Research Article

Addressing the Safety of Transportation Cyber-Physical Systems: Development and Validation of a Verbal Warning Utility Scale for Intelligent Transportation Systems

Yiqi Zhang,[1] Changxu Wu,[1] Chunming Qiao,[2] Adel Sadek,[3] and Kevin F. Hulme[4]

[1]*Department of Industrial & Systems Engineering, The State University of New York (SUNY), Buffalo, NY 14260, USA*
[2]*Department of Computer Science Engineering, The State University of New York (SUNY), Buffalo, NY 14260, USA*
[3]*Department of Civil, Structural and Environmental Engineering, The State University of New York (SUNY), Buffalo, NY 14260, USA*
[4]*NYs Center for Engineering Design & Industrial Innovation, The State University of New York (SUNY), Buffalo, NY 14260, USA*

Correspondence should be addressed to Changxu Wu; changxu.buffalo@gmail.com

As an important application of Cyber-Physical Systems (CPS), advances in intelligent transportation systems (ITS) improve driving safety by informing drivers of hazards with warnings in advance. The evaluation of the warning effectiveness is an important issue in facilitating communication of ITS. The goal of the present study was to develop a scale to evaluate the warning utility, namely, the effectiveness of a warning in preventing accidents in general. A driving simulator study was conducted to validate the Verbal Warning Utility Scale (VWUS) in a simulated driving environment. The reliability analysis indicated a good split-half reliability for the VWUS with a Spearman-Brown Coefficient of 0.873. The predictive validity of VWUS in measuring the effectiveness of the verbal warnings was verified by the significant prediction of safety benefits indicated by variables, including reduced kinetic energy and collision rate. Compared to conducting experimental studies, this scale provides a simpler way to evaluate overall utility of verbal warnings in communicating associated hazards in intelligent transportation systems. This scale can be further applied to improve the design of warnings of ITS in order to improve transportation safety. The applications of the scale in nonverbal warning situations and limitations of the current scale are also discussed.

1. Introduction

Warning plays an important role in the communication of the information with regard to potential hazards to avoid accidents and injuries. There has been a significant increase in research on communication of the road safety in the last three decades [1]. In order to improve driving safety, recent advances in Transportation Cyber-Physical Systems (CPS) aim to establish a connected transportation environment connecting cyber world (e.g., information, communication, and intelligence) and physical world (e.g., sensors and actuators) and provide the integrated real-time information among multiple levels, including vehicles to vehicle communication, vehicle to infrastructures communication, and in-vehicle information communication [2]. Compared to traditional transportation environment, the connectivity of the CPS

allows drivers to learn about the traffic status out of their sight and provides them with more time to respond to warnings regarding potential hazards [3–5]. As an important application of Transportation CPS, most of the intelligent transportation systems (ITS) researches proposed algorithms to schedule warnings based on warning utility. With various warnings provided by the ITS, it is increasingly important to evaluate the warning utilities in the design of Transportation CPS.

Warning utility refers to the degree to which a warning enhances user performance through its presence. Shackel [6] defined the utility of system as whether the system does what is needed functionally. The definition of utility is similar to the definition of system effectiveness defined in Regan et al.'s work [7]. Regan et al. [7] clarified the definitions of effectiveness, usefulness, and ease of use. In their work, effectiveness

was defined as "the question of whether the system works in accordance with its functional description"; perceived usefulness was defined as "the degree to which a person believes that using a particular system will enhance his/her performance"; and ease of use was defined as "the degree to which a person believes that using a particular system would be free of effort." As it was discussed in [8], acceptance was defined with two dimensions: usefulness and satisfaction. However, further studies suggested that acceptance also contains dimensions such as effectiveness and social influence [7]. Adell [9] categorized different definitions of acceptance of system and defined it as "the degree to which an individual intends to use a system and, when available, incorporates the system in his/her driving." Previous studies have indicated that subjective assessment of perceived usefulness did not always reflect enhancement on driving performance when using the system [10, 11]. Specifically, a too late or too early warning may still be perceived as being useful for drivers, even without significantly improving driver safety. On the basis of such difference, our previous work showed warning utility may serve as a better construct to assess expected objective benefits [12]. Compared to the definitions of usefulness and acceptability, the definition of utility focuses on user performance rather than user attitude.

To date, much work has been done to subjectively evaluate warnings regarding different constructs including acceptance, usefulness, and ease of use. Davis [13] proposed the Technology Acceptance Model and developed a scale to evaluate acceptance with two dimensions: perceived usefulness and ease of use. Segars and Grover [14] examined this scale with a factor analysis and found there was an additional factor "effectiveness" besides perceived usefulness and ease of use. Van Der Laan et al. [8] developed a scale to measure acceptance with two dimensions: usefulness and satisfaction. This scale has been applied to estimate the usefulness and satisfaction of the system for different systems. Based on the Unified Theory of Acceptance and Use of Technology (UTAUT), Adell [9] developed a questionnaire to measure acceptance with the following dimensions: performance expectancy, effort expectancy, social influence, and behavioural intention to use the system. In terms of subjective assessment of usefulness, there were studies using the word "usefulness" directly to measure system usefulness [15, 16]. As the above subjective assessment tools focused on evaluating the users' attitude to systems, the warning utility with a focus on expected objective benefit (e.g., reduction in accident risk) was seldom measured by scales. The behavioural approach to assess the utility of the warning systems can be highly task-depended and, in the meantime, time consuming and of high cost. The current study will address this problem by developing a scale to evaluate the verbal warning utility with regard to improving the effectiveness of communication in ITS.

The warnings mainly focused on in this work are auditory warnings which have been widely applied to many warning systems nowadays. Compared to visual warnings, auditory warnings have an advantage over visual warnings that human hearing cannot be shut off in the way human vision can. Studies have demonstrated that auditory warnings have higher levels of compliance than visual warnings (e.g., Wogalter and Young [17]). Auditory warnings are able to attract the human's attention regardless of where their attention is directed and when people are working under conditions with high workload, especially high visual workload, and/or when the operator has to move about a lot or visual conditions are bad.

Among auditory warnings, the current development of the scale mainly focused on verbal auditory warnings, which are user friendly since the users may easily understand and differentiate them without specific trainings. Compared to verbal auditory warnings, nonverbal auditory warnings have the drawback that their meanings need to be learned, remembered, and recognized at the time that they sound [18]. Research has indicated that humans are quite bad at remembering and recognizing nonverbal auditory warnings. For example, one study showed that people working in an operating room and recovery room in a teaching hospital were unable to recognize more than half of the alarms currently in use [19]. Patterson's study shows that people are able to learn the first few of a set of auditory warnings more or less as fast as they can be presented, but that progress slows beyond six or seven warnings [20]. Moreover, identification of nonverbal auditory warnings in operational situations is likely to be worse than in the laboratory setting since recognition on an absolute basis is likely to be lower than on a relative basis [21]. A previous work found that verbal warnings leaded to a faster reaction time than nonverbal warnings did [22], especially for complicated road conditions [23]. Studies also suggested the verbal warning led to smaller crash rate than nonverbal warnings in intersection collision warning systems [24], especially for older drivers [25]. There are disparate recommendations of the application of verbal warnings in critical safety situations. The current ISO working draft [26] suggested verbal warnings should not be used for safety critical warnings since such warnings took more time to present, whereas Noyes et al. [27] suggested verbal warnings should be used in safety critical situations since drivers responded more quickly and have more chance to react accurately to verbal warnings than nonverbal warnings. Generally speaking, verbal auditory warnings can be widely applied without additional training, which makes it especially suitable for transportation assistant systems.

