewing and eye fixation measurements in the lab testing and the simulator testing are not the same as the actual driver glance time while driving on the roads. The PATH team and the Caltrans team developed the following rebuttal letter to the FHWA in response to the denial of the request.

- Firstly, the objectives of conducting the lab testing and the simulator testing were to explore drivers' understanding of alternative designs for each proposed message. The methodology that was used in the lab testing and simulator testing was to serve the purpose of evaluating the understandability of various designs. The viewing time observed in the lab testing and the simulator testing should not be directly applied to infer the viewing time that drivers will need in the real-world driving scenario.

In the lab testing, test subjects were encouraged to understand the overall meaning of each design as well as the detailed design elements (e.g., the meaning of symbols). Test subjects had full control of when to start and end the evaluation of each design. Hence, they don't have the type of time urgency while reading the signs in the lab in comparison with reading the signs on the road. The objective of the simulator testing was to further evaluate the understandability of the signs in the simulated driving environment and evaluate the proposed design elements based on the results of the lab testing. In the simulator testing, fixed viewing time was set for each category of message (e.g., 15 seconds for three-line messages and 20 seconds for GRIPs) with the purpose to provide sufficient, rather than limited, evaluation time for each test subject. Compared with the lab testing, the driving simulator testing offered higher experimental validity by providing a realistic representation of the road geometry and traffic conditions where the messages are displayed and requiring the test subjects to spend both physical and mental efforts onto the driving task while looking at the signs. The project team concurs that "there is a limited amount of time available in which a driver would be able to view the sign while driving at highway speed", and have proposed to evaluate drivers' processing of different messages under a limited amount of viewing time through the on-road testing.

In addition, since the kick-off stage of this research project, FHWA representative (Mr. Kevin Korth, Mr. Steve Pyburn) was part of the research team and did not express any concerns regarding the methodology design for either the lab testing or the simulator testing.

In short, with the objectives on the measured data to address the research questions of understandability and alternative designs, the applied methodology allows test subjects to comprehend alternative designs of various messages in a leisurely manner, which is much different in comparison with processing signs while driving on the road. Therefore, the viewing time observed in the lab testing and the simulator testing should not be directly applied to infer the time that drivers will need to process the messages in the real-world driving scenario.

- Secondly, the total accumulative viewing time in the lab or in the simulator testing is not the same as the glance time, to which the 2.0 seconds refers. Hence, the 2.0-second guideline is not applicable for the viewing time that was observed in the lab testing and deployed in the simulator testing.

The "2.0 seconds", which was mentioned in FHWA's reply, refers to the time for each glance rather than the total viewing time for each sign. Drivers on the road may take multiple glances at one sign. Results in the lab testing indicated that the average viewing time for the BART travel time messages was between 14.12 and 18.98 seconds. The average viewing time for the single-link GRIP messages was between 15.88 and 17.84 seconds. The above results were consistent with other studies that applied similar testing methodology. As one example, in a study sponsored by FHWA (Ullman et al., 2009), the self-paced average viewing time for a single work zone symbol is about 11.8 seconds. The average viewing time for a more complicated sign (e.g., HOV Lane Identification) ranges between 20.7-36.2 seconds. As another reference, the SAE J2365[11] has specified the standards that the maximum time for drivers to complete a navigation-related task involving both visual displays and manual controls in a moving vehicle should be less than 15 seconds. In neither reference, the 2.0-second guideline is applicable. Likewise, the 2.0-second guideline is not applicable for the viewing time that was observed in the lab testing and deployed in the simulator testing.

- Thirdly, there are significant learning or familiarity effects in viewing signs for travelers. As the IDBs are aimed to provide traveler information for local commuters, once the commuters get familiar with the design and information layout of the sign, it is anticipated that they will spend much less time looking for the information that is specifically related to their destinations.

