SUPPORTING JUSTIFICATION – Part B
Head-Up Display Study; OMB No. 2130-NEW
Form Number FRA F 6180.158

1. Description of sampling method to be used.

FRA will recruit approximately 40 railroad engineers for the experiment.

A convenience sampling method will be utilized. The population of interest is railroad engineers. They will be recruited from the Brotherhood of Locomotive Engineers and Trainmen (BLET) of those engineers located in and around the Boston, MA geographical area. FRA is working with rail labor unions on this data collection effort.

2. Description of procedures for information collection, including statistical methodology for stratification and sample selection.

2.1 Statistical Methodology:

The purpose is to explain the statistical methodology and, in particular, to explain the difference between conducting an experimental test of a hypothesis (what we are planning to do) and trying to generalize the results from a survey. It is important to discriminate between these two purposes because they lead to very different estimates of the sample size. (a) The number of subjects needed for testing an hypothesis (e.g., the means of two groups of engineers are equal) depends on the predicted effect size, probability of a Type 1 error, and probability of a Type II error. (b) The number of subjects needed for generalizing the results from a survey depends on the size of the population, the level of confidence, and the precision.

For the evaluating the hypothesis that we are testing, we need about 40 subjects, for generalizing the results from a survey we need about 400 subjects. In the cover letter, we have discussed in detail the actual computations required to obtain the sample size for the test of our hypothesis. Below, we point to the use of a calculator available on the internet which can be used to determine the sample size for the test of a hypothesis. This is the sample size we need. We also point to an online sample size calculator for computation the sample size to obtain a precise estimate of a population statistic.

a) Experiments: Testing the Difference in Proportions
We estimate that we will need 19 participants to test the hypothesis that the performance of the HUD and radio groups differ. We assume that the probability of a Type 1 error is 0.05, that the probability of a Type II error is 0.20 (power is equal to 0.80), and that the radio group will detect only 30% of the safety-critical events whereas the HUD group will detect 70% of the safety-

critical events. Again, the calculations are in the cover letter. It is also possible to use the online calculator below.

https://www.stat.ubc.ca/~rollin/stats/ssize/b2.html

Inference for Proportions: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values (as decimal fractions) for p1 (proportion in population 1) and p2 (proportion in population 2) and, if calculating power, a sample size (assumed the same for each sample). You may also modify α (type I error rate) and the power, if relevant. After making your entries, hit the **calculate** button at the bottom.

- ◉ Calculate Sample Size (for specified Power)
- ○ Calculate Power (for specified Sample Size)

Enter a value for p1: .7

Enter a value for p2: .3

- ◉ 1 Sided Test
- ○ 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size (for each sample separately) is: 19

b) Surveys: Precise Estimates of Proportions

Below, we show the results from an online calculator that computes the sample size for surveys. Note that now we need to know the population size. We did not need to know this for the experimental test of the hypotheses (above). With a population of 36,000, a confidence interval of +/- 5, and a confidence level of 95%, we would need roughly 400 participants. <u>We are not doing this type of research</u>.

https://www.surveysystem.com/sscalc.htm#one

Sample Size Calculator

This Sample Size Calculator is presented as a public service of Creative Research Systems survey software. You can use it to determine how many people you need to interview in order to get results that reflect the target population as precisely as needed. You can also find the level of precision you have in an existing sample.

Before using the sample size calculator, there are two terms that you need to know. These are: **confidence interval** and **confidence level**. If you are not familiar with these terms, click here. To learn more about the factors that affect the size of confidence intervals, click here.

Enter your choices in a calculator below to find the sample size you need or the confidence interval you have. Leave the Population box blank, if the population is very large or unknown.

Determine Sample Size

Confidence Level: ◉ 95% ○ 99%

Confidence Interval: 5

Population: 36000

Calculate Clear

Sample size needed: 380

Find Confidence Interval

Confidence Level: ◉ 95% ○ 99%

Sample Size:

Population:

Percentage: 50

2.2 Procedures (All time calculations have been included in the burden estimates):

1. Each participant will arrive at the Volpe Center. Prior to commencement of any study procedures, the participant will read and sign a consent form, and then be assigned to an experimental condition.

2. Once the participants are familiar with the technologies and understand all of the job briefing, they will do a complete practice (training) run of the territory on the train simulator (CTIL). The practice run lasts approximately 40 minutes. This practice run will allow the participants to become familiar with the territory and the simulator controls, and the engineer will be operating on all clear signals with no radio communications.

3. After the practice run, participants will be allowed a 15-minute break, during which they can stretch, consume refreshments, and use the restroom facilities. When participants return, they will complete their first experimental run on the train simulator. The experimental run will last approximately 60 minutes. Once the first experimental run is complete, participants will be allowed a 15-minute break again. The second experimental run is over the same territory as the first experimental run. Upon completion of all three runs (training, Experimental 1, Experimental 2), participants will be allowed a final break.

4. After the simulator portion of the experiment, participants will fill out a brief demographic, usefulness and usability questionnaires. This will take approximately 20 minutes. The questionnaire following the simulator experiment will help us gather preliminary information on the usability of the HUD. This may be helpful in future experiments. The answers to this questionnaire are not expected to be generalizable to the population of locomotive engineers, and are solely to provide information to the research team that may inform future experimental technology design.