The purpose of the present research was to develop a scale to evaluate the utility (effectiveness) of verbal warnings in order to achieve effective communication in intelligent transportation systems. The evaluation of the warnings utility is mainly based on two aspects: how well it attracts human attention and how well it provides understandable information needed for audiences [1, 28]. By summarizing items that may influence establishing warning utility, the dimensions of Verbal Warning Utility Scale (VWUS) will be selected through a subjective evaluation. After the scale was developed, the reliability of the scale was tested based on the split-half reliability analysis and a factor analysis was conducted to explore the structure of the scale. An experimental study was then conducted to test the predictive validity of the scale in a simulated environment. If the verbal warning utility score can be used to predict safety benefits brought by the designed warnings, it is suggesting that the VWUS can be applied to evaluate the warning effectiveness in improving driver safety.

The validated scale can be utilized to assess the utility of verbal warnings in the intelligent transportation systems instead of performing behavioral experiments.

2. Development of the Scale

To develop the verbal warning utility scale, we referred to the development process of the NASA Task Load Index (NASA-TLX), which is also a multidimensional scale [29]. The steps in developing the NASA-TLX include exploring relevant factors to the concept being tested (i.e., workload), conducting subjective rating to select salient factors with a certain criteria, and performing experimental studies to validate the scale.

2.1. Method

2.1.1. Participants. One hundred and four students (72 males and 32 females) from the University at Buffalo finished the exploratory questionnaire. Participants were recruited through the flyers on campus. All had a valid driver's license and were English native speakers. Participants' ages ranged from 18 to 69 years with an average age of 29.66 years (SD = 12.09). In terms of driving experience, the average years since obtaining a valid US driver's license were 11.11 years (SD = 11.47), while the average mileage was 9355.77 miles (SD = 7037.41).

2.1.2. Material. Since many warning characteristics may combine to creating a subjective warning utility, we reviewed warning characteristics that may influence the effectiveness of warnings. To our best knowledge, we find that the following warning characteristics could influence the warning effectiveness: Category one is regarding warning effectiveness in representing the hazard/task: Representation of Event Urgency [30–32], the degree to which the event urgency is represented by a warning; Time to Display [33], the degree to which a warning occurs at a favourable time; Number of Replications [34]; Distinctive Features [35], the degree to which a warning can be differentiated from other warnings in the system. Category two is regarding features relevant to the warning itself: Modality [36], the way a warning is presented; Loudness [37], the attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud; Length of the Warning, the number of words in a warning; Voice Type, whether a warning is presented with a female or male voice [38]; Voice Quality [34], the degree to which a warning can be recognized; Rate of Speech, the number of words presented in terms of time. Category three is regarding the human cognitive process of the verbal warnings: Alertness [30], the degree to which the user's attention is raised; Loss of Vigilance, the degree to which the ability to maintain attention is impaired over prolonged periods of time; Accuracy of Understanding [39], the degree to which an individual understands the hazard with the presence of a warning; Familiarity [40], the degree to which an individual is being familiar of a warning; Annoyance [30], the degree to which a warning annoys an individual; Likeness/Preference, the degree to which an individual prefers the presence of
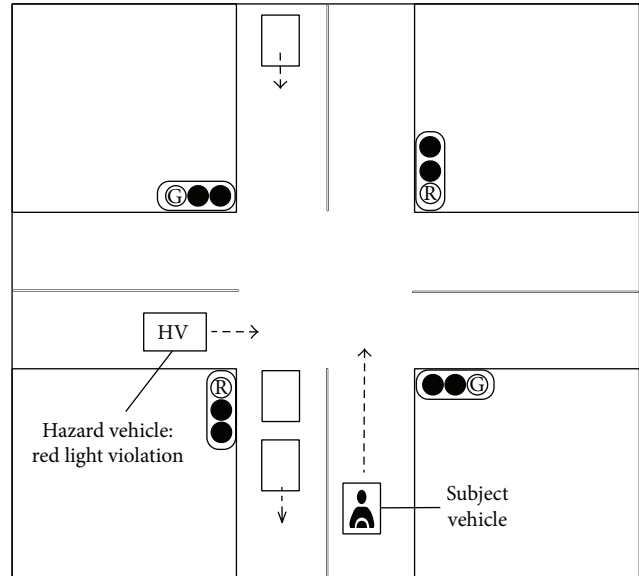

Figure 1: The example of the test scenario.

a warning; Comfort: the degree to which an individual feels comfortable in responding to a warning; Trust [41], the degree to which an individual trusts a warning to be reliable; Reduced Workload [37], the degree to which the workload is reduced with a presence of a warning; Forgetting [34], the probability of a warning being forgotten by an individual after its presence.

A questionnaire was used to explore salient items to be subjectively equivalent to warning utility. The questionnaire listed the name of each item along with its explanation and asked the participants to identify its relationship with the overall verbal warning utility (i.e., "a": subjectively equivalent, "b": related, and "c": unrelated). An item being subjectively equivalent to the overall verbal warning utility indicates that this item could represent the utility of verbal warnings. An item being related to the overall verbal warning utility indicates that this item contributes to the utility of verbal warnings. An item being unrelated to the overall verbal warning utility indicates that this item does not contribute to the utility of verbal warnings. An example item is "Representation of event urgency: How urgent is the event conveyed by the warning?"

2.1.3. Procedure. The study was conducted at the University at Buffalo. Participants were asked to complete a questionnaire regarding demographic information (e.g., gender, age, and driving history) and the exploration questionnaire then was administrated. The participants were asked to imagine the scenario in which a warning was provided to describe the coming dangerous events and help them to decrease the probability of collision. As it is shown in Figure 1, an example was presented that an in-vehicle warning system alarms the driver on the subject vehicle about a hazard caused by a hazard vehicle running the red light from its left.

The warning is, for instance, "a vehicle at your front-left is running the red light at the intersection." The questionnaire

Table 1: Mean score of each factor and frequency of the three kinds of relationships between each factor and the warning utility ($N = 104$).

Items	Representation of Event Urgency	Time to Display	Number of Replications	Distinctive Features	Warning modality
Equivalent	59	63	25	45	38
Related	38	33	57	46	40
Unrelated	7	8	22	13	26
χ^2	25.62***	34.74***	4.04	4.62	0.48
Items	Loudness	Length of the Warning	Voice Type	Voice Quality	Rate of speaking
Equivalent	43	37	14	34	38
Related	53	50	37	51	57
Unrelated	8	17	53	19	9
χ^2	3.01	0.24	18.48***	0.02	0.48
Items	Alertness	Loss of Vigilance	Accuracy of Understanding	Familiarity	Annoyance
Equivalent	62	31	52	24	21
Related	31	61	41	64	48
Unrelated	11	12	11	16	35
χ^2	32.33***	0.58	13.00***	4.92	8.08*
Items	Likeness	Comfort	Trust	Workload	Forgetting
Equivalent	19	23	41	25	32
Related	40	50	44	43	39
Unrelated	45	31	19	36	33
χ^2	10.62*	5.59	1.74	4.04	0.31

Note: *$p < .017$, ***$p < .001$.
The p value of .017 was applied with Bonferroni correction for p value of .05.

listed the name of each item along with its explanation and asked the participants to identify its relationship with the overall verbal warning utility (i.e., "a": subjectively equivalent, "b": related, and "c": unrelated). An item being subjectively equivalent to the overall verbal warning utility indicates that this item could represent the utility of verbal warnings. An item being related to the overall verbal warning utility indicates that this item contributes to the utility of verbal warnings. An item being unrelated to the overall verbal warning utility indicates that this item does not contribute to the utility of verbal warnings. An example item is "Representation of event urgency: How urgent is the event conveyed by the warning?"