There has been extensive relevant research that could support the learning effects on viewing signs for travelers. It was found that drivers' familiarity with the sign was significantly correlated with their response time as well as their comprehension accuracy (Ng & Chan, 2008; Shinar et al., 2003). In our study, the learning effect was also noticeable between the lab testing and the simulator testing, even though the viewing experience was limited. For the single-link GRIP, the understanding accuracy in the initial stage (the lab testing) ranges from 83% to 91%. In the simulator testing, many participants mentioned that it was much easier to understand the single-link GRIP. The understanding accuracy for single-link GRIP ranges from 97% to 100%. Similar to the learning effect that we have identified in the lab testing and simulator testing, as one reference, Alkim (2000) pointed out that "comprehension of the information presented is initially quicker for regular CMS. But drivers get used to GRIP rapidly and the difference is made up quickly".

- Ultimately, the objectives of implementing these new signs include reducing their reliance on personal communication devices for similar information, which leads to longer time off the road.

In addition, GRIPs including more complicated network-based GRIP and link-based GRIP as shown in Figure 33, have been deployed in other countries such as Japan, China, Australia, Germany, France, and United Kingdom et al. These deployment experiences in other countries illustrate the potential of using GRIP displays to provide traffic information. Hence, it will be beneficial to evaluate drivers' behaviors in response to these signs and gain further insights by conducting roadway testing as a continuation of this current study, with the selected signs that have been screened and revised based on the human-factor findings from the lab and simulator experiments.

[11] SAE J2365, Calculation and Measurement of the Time to Complete In-Vehicle Navigation and Route Guidance Tasks. Accessed from https://saemobilus.sae.org/content/j2365_201607

Figure 33. Network-based GRIP Used in Japan and Link-based GRIP Used in France

Goals for these displays include influencing drivers' behavior to consider other modes of transportation and to reduce reliance on personal communication devices for similar information. Please note that we are only asking to do testing, not permanent display, on the corridor. In addition, the messages will be displayed during congested periods when drivers will have repeated and extended opportunities to view the displays. Caltrans will utilize CMS and PCMS upstream to alert the drivers the IDB is under test. Caltrans will coordinate with CHP and other local agency partners and involve public affairs teams as needed. On-road testing was not accomplished as FHWA denied the field-testing request from Caltrans.

References

[1] Alkim, T. P. (2000). Graphical route information on variable message signs. Tenth International Conference on Road Transport Information and Control, 2000, 32–36. https://doi.org/10/b6kqxp

[2] Ng, A. W. Y., & Chan, A. H. S. (2008). The effects of driver factors and sign design features on the comprehensibility of traffic signs. Journal of Safety Research, 39(3), 321–328. https://doi.org/10/fh85jx

[3] Shinar, D., Dewar, R. E., Summala, H., & Zakowska, L. (2003). Traffic sign symbol comprehension: A cross-cultural study. Ergonomics, 46(15), 1549–1565. https://doi.org/10/c2m6n5

[4] Ullman, B. R., Trout, N. D., Dudek, C. L., Institute, T. T., of Transportation, T. D., & Administration, F. H. (2009). Use of graphics and symbols on dynamic message signs: Technical report (No. 2; Vol. 7, p. 192p). https://trid.trb.org/view/890543

Part V: Research Topics for Future Studies

Introduction

The planned field testing and focus group discussions could not be implemented as the result of FHWA's denial of the request for permission to experiment. The PATH team suggested doing another lab testing in order to address FHWA's comments. The lab testing was proposed to use fixed display time (e.g., 6 seconds, or 8 seconds) controlled by the testing program rather than free viewing time controlled by the participants. Besides, the revised lab testing was planned to test the same signs in multiple times in order to prove the learning effect. The PATH team has developed the testing program to control the display time for the revised lab testing. However, the testing could not be carried out due to the COVID-19 outbreak, and the following statewide working from home order started in middle March of 2020. Besides the revised lab testing, the PATH team took a further look at other potential issues and concerns that were not fully answered for the design of IDB signs and suggested the following two topics for future studies beyond this project.