Detailed Description of Simulation:

GLANCES DURING RADIO COMMUNICATIONS

The simulator allows for meticulous control of track locations where events occur during the experiment and when messages are sent. This enables us to control presentation of the radio communications and safety critical events during the simulation. Therefore, since we control both where a safety critical event occurs on the tracks and when a message is presented (either spoken by the dispatcher or displayed as a yellow icon on the HUD), we have arranged it so that safety critical events occur as many times while an engineer is listening to a radio message (in the Radio condition) and as they do when a yellow icon is displayed on the windshield (in the HUD condition).

With this level of control, we believe that there are no confounds present when we compare the engineers' ability to detect a safety critical event in the HUD and Radio conditions during the presentation of a non-urgent message and that there are no confounds present when we compare the engineers' glances away from the forward track in the HUD and Radio conditions when reading back the non-urgent message. If we find that the HUD performs better in both cases, then we believe we have a solid comparison of the HUD with the Radio conditions when non-urgent messages are presented and when they are read back.

A critical piece of the simulation protocol is the behavior of the engineer listening to a radio message and then reading it back. One might initially think that an engineer listening to a dispatcher's message could keep his or her eyes focused on the forward track throughout the communication of that message. However, based on our studies here on the CTIL with actual locomotive engineers, the engineers need to glance away from the forward track for three reasons in the process of receiving and reading back a radio message. First, when the non-urgent message is first delivered it will contain content which the engineer frequently needs to write down on paper inside the cab. For example, the engineer might be told to report broken track at milepost 32.4. Second, when the engineer confirms the message by reading it back to the dispatcher he or she needs to look inside to press the talk button. And third, the engineer will often continue to look inside the cab to read back the exact milepost and any other information which was included in the original communication and which he or she wrote down.

DETECTING SAFETY CRITICAL EVENTS(SCE): HUD AND RADIO CONDITIONS

The above discussion makes it clear why the engineer should be more likely to detect a safety critical event (SCE) in the HUD condition than in the Radio condition. In brief, in the Radio condition, the engineer is both listening to the message and looking inside the cab to write down the details of the message when the safety critical event occurs. Looking inside the car to write down the content of the message takes the engineer's eyes away from the safety critical event occurring outside. In the HUD condition, a yellow icon will appear on the HUD. The engineer can therefore detect the safety critical event since the yellow icon does not compete for his or her visual attention (which remains on the forward track).

The exact timeline for the events in the HUD and Radio conditions are displayed below in Table 1. In the Radio condition the dispatcher will let the engineer know that a message is about to be communicated (start of initial radio communication). This initial communication is only to establish that the dispatcher has something to relay. There is no message at this point, but the engineer can prepare to listen to a message. The actual message continues 2.5 seconds before the SCE and continues 2.5 seconds after the SCE (each message is approximately 5 seconds long). The engineer will typically start glancing inside the cab to write down the content of the message. Readback will take place after the message has been delivered. In the HUD condition, the only indication that a non-urgent message has arrived is the display of a yellow icon on the windshield.

Time relative to onset of Safety Critical Event (SCE)	**Event (Radio condition)**	**Event (HUD condition)**
12.5 seconds before SCE	Start of initial communication	(none)
2.5 seconds before SCE	Start of Radio Message Content	Yellow Light
>2.5 seconds before SCE	Engineer listening to content (writing down message)	Content available for reading (engineer's decision)
SCE	**Engineer listening to content (writing down message)**	**Content available for reading (engineer postpones)**
>2.5 second after SCE	Engineer readback of content.	Content available for reading (engineer's decision)

Table 1. Timeline of events in Radio and HUD conditions

READING BACK MESSAGES: HUD AND RADIO CONDITIONS

Above we noted that the safety critical event always occurs while the non-urgent message notification icon is present (also see Table 1). It is true that the safety critical event never occurs while the engineer is reading a non-urgent message in the HUD condition. But no safety critical event occurs while the engineer is reading back a non-urgent message in the Radio condition. We will compare glance behaviors with the use of an eye-tracker device in both conditions.

3. <u>Description of methods to maximize response rate and to deal with non-response issues</u>.

Compensation for each participant's time and travel will be given as an incentive to participate in the experiment.

4. <u>Describe any tests for procedures or methods to be undertaken</u>.

A pilot study with 2 or 3 engineers will be conducted to refine the data collection procedures and instruments. Because this pilot is designed solely to test the study methods and not for analysis of the data, we will select the pilot participants. This pilot study will include a complete experimental run on the simulator. The engineers will then provide the research team with informal feedback regarding the experimental design.

Limitations

This is primarily an exploratory study of the feasibility of innovative technologies for use in a train cab that may mitigate radio distraction. Given this, the work includes several limitations that may affect the generalizability of the results. First, this is a convenience sample utilizing forty (40) locomotive engineers from the Boston, MA area, therefore, the results cannot be necessarily generalized to the entire population of locomotive engineers across the United States. Also, this study will be conducted in a simulator and hence the results cannot necessarily be directly generalized to the railroad operating community without further study. Future studies would need to be conducted to further investigate the new technologies we propose to study, and to further investigate if HUDs and/or digital transmission of communications may be useful in mitigating radio distraction. FRA will not be making recommendations as a result of this study and FRA will be publishing the study independent of whatever results are obtained.

5. Provide name and phone number of individuals consulted on statistical aspects of study design and other persons who will collect/analyze information for agency.

Donald Fisher, Ph.D. (617) 494-3418. Volpe (Collect and Analyze)
Rachel Grice. (202) 493-8005. Federal Railroad Administration (Collect and Analyze)
Matthew Isaacs. (617) 494-2474. Volpe (Collect and Analyze)