2.2. Results. Three relationship options, including subjectively equivalent, related and unrelated to verbal warning utility, were recorded. The frequency of each relationship for all twenty items was shown in Table 1. The chi-square goodness-of-fit test was conducted for each factor to test whether the frequency of subjectively equivalent is significantly more than the frequency of the other two relationships. Applying the Bonferroni correction [42], $p = .05$, was divided by the number of tests (3) to get the Bonferroni critical value, so a test would have to have $p < .017$ to be significant. Under that criterion, only the tests for Representation of Event Urgency, Time to Display, Alertness, and Accuracy of Understanding showed significantly higher proportion of equivalency to warning utility than being related or unrelated to warning utility. In the meantime, these four items (i.e., Representation of Event Urgency, Time to Display, Alertness, and Accuracy of Understanding) were considered to be

subjectively equivalent to warning utility by more than 50% (frequency of equivalent ≥ 52) [43].

Finally, combining the above analysis results, these four most salient items were then confirmed to consist of Verbal Warning Utility Scale (VWUS). The definitions of the four items were presented as follows.

(i) Representation of event urgency (item 1): how well is the event urgency conveyed by the warning?

(ii) Time to display (item 2): whether it is an appropriate time to display the warning to users?

(iii) Alertness (item 11): how well does the warning alert users to the dangerous event when they received the warning?

(iv) Accuracy of understanding (item 13): how well does the warning let users understand the event that just happened?

3. Validation Study

An empirical study was conducted to test the correlation between VWUS scores and behavior measurements in avoiding potential collisions in a simulated driving task, which would lend support to the predictive validity of VWUS. It was hypothesized that scores of VWUS developed in the present research would significantly predict the human performance in avoiding a potential accident with auditory warnings broadcasted.

Warning lead time plays an important role in determining the effectiveness of warnings in Transportation Cyber-Physical Systems [44, 45]. The Society of Automotive Engineers (SAE) proposed the four levels of time urgency to define priority order index. Emergency level was defined with lead time ranging from 0 to 3 s, immediate level was defined with lead time ranging from 3 to 10 s, near term level was defined with lead time ranging from 10 to 20 s, and preparatory level was defined with lead time ranging from 20 to 120 s [46]. Previous empirical study focuses more on the emergency and immediate level of warning lead time. Yan et al. [47] studied verbal warnings with the 7 levels of warning lead time (i.e., 2.5 s, 3 s, 3.5 s, 4 s, 4.5 s, 5 s, and 5.5 s). Results suggested early warnings led to a shorter brake reaction time, smaller crash rates, and smaller deceleration, indicating more timely and gradual responses to avoid the hazards compared with late warnings. Yan et al. [48] explored the impact warning lead time (3 s versus 5 s) and warning content (with or without direction information) on driving performance. Zhang et al. [49] also investigated the impact of directional and nondirectional verbal warning on driving behaviors at warning timing (lead time to collision) of 7 s. In order to measure the verbal warning utility completely, a wider range of lead times was applied to measure warning utility, including extremely short and long lead times.

The behavioral measurements quantifying the effectiveness of verbal warnings in the current study included the collision and the impact reduction. The "collision" was coded as a dichotomous variable specifying whether there was a collision between a subject's vehicle and hazard vehicle (get collided: 1 and avoid collision: 0). The reduced kinetic energy of the subject's vehicle specified the impact reduction led by the warnings. Because the mass of the vehicle can vary in reality, we will study the reduced kinetic energy of a vehicle with unit mass.

3.1. Method

3.1.1. Participants. Thirty-two participants (24 males and 8 females) took part in the current experiment with ages ranging from 18 to 26 years (mean = 21.13, SD = 2.54). In terms of driving experience, the average years since obtaining a valid US driver's license were 3.5 years (SD = 2.36) and the average mileage was 8343.75 miles (SD = 6438.0). All participants recruited had normal or corrected-to-normal vision and valid driver's licenses and are asked to sign an informed consent form before the experiment. Each driver was paid $20 for the time taken to complete the experiment.

3.1.2. Apparatus. The driving task was completed using a STISIM driving simulator (STISIMDRIVE M100K, Systems Technology Inc., Hawthorne, CA). The driving simulator consists of a Logitech Momo steering wheel with force feedback, a gas, and a brake pedal (Longitech Inc., Fremont, CA). The resting position of the throttle pedal is 38.2° (the angle between the pedal surface and the ground) and the maximal throttle input is 15.2°. For the brake pedal, the resting position is 60.1° and the maximal brake input is 28.6°. The STISIM simulator was installed on a Dell Workstation (Precision 490, Dual Core Intel Xeon Processor 5130 2 GHz) with a 256 MB PCIe × 16 nVidia graphics card, Sound Blaster X-Fi system, and Dell A225 Stereo System. The driving scenario was presented on a 27-inch LCD with 1920 × 1200 pixels resolution.

3.1.3. Material. The Verbal Warning Utility Scale (VWUS) consisted of a rating scale and a weighting scale (see Appendix 1 in the Supplementary Materials available online at http://dx.doi.org/10.1155/2015/126947). The rating scale comprised of the ratings of four items, "Representation of Event Urgency, Time to Display, Alertness, and Accuracy of Understanding" and a question regarding the overall utility of the warning. Participants were firstly asked how useful the warning heard in the experiment was in helping them to avoid the hazard by rating each factor on a ten-point scale (e.g., "0": not well at all to "9": extremely well). An example item is "How well did the warning represent the event urgency?" At the end of the rating scale, the participants were also asked about the overall utility of the warning using the item, "In general, how useful was the warning?"

In order to get more accurate ratings, a weighting scale was introduced. The weights obtained from weighting scales account for two potential sources of variability in the rating of verbal warning utility. It includes the differences in definition of verbal warning utility between subjects for the same warning and the differences in the sources of verbal warning utility between warnings. After finishing the rating scales, subjects were then asked to compare each pair of properties of the warning they just rated and select the one that contributed the most to the warning utility using the weighting scale. By comparing each dimension with the other three dimensions, each dimension would get three weighted scores ("0": contributes less to the warning utility and "1": contributes most to the warning utility score). The weighted score for each dimension was calculated by taking the average of the three scores. The total score of warning utility was calculated by adding conducts of rating score and weighted score of each dimension together.

3.1.4. Experiment Design. The current experiment adopted a single-factor within-subject design with lead time as the independent variable. Sixteen levels of the lead time were designed in the experiment (0 s, 1 s, 1.5 s, 2 s, 2.5 s, 3 s, 3.5 s, 4 s, 4.5 s, 5 s, 6 s, 8 s, 10 s, 15 s, 30 s, and 60 s). The lead time indicated the time to collision if the vehicles driven by the participant continued to travel at their current relative position, velocity, and acceleration. The reason why we design such a wide range for lead time is that the development of intelligent transportation system nowadays has enabled the drivers to acquire information of the potential hazard through warnings in advance.