Legibility Distance for Symbols, Texts, and Numbers

According to the guideline of legibility distance in CAMUTCD[12], messages should be designed to be legible from a minimum distance of 600 feet for nighttime conditions and 800 feet for normal daylight conditions. When the legibility distance could not be achieved, messages composed of fewer units of information should be used. The IDBs have a form factor that differs from typical Model 500 CMS, with the visible optical area of 157 1/2" x 196 7/8", with a resolution of 216 x 270 pixels full matrix. The IDBs are capable of providing better legibility and enhanced sign recognition. However, there is a lack of fundamental research to specify the legibility distance of both symbols and texts displayed on the IDBs.

On the one hand, in order to design legible IDB signs, the Caltrans D4 team designed the signs with different characteristics (e.g., different font sizes and spacing) and tested the legibility distance for each design on the freeway. On the other hand, the symbol is one very important element for the IDB. As there are no symbols on traditional CMS, there are no guidelines regarding legibility distance for symbols for CMS. Hence, there is no guideline to follow for the legibility distance of symbols on the IDB. What is more, considering the different combinations of background colors and symbol colors, the legibility distance for symbols of the same size could be complicated. Therefore, we propose a study to have a comprehensive evaluation of the legibility distance for symbols and texts in different sizes and colors displayed on the IDB. The study could be done in Richmond Field Station by setting up an IDB on a smaller scale in comparison with the actual IDB in the field but allowing for the same visual effect. As a summary, we propose to test the legibility distance of the IDB, considering the following design factors:

- Texts with different letter sizes, different spacing between text lines;
- Numbers with different font sizes, different spacing between letters;
- Symbols with different sizes and different color combinations;
- Daytime and nighttime lighting conditions;
- Under different driving speeds.

[12] California MUTCD 2014 Edition, Section 2L.03 Legibility and Visibility of Changeable Message Signs

Through the field testing in Richmond Field Station, we could gain an understanding of font sizes for letters and numbers, spacing between lines, range of sizes for symbols with different color combinations that will meet the guideline of legibility distance from CAMUTCD. This understanding can be served as a guideline for designing new IDB signs in the future.

Message and Display Design for Public Transportation Messages

There are two categories of messages about the provision of information for public transportation. One is the public transportation related special messages. The other one is the transit travel time messages. These public transportation messages are designed to provide information on alternative transportation modes in place of driving. Based on the findings from the lab testing in the IDB study, many participants mentioned that they didn't understand why those messages needed to be seen while they were driving on the freeway. Besides, in order for travelers to make the decision of taking public transportation rather than driving, extra information such as parking availability, bus or BART arrival time, travel time to the destinations, and travel time from the transit station to their final destinations would be desirable and helpful for users.

Hence, in order to make the public transportation message useful for travelers, firstly, it is critical to understand the following several questions:

- What are the scenarios that drivers drive to the terminal and then take public transportation?
- Will travelers drive to a bus station and then take a bus? Or, will most likely travelers drive to the BART station only?
- What is the necessary information that will be needed for the travelers to make the decision between taking the BART and driving?

Either interview or focus group discussion could be conducted in order to answer the above questions. Travelers who drive to the bus terminals or BART stations and take the bus or BART could be recruited to participate in the interview or focus group discussion. More detailed questions could be asked about their (1) behavior of driving from home to the bus terminals or BART stations; (2) experience while transferring at the bus terminals or BART stations; (3) information that they will need for planning purposes (e.g., change their originating station); and (4) other specific experience that they had regarding taking the public transportation.

After figuring out the above questions, if more information is needed according to the feedback from most of the participants, the next step would be to investigate how to present all the necessary information on one IDB sign. We propose designs of the IDB signs considering using (1) transit logos; (2) signs with two virtual phases; (3) signs with two actual phases (e.g., one to show the travel time, the other one to show the parking information). After having the proposed designs, either a lab testing or a simulator testing could be used to further test the understandability of the messages.