Each participant would go through sixteen trials assigned with sixteen levels of lead time, each of which involves a hazard leading to potential collision and its corresponding warning. The orders of 16 lead times and 16 collision events were randomized. As it is shown in Figure 2, sixteen collision scenarios along with their warning were designed and programed with the driving simulator to represent the potential collision events in reality.
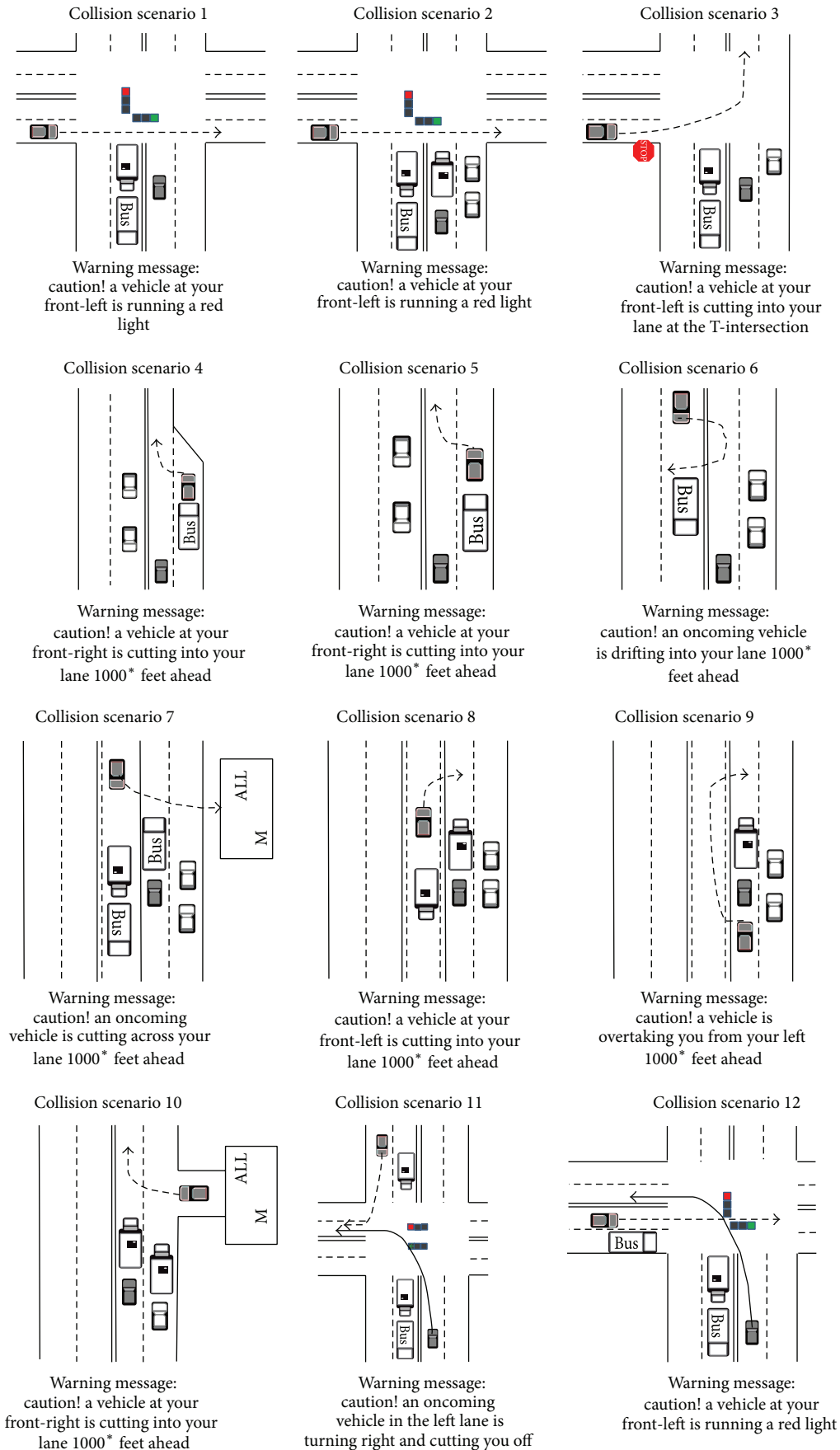
Collision scenario 1
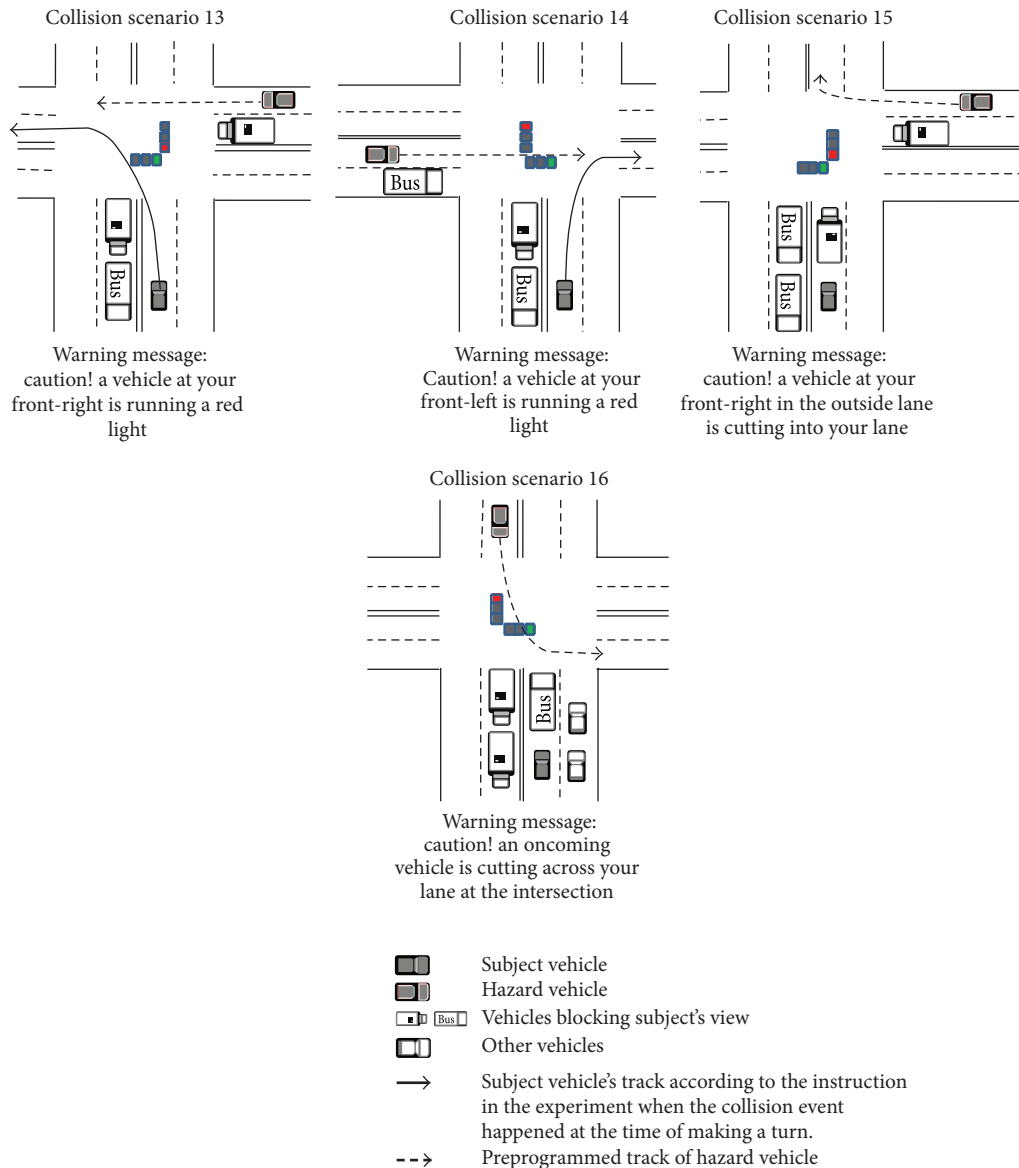
Warning message:
caution! a vehicle at your
front-left is running a red
light

Collision scenario 2

Warning message:
caution! a vehicle at your
front-left is running a red light

Collision scenario 3

Warning message:
caution! a vehicle at your
front-left is cutting into your
lane at the T-intersection

Collision scenario 4

Warning message:
caution! a vehicle at your
front-right is cutting into your
lane 1000* feet ahead

Collision scenario 5

Warning message:
caution! a vehicle at your
front-right is cutting into your
lane 1000* feet ahead

Collision scenario 6

Warning message:
caution! an oncoming vehicle
is drifting into your lane 1000*
feet ahead

Collision scenario 7

Warning message:
caution! an oncoming
vehicle is cutting across your
lane 1000* feet ahead

Collision scenario 8

Warning message:
caution! a vehicle at your
front-left is cutting into your
lane 1000* feet ahead

Collision scenario 9

Warning message:
caution! a vehicle is
overtaking you from your left
1000* feet ahead

Collision scenario 10

Warning message:
caution! a vehicle at your
front-right is cutting into your
lane 1000* feet ahead

Collision scenario 11

Warning message:
caution! an oncoming
vehicle in the left lane is
turning right and cutting you off

Collision scenario 12

Warning message:
caution! a vehicle at your
front-left is running a red light

FIGURE 2: Continued.

Collision scenario 13

Warning message:
caution! a vehicle at your
front-right is running a red
light

Collision scenario 14

Warning message:
Caution! a vehicle at your
front-left is running a red
light

Collision scenario 15

Warning message:
caution! a vehicle at your
front-right in the outside lane
is cutting into your lane

Collision scenario 16

Warning message:
caution! an oncoming
vehicle is cutting across your
lane at the intersection

◼◻ Subject vehicle
◼◻ Hazard vehicle
◼◻ Bus Vehicles blocking subject's view
◻◻ Other vehicles

⟶ Subject vehicle's track according to the instruction
in the experiment when the collision event
happened at the time of making a turn.
--→ Preprogrammed track of hazard vehicle

FIGURE 2: The designed scenarios and verbal warnings.

In the experiment, the sight of the participant regarding the potential hazard was blocked in purpose so that they could only rely on the auditory warning to learn about the upcoming collision event until the last minute (i.e., they will not be able to see the hazard until the time when they cannot avoid the hazard successfully even with a full braking when they confirm the hazard by vision). The behavioural measurements in this case indicated the safety benefits brought by the warnings without confounding participant's judgement of the hazards. In the meantime, the nonwarning messages (e.g., news and weather forecast) were randomly assigned during the experiment in case participants would brake the vehicle whenever there is a message. Therefore we could assume that drivers' responses to the verbal warnings are based on their understanding of the warning content. The nonwarning messages were designed carefully without causing interference with warnings. They were serving as distracted sources in real driving situations.

3.1.5. Procedure. Upon arrival at the laboratory, participants were asked to sign an informed consent form and then complete the demographic questionnaire. In the experiment, they will experience one potential collision scenario with its warning broadcasted to them in each trial. After each trial of the test block, they were asked to complete the Verbal Warning Utility Scale (VWUS) to evaluate the utility of the verbal warnings in helping them avoid the corresponding collision event.

Before the experiment, participants were trained by completing a practice block to get familiar with the operation of the driving simulator and the driving environment. During

TABLE 2: Mean and standard deviation for demographic measures and Spearman correlation coefficient.

Variables	Mean (SD)	Spearman's rho	p
Age (years)	21.13 (2.54)	−.01	.99
Gender (proportion of male)	0.75	.02	.63
Annual millage (miles)	8343.75 (6437.95)	.01	.81
License year (years)	3.50 (2.36)	−.01	.77
Reduced Kinetic Energy (J)	314.92 (152.09)	.56	<.001
VWUS total score	5.88 (2.51)		

the practice block, participants were asked to drive a four-mile distance with two randomly selected collision events (with corresponding verbal warnings) and five nonwarning messages broadcasted. They were informed that there would be collision events with corresponding verbal warnings and that they could respond based on their own driving experience. Subjects were instructed to drive in the inside lane and were informed that there would be messages broadcasting during the driving task. The scenario in the practice block was designed similarly with the one in the test block.

Following the practice session, participants completed the test block comprising of sixteen trials under a 2-lane (in each direction) local environment. Before the formal experiment, all participants were suggested to adjust the seat until they felt comfortable and their feet could come in full contact with the surface of the pedal. In the formal experiment, all participants were required to follow normal traffic laws and try to keep the speed at 45 mile/h. They would be advised to adjust the speed if their own speed was lower than 40 mile/h or higher than 50 mile/h when there was no warning, turn, or stop sign or red light.

3.2. Results

3.2.1. Descriptive Analysis. In total, we obtained 512 VWUS total scores of warnings (32 participants × 16 warnings each). Table 2 outlines descriptive statistics and obtained the associations between the VWUS and demographic variables via correlations between its total scores with those variables such as age, gender, annual mileage, and license year. All Spearman correlations coefficients between VWUS total scores and demographic characteristics were not significant ($p > .05$) with low correlation coefficients ($r < .06$), indicating the general criteria for evaluating warning utility. The reduced kinetic energy had a significant correlation with VWUS total scores.

3.2.2. Split-Half Reliability. Since the VWUS only had one question for each item, we were unable to conduct internal consistency for the scale. In order to test its reliability, the split-half reliability analysis was conducted for the 32 participants who completed the VWUS during the experiment. The Cronbach's α for each half was .762 and .777, respectively, and the split-half reliability for the VWUS was high with a Spearman-Brown Coefficient of .873.

3.2.3. Factor Analysis. A factor analysis was conducted on the 4 VWUS items with varimax rotation. Bartlett's test of

TABLE 3: Loading values on items of VWUS.

Scale items	Factor 1	Factor 2
Time to Display	−0.84	−0.51
Accuracy of Understanding	0.75	
Representation of Event Urgency	0.60	
Alertness		0.97
Initial eigenvalues	1.66	1.27
Variance explained (%)	41.51	31.7

sphericity $\chi^2(6) = 841.37$, $p < .001$, indicated that correlations between items were sufficiently large enough for the analysis. An initial analysis was run to obtain eigenvalues for each component in the data. Two components were obtained with the eigenvalues over 1.0 according to Kaiser's criterion [50] and in combination explained 73.22% of the variance.

The 2-factor extraction accounted for 73.22% of the variance with communalities ranging from 0.51 to 0.97 and 1 item cross-loadings. The 1-factor extraction accounted for 41.51% of the variance with communalities ranging from 0.60 to 0.84, which explained less of the variance. The 3-factor extraction accounted for 97.68% of the variance with communalities ranging from 0.51 to 0.97 and 3-item cross-loadings. Based on Osborne and Waters's [51] criteria for factor extraction the 2-factor solution was selected because it had the least item cross-loadings and explained relative higher variance. Table 3 showed the initial eigenvalues and variance explained by each factor. The summary of the loadings values of each item in the rotated factor matrix was also provided with loading values less than .30 being suppressed [51]. By examining the item loading on these factors, specific themes were defined based on the content of items among each factor. The first factor which was defined as "Accuracy of Hazard Representation" contained 3 items, including "Representation of Event Urgency," "Time to Display," and "Accuracy of Understanding". The second factor contained the item Alertness.

3.2.4. Criterion Validity. Figure 3 illustrated a significant correlation between VWUS total utility score (calculated by each item's rating and weighting scores in VWUS) and overall utility score ($r = .851$, $p < .001$) obtained directly from the studies. It is suggested that the calculated total utility score using VWUS is a good fit for the overall utility score in assessing the warning utility in general.

The intercorrelation among scores of each factor and the overall utility was reported with Spearman correlation

TABLE 4: Intercorrelation of scores of VWUS factors and the overall utility.

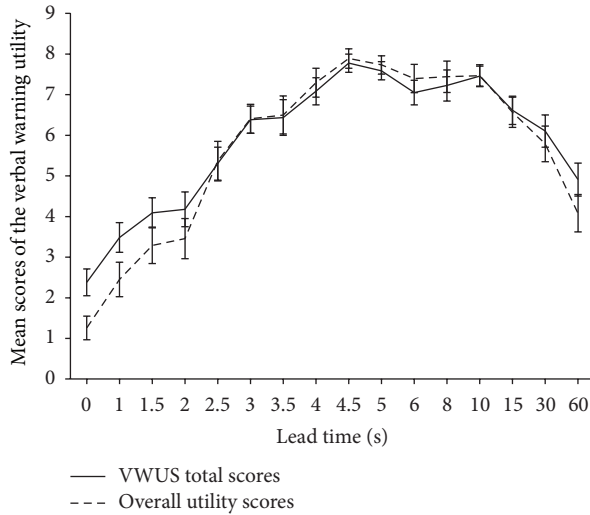| Variables	1	2
Factor 1: Accuracy of Hazard Representation	—	
Factor 2: Alertness	.07	—
Overall utility score	−.10***	.86***

Note: *** $p < .01$.

FIGURE 3: The correlation between the VWUS total scores and the overall utility scores (error bar: ±1 SE).

coefficients since the assumption of normal distribution was violated. As it is shown in Table 4, the overall utility score was significantly correlated with scores of factor "Accuracy of Hazard Representation" and factor "Alertness." The intercorrelation between two factors was not significant ($r_s = .07$, $p = .12$). In terms of the items on factor "Accuracy of Hazard Representation," the intercorrelation between Representation of Event Urgency and Time to Display was ($r_s = −.35$, $p < .01$). That between Representation of Event Urgency and Accuracy of Understanding was ($r_s = −.15$, $p < .01$) and that between Time to Display and Accuracy of Understanding was ($r_s = −.45$, $p < .01$).

Two sequential linear regression analyses were conducted to determine whether the VWUS could significantly predict the warning utility using criterion validity measures (i.e., reduced kinetic energy) obtained from the behavioural experiment. This behavioural measurement served as an indicator of safety benefits brought by the warnings. The collision rate was defined as the proportion of drivers getting into the collision among the total drivers under each level of the lead time. Gender, age, and license year were entered at Step 1; order of lead time, scenario type, and warning lead time were entered at Step 2; VWUS total scores of each warning ($N = 502$) or scores of two factors (Accuracy of Hazard Representation and Alertness) were entered at Step 3 in each analysis, respectively.

The results of the first model with the examination of the VWUS total score in predicting safety benefits brought by

the verbal warnings was presented in Table 5. The results indicated that the VWUS total score calculated by rating score and weighted score of each dimension significantly predicted reduced kinetic energy ($t = 15.50$, $p < .001$). Specifically, more kinetic energy was reduced when drivers responded to warnings with higher VWUS total scores. The results of the second model with scores of two factors of VWUS suggested that the accuracy of hazard representation significantly predicted the reduced kinetic energy ($t = −2.36$, $p < .05$). It is suggested that the more accurate the warning described the hazard, the more the kinetic energy was reduced as a result of the braking response to warnings. Both results suggested the validity of the VMUS in predicting the effectiveness of verbal warnings.

The logistic regression analysis was then conducted to determine whether the VWUS could significantly predict the warning utility using collision as criterion validity measures. The collision (avoid collision: 0 and collision: 1) served as another indicator of safety benefits brought by warnings. Gender, age, and license year were entered at Step 1, order of lead time, scenario type, and warning lead time were entered at Step 2, and VWUS total scores of each warning ($N = 512$) were entered at Step 3. Another logistic regression analysis was performed with scores of two factors of VWUS entered at Step 3. However, the test of the logistic regression model with scores of two factors of VWUS against a constant only model was not statistically significant ($\chi^2(8) = 12.53$, $p = .13$). Therefore, only the model with VWUS total scores as predictors was reported. The results of the logistic regression analysis with the examination of the VWUS total score in predicting safety benefits brought by the verbal warnings were presented in Table 6.

The test of the first logistic regression model against a constant only model was statistically significant, indicating that the predictors as a set reliably significantly predict the probability of collision ($\chi^2(8) = 15.62$, $p < .05$). Nagelkerke's R^2 of .50 indicated a moderate relationship between prediction and grouping. Successful prediction of probability of collision was 80%. The Wald criterion demonstrated that VWUS total score made a significant contribution to prediction ($p < .001$). Age and warning lead time also significantly predicted collision. In particular, warnings with higher VWUS total scores significantly reduced the probability of collision.

3.2.5. Scale Sensitivity. The scale sensitivity was tested with the comparison of two ANCOVA tests with warning lead time as the independent variable, scenario type as the covariate, and scale scores and reduced kinetic energy as the dependent variables, respectively. The results indicated significant main effects of lead time on scale scores ($F(15, 490) = 21.68$, $p < .001$, partial $\eta^2 = .40$) and reduced kinetic energy ($F(15, 490) = 28.49$, $p < .001$, partial $\eta^2 = .47$) after controlling for the effect of scenario type. As it is shown in Figure 4, the changes of subjective warning utility (i.e., VWUS total score) and objective warning utility (i.e., reduced kinetic energy) followed a fairly similar pattern as the lead time changes.

TABLE 5: Linear regression results for the prediction of Reduced Kinetic Energy with VWUS scores.

Variables	R^2	F for change in R^2	B	SE	Beta	t
Model 1						
Gender			−10.21	12.76	−.03	−.80
Age			−6.46	2.47	−.11	−2.62*
License obtaining year	.02	2.72*	7.50	2.67	.12	2.82**
Order of lead time			−.63	1.20	−.02	−.52
Scenario type			−2.15	1.20	−.07	−1.80
Lead time	.03	2.35	.01	.01	.08	2.09*
VWUS total score	.35	240.33***	34.36	2.22	−.57	15.50***
Model 2						
Gender			−4.09	15.37	−0.01	−0.27
Age			−6.57	2.99	−0.11	−2.20*
License obtaining year	0.02	2.72*	8.79	3.23	0.14	2.72*
Order of lead time			1.17	1.44	0.04	0.81
Scenario type			−3.67	1.44	−0.11	−2.55*
Lead time	0.03	2.25	0.01	0.01	0.03	0.77
Accuracy of Hazard Representation			−72.44	30.65	−0.25	−2.36*
Alertness	0.05	4.70**	−41.10	31.16	−0.14	−1.32

Note: $^*p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

TABLE 6: Logistic regression results for the prediction of collision in the VWUS scores.

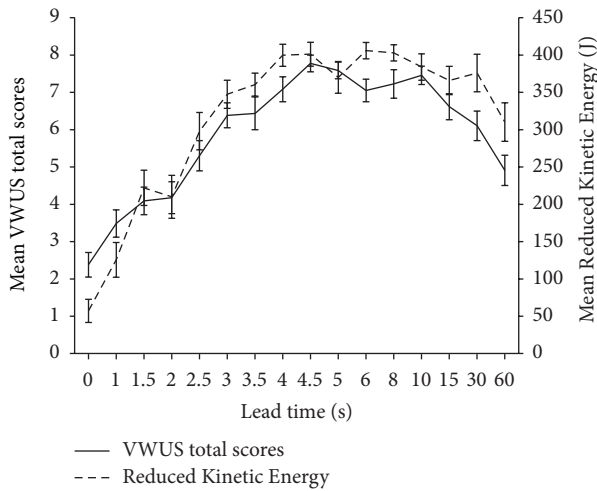|	B	S.E.	Wald χ^2	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Model 1						
Gender	−0.19	0.27	0.47	0.83	0.49	1.41
Age	0.11	0.05	4.37*	1.11	1.01	1.23
License obtaining year	−0.08	0.06	1.86	0.93	0.83	1.03
Order of treatment	−0.02	0.03	0.35	0.99	0.94	1.04
Scenario type	−0.04	0.03	2.60	0.96	0.91	1.01
Lead time	−0.03	0.01	12.25***	0.97	0.96	0.99
VWUS total score	−0.62	0.06	114.87***	0.54	0.48	0.60

Note: $^*p < .05$, $^{***}p < .001$.

FIGURE 4: The changes of VWUS total score and objective warning utility as a function of lead time (error bar: ±1 SE).

Specifically, the results of the post hoc tests indicated that the scale scores on warning utility and kinetic reduced energy were significantly different on the following pairs of lead time levels ($p < .05$): lead time of 0 s and lead time levels ranging from 1 s to 60 s; lead time of 1 s and lead time levels ranging from 2.5 s to 60 s; lead time of 1.5 s and lead time levels ranging from 3 s to 30 s; lead time of 2 s and lead time levels ranging from 3 s to 30 s; lead time of 2.5 s and lead time levels ranging from 4 s to 8 s. The only difference in the post hoc tests was that the scale scores were different between lead time of 60 s and lead time levels ranging from 3 s to 30 s, whereas the reduced energy was different between lead time of 60 s and lead time levels ranging from 4 s to 30 s.

4. Discussion

By drawing on the experience of the development process of NASA-TLX, we developed the Verbal Warning Utility

Scale (VWUS) to evaluate the warning utility in general. The VWUS consisted of two parts: the total utility divided into four items serving as one part of the scale (including Representation of Event Urgency, Time to Display, Alertness, and Accuracy of Understanding) and the weighting of these items by comparing these items pairwise based on their perceived importance. The reliability analysis indicated excellent split-half reliability for the VWUS. Similar to NASA-TLX, each item of VWUS merely had one rating question so that we are unable to test the internal consistency for VWUS. However, scales with multiple questions for each dimension may cause the memory of the warning utility to decay during the evaluation process, resulting in less accurate or reliable ratings. Therefore, from the application perspective, this Verbal Warning Utility Scale developed in the current work with each rating question in each dimension provides a simple way to evaluate warning effectiveness without too much memory decay taking place.

The results from the validation study clearly indicated that Verbal Warning Utility Scale (VWUS) could significantly evaluate warning utility regarding objective safety benefits. It is suggested that a higher score of the VWUS revealed the warning with a higher utility, which had a higher chance to reduce the probability of potential risk during driving tasks. The linear and logistic regression analyses using the VWUS total scale suggested that it was a significant predictor of the objective warning utility indicators obtained from the experiment, including the reduced kinetic energy and the probability of collision. More specifically, a higher score of VWUS indicated that more reduced kinetic energy and lower collision rate could be achieved with the presence of the warning, which further suggested a higher utility of the warning. The relative low R square reported in the second linear regression model with two factors of VWUS as predictor can result from the fact that the accuracy of hazard representation was the only significant predictor of warning utility in the current validation experiment (see model 2 in Table 5). One reason for such results may be that the independent variable (lead time) designed in the current validation experiment mainly addressed the accuracy of hazard representation. Further validation studies need to be performed with warning alertness levels being addressed.

There is a significant increase in research in terms of warning design to improve the safe communication of the various systems using verbal warnings. In order to achieve a safe and successful interface between human and the system operation, the issues regarding warning effectiveness have raised researchers' attention. To date, much work has been done to subjectively evaluate warnings regarding different constructs including acceptance, usefulness, and ease of use [8, 9, 13]. However, previous studies have indicated subjective assessment of acceptance and perceived usefulness did not always reflect enhancement on driving performance when using the system [10, 11]. Laid on the foundation of previous works, the scale developed in the current study aims to evaluate the effectiveness of the verbal warnings with a focus on driver performance. The establishment of the Verbal Warning Utility Scale enables us to assess whether the verbal warnings designed in the system is effective regarding the improvement of driver performance rather than drivers' subjective opinion on the warnings.

The current development of VWUS is mainly based on verbal warnings. Since all the dimensions chosen in the current version of the scale are quite general, this scale may also be suitable for examining the utility of nonverbal warnings. Compared to nonverbal warnings, verbal warnings are much easier to identify, understand, and remember. However, verbal warnings may take more time to respond, which makes nonverbal warnings a better choice if the situation is emergent or if the human is trained to use the warning system. In the warning design of the system, both types of warnings can be chosen to alarm humans with potential risks. The current validation experiment has already indicated the validity of the scale applied to verbal warnings. Therefore, further work could validate its effectiveness in testing the utility of nonverbal warnings in order to increase its universality. In the meantime, the current validation study suggested the validity of the scale to evaluate warnings of transportation systems. In order to generalize the testing power of VWUS, further validation studies are also necessary to examine the correlations between VWUS and human performance to avoid potential hazards with different warnings broadcasting in diverse domains.

The scale sensitivity was tested with the comparison of two ANCOVA tests on scale scores and reduced kinetic energy, respectively. The results indicated that the VWUS scale was sensitive to the changes of warning effectiveness with a lead time change of 1 s. The changes of subjective warning utility (i.e., VWUS total score) and objective warning utility (i.e., Reduced Kinetic Energy) follow a fairly similar pattern as the lead time changes. The results from the sensitivity analysis suggested that the VWUS scale was sensitive to the changes of warning effectiveness with a change in warning timing of as little as 1 s. The studies regarding warning timings indicated a change in warning timing of 1.5 s for nonverbal warnings [52] and 2.5 s for verbal warnings [47] was enough to change drivers' collision avoidance performance. Compared to previous work, the scale developed in the current work showed adequate sensitivity to the changes in warning effectiveness.

5. Conclusion

The Verbal Warning Utility Scale (VWUS) developed in the current work was suggested as an effective tool to evaluate the utility of verbal warnings with a relatively good reliability and validity obtained from the validation study. Compared to previous works using behavioural experiments, the development of the scale provides a simpler way to assess the effectiveness of verbal warnings in intelligent transportation systems. This scale can be further applied to evaluate the design of warnings in order to improve the communications in the intelligent transportation system.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] K. R. Laughery, "Safety communications: warnings," *Applied Ergonomics*, vol. 37, no. 4, pp. 467–478, 2006.

[2] F.-J. Wu, Y.-F. Kao, and Y.-C. Tseng, "From wireless sensor networks towards cyber physical systems," *Pervasive and Mobile Computing*, vol. 7, no. 4, pp. 397–413, 2011.

[3] F. Ye, M. Adams, and S. Roy, "V2V wireless communication protocol for rear-end collision avoidance on highways," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '08)*, pp. 375–379, May 2008.

[4] M. Jerbi, P. Marlier, and S. M. Senouci, "Experimental assessment of V2V and I2V communications," in *Proceedings of the IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS '07)*, pp. 1–6, IEEE, Pisa, Italy, October 2007.

[5] J. Santa, A. F. Gómez-Skarmeta, and M. Sánchez-Artigas, "Architecture and evaluation of a unified V2V and V2I communication system based on cellular networks," *Computer Communications*, vol. 31, no. 12, pp. 2850–2861, 2008.

[6] B. Shackel, "Usability—context, framework, definition, design and evaluation," in *Human Factors for Informatics Usability*, pp. 21–37, Cambridge University Press, New York, NY, USA, 1991.

[7] M. A. Regan, E. Mitsopoulos, N. Haworth, and K. Young, "Acceptability of in-vehicle intelligent transport systems to Victorian car drivers," Public Policy Report 02/02, Monash University Accident Research Centre, 2002.

[8] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 1, pp. 1–10, 1997.

[9] E. Adell, "Acceptance of driver support systems," in *Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems*, Berlin, Germany, 2010.

[10] F. Naujoks and A. Neukum, "Timing of in-vehicle advisory warnings based on cooperative perception," in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting*, pp. 193–206, 2014.

[11] J. Werneke, A. Kleen, and M. Vollrath, "Perfect timing: urgency, not driving situations, influence the best timing to activate warnings," *Human Factors*, vol. 56, no. 2, pp. 249–259, 2014.

[12] Y. Zhang, C. Wu, J. Wan, and C. Qiao, "Development and validation of warning message utility scale (WMUS)," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, pp. 1179–1183, 2013.

[13] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[14] A. H. Segars and V. Grover, "Re-examining perceived ease of use and usefulness: a confirmatory factor analysis," *MIS Quarterly: Management Information Systems*, vol. 17, no. 4, pp. 517–525, 1993.

[15] J. R. Sayer, D. J. LeBlanc, M. L. Mefford, and J. Devonshire, "Field test results of a road departure crash warning system: driver acceptance, perceived utility and willingness to purchase," in *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Stevenson, Wash, USA, July 2007.

[16] M. Horst, M. Kuttschreuter, and J. M. Gutteling, "Perceived usefulness, personal experiences, risk perception and trust as determinants of adoption of e-government services in The Netherlands," *Computers in Human Behavior*, vol. 23, no. 4, pp. 1838–1852, 2007.

[17] M. S. Wogalter and S. L. Young, "Behavioural compliance to voice and print warnings," *Ergonomics*, vol. 34, no. 1, pp. 79–89, 2013.

[18] J. Edworthy and R. Hards, "Learning auditory warnings: the effects of sound type, verbal labelling and imagery on the identification of alarm sounds," *International Journal of Industrial Ergonomics*, vol. 24, no. 6, pp. 603–618, 1999.

[19] B. T. K. Momtahan, R. Hétu, and B. Tansley, "Audibility and identification of auditory alarms in the operating room and intensive care unit," *Ergonomics*, vol. 36, no. 10, pp. 1159–1176, 1993.

[20] R. D. Patterson, *Guidelines for Auditory Warning Systems on Civil Aircraft*, 1982.

[21] M. S. Sanders and E. J. McCormick, *Human Factors in Engineering and Design*, Mcgraw-Hill Book, New York, NY, USA, 6th edition, 1987.

[22] J. Serrano, L. L. Di Stasi, A. Megías, and A. Catena, "Effect of directional speech warnings on road hazard detection," *Traffic Injury Prevention*, vol. 12, no. 6, pp. 630–635, 2011.

[23] S.-H. Chang, C.-Y. Lin, C.-P. Fung, J.-R. Hwang, and J.-L. Doong, "Driving performance assessment: effects of traffic accident location and alarm content," *Accident Analysis & Prevention*, vol. 40, no. 5, pp. 1637–1643, 2008.

[24] S.-H. Chang, C.-Y. Lin, C.-C. Hsu, C.-P. Fung, and J.-R. Hwang, "The effect of a collision warning system on the driving performance of young drivers at intersections," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 5, pp. 371–380, 2009.

[25] C. L. Baldwin, J. F. May, and R. Parasuraman, "Auditory forward collision warnings reduce crashes associated with task-induced fatigue in young and older drivers," *International Journal of Human Factors and Ergonomics*, vol. 3, pp. 107–121, 2014.

[26] I. S. O. W. Draft, "Road vehicles—ergonomic aspects of transport information and control systems—simulated lane change test to assess driver distractio," ISO TC 22, 2004.

[27] J. M. Noyes, E. Hellier, and J. Edworthy, "Speech warnings: a review," *Theoretical Issues in Ergonomics Science*, vol. 7, no. 6, pp. 551–571, 2006.

[28] R. G. Mortimer, "Effectiveness of a warning as measured by behavior change," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 51, no. 9, pp. 574–577, 2007.

[29] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," in *Human Mental Workload*, North Holland Press, Amsterdam, The Netherlands, 1988.

[30] C. L. Baldwin, "Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance," *Ergonomics*, vol. 54, no. 4, pp. 328–337, 2011.

[31] E. Hellier, J. Edworthy, B. Weedon, K. Walters, and A. Adams, "The perceived urgency of speech warnings: semantics versus acoustics," *Human Factors*, vol. 44, no. 1, pp. 1–17, 2002.

[32] B. Weedon, E. Hellier, J. Edworthy, and K. Walters, "Perceived urgency in speech warnings," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, pp. 690–693, August 2000.

[33] J. D. Lee, D. V. McGehee, T. L. Brown, and M. L. Reyes, "Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator," *Human Factors*, vol. 44, no. 2, pp. 314–334, 2002.

[34] J. Fleming, P. Green, and S. Katz, "Driver performance and memory for traffic messages: effects of the number of messages, audio quality, and relevance," UMTRI Technical Report 98-22, University of Michigan, Ann Arbor, Mich, USA, 1998.

[35] M. S. Wogalter, S. S. Godfrey, G. A. Fontenelle, D. R. Desaulniers, P. R. Rothstein, and K. R. Laughery, "Effectiveness of warnings," *Human Factors*, vol. 29, no. 5, pp. 599–612, 1987.

[36] J. J. Scott and R. Gray, "A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving," *Human Factors*, vol. 50, no. 2, pp. 264–275, 2008.

[37] E. E. Wiese and J. D. Lee, "Auditory alerts for in-vehicle information systems: the effects of temporal conflict and sound parameters on driver attitudes and performance," *Ergonomics*, vol. 47, no. 9, pp. 965–986, 2004.

[38] M. Polkosky and J. Lewis, "The function of nonspeech audio in speech recognition applications: a review of the literature," Tech. Rep. TR 29.3405, IBM Voice Systems, 2001.

[39] M. F. Lesch, "Comprehension and memory for warning symbols: age-related differences and impact of training," *Journal of Safety Research*, vol. 34, no. 5, pp. 495–505, 2003.

[40] G. M. Goldhaber and M. A. DeTurck, "Effectiveness of warning signs: 'familiarity effects'," *Forensic Reports*, vol. 1, no. 4, pp. 281–301, 1988.

[41] M. L. Cummings, R. M. Kilgore, E. Wang, L. Tijerina, and D. S. Kochhar, "Effects of single versus multiple warnings on driver performance," *Human Factors*, vol. 49, no. 6, pp. 1097–1106, 2007.

[42] J. H. McDonald, *Handbook of Biological Statistics*, vol. 2, Sparky House Publishing, Baltimore, Md, USA, 2009.

[43] N. Hancock and P. A. Meshkati, *Human Mental Workload*, North-Holland, Amsterdam, The Netherlands, 1988.

[44] A. Tang and A. Yip, "Collision avoidance timing analysis of DSRC-based vehicles," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 182–195, 2010.

[45] H. Sohn, J. D. Lee, D. L. Bricker, and J. D. Hoffman, "A dynamic programming algorithm for scheduling in-vehicle messages," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 226–234, 2008.

[46] I. T. S. Sae, *In-Vehicle Message Priority, Troy*, vol. J2395, MI: Safety and Human Factors Committee, 2002.

[47] X. Yan, Y. Zhang, and L. Ma, "The influence of in-vehicle speech warning timing on drivers' collision avoidance performance at signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 231–242, 2015.

[48] X. Yan, Q. Xue, L. Ma, and Y. Xu, "Driving-simulator-based test on the effectiveness of auditory red-light running vehicle warning system based on time-to-collision sensor," *Sensors*, vol. 14, no. 2, pp. 3631–3651, 2014.

[49] Y. Zhang, X. Yan, and Z. Yang, "Discrimination of effects between directional and nondirectional information of auditory warning on driving behavior," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 980281, 7 pages, 2015.

[50] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141–151, 1960.

[51] J. W. Osborne and E. Waters, "Four assumptions of multiple regression that researchers should always test," *Practical Assessment, Research & Evaluation*, vol. 8, no. 2, 2002.

[52] D. V. McGehee, T. L. Brown, J. D. Lee, and T. B. Wilson, "Effect of warning timing on collision avoidance behavior in a stationary lead vehicle scenario," *Transportation Research Record*, vol. 1803, no. 1, pp. 1–6, 2002.