Hazard Level Perceptions of Current and Proposed Warning Sign and Label Panels

Michael J. Kalsher
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Michael S. Wogalter
North Carolina State University
Raleigh, NC 27695-7801

Blair M. Brewster
Electromark Company
Wolcott, NY 14590

Marilyn E. Spunar
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

ABSTRACT

A growing number of studies have investigated factors associated with various measures of warning effectiveness, including noticeability, comprehension, and most importantly, compliance. Some research has begun to examine the components comprising signs and labels (e.g., signal words, color). However, there has been virtually no research on people's perceptions of sign/label configurations that are currently found on warnings. The present study evaluates the warning styles that are specified in ANSI (1991) Z535.2 and Z535.4 standards, as well as a set of proposed styles. The results confirmed several specifications in current standards (e.g., the signal word DANGER was perceived as more hazardous than the other currently-specified signal words), whereas other specifications were not fully confirmed (e.g., the WARNING configuration as indicating higher hazard than the CAUTION configuration). Some newly developed warning styles (e.g., using the signal word DEADLY and a skull icon) show promise for better signaling highly hazardous conditions.

INTRODUCTION

A growing body of evidence suggests that certain characteristics of warnings, such as the use of signal words and icons, color, shape, borders and other features can enhance the effectiveness of warning signs and labels, by increasing noticeability and perceived hazard, facilitating comprehension, and motivating compliance (see Laughery, Wogalter, and Young, 1994). Although research has begun to explore the relative contributions of component features of warning signs and labels, very little work has examined the effects of combinations of features that comprise common multi-faceted warning configurations.

The present research examines peoples's perceptions of warning configurations specified by existing standards and guidelines (ANSI, 1991; FMC, 1985; Westinghouse, 1981). Currently, some of these recommendations are inconsistent. For example, the warning sign standard ANSI Z535.2 recommends that warning headers use shape borders around the signal words DANGER (an oval) and WARNING (a diamond). Presumably shape serves as a redundant cue in cases where an individual might not be able to understand the signal word (e.g., the person can not read the language) or see the colors (e.g., under dim light). In contrast, the product label standard ANSI Z535.4 specifies header configurations with the same signal words and colors but recommends a different shape cue, a signal icon (an exclamation point inside a triangle) preceding the signal word.

A related issue is the degree of hazard conveyed by each of the ANSI-recommended signal words. Current standards describe DANGER as connoting high hazard, followed by WARNING which in turn is followed by the term CAUTION. Whereas, research has consistently found DANGER to connote higher hazard levels than the other two terms, most research shows that people do not distinguish between WARNING and CAUTION with regard to connoted hazard levels (e.g., Wogalter and Silver, 1990, in press). Other research (e.g., Chapanis, 1994) shows that the component colors orange and yellow which are respectively tied to the terms WARNING and CAUTION in the standards, are also not readily distinguished on the hazard dimension.

Interestingly, most of the design features described in the standards and guidelines are not based on empirical research. Thus, it is possible that some of the specifications do not produce the best kinds of warnings. Alternative designs, that make use of different signal words, other color combinations, and include well-designed icons and pictorials might better signal hazardous conditions.

The present research compares people's hazard perceptions and rated noticeability of three sets of warning configurations. Two were constructed in accordance to existing environmental sign (ANSI Z535.2) and product label (ANSI Z535.4) standards, and the third was a newly developed set of designs that has been proposed as an alternative to the existing standards.

METHOD

Participants

Two groups of volunteers served as participants. The first group consisted of 36 (16 males, 20 females) college

students. The students ranged in age from 18 to 23 years with a mean of 20.1. The second group consisted of 124 non-students from various public locations in upstate New York and western Massachusetts. The non-students (65 males, 59 females) ranged in age from 18 to 84 with a mean of 42.6.

Stimulus Materials

Gray-scale representations of the warning stimuli are shown in Figure 1. Table 1 lists some of the major characteristics of the warning stimuli. Stimuli consisted of 16 magnetic cards of complete warnings, except that a series of X's was inserted into the space typically occupied by the warning message. The cards were all identical in size (12.7 x 17.8 cm; 5 x 7 in.) and were constructed to resemble existing warning designs specified in current standards or they were newly-constructed as possible improvements (proposed). The cards differed on one or more of the following dimensions: the signal word; the color used for the foreground figures and print as well as the background header and message panels; capitalization (all versus small caps), letter size and boldness, left vs. center justification, and the presence of shape borders or icons (e.g., triangle/exclamation point or skull).

The five cards comprising the first set were consistent with ANSI 535.2 standards for environmental sign warnings having the signal words in the header: DANGER, WARNING, CAUTION, NOTICE and SAFETY FIRST. The four cards in the second set were consistent with ANSI Z535.4 standards for product labels with the signal words: DANGER, WARNING, CAUTION, and NOTICE. The third set consisted of seven newly-constructed warning headers. Four had the signal words: DANGER, WARNING, CAUTION, and NOTICE. The proposed headers differed from their counterparts in the first two sets in several ways: (1) the letters comprising the signal words were larger and bolder, and the first capitalized letter of the signal word was slightly larger than the remaining capitalized letters (i.e., small caps); (2) there was somewhat more color space in the header (i.e., greater surface space of red in the DANGER header); (3) all text is left justified; and (4) there is no shape or icon in the headers for the four conventional signal words. The proposed set also contained three signs with the term "deadly" in the header—a term that received the highest hazard ratings in previous research (e.g., Wogalter and Silver, 1990; in press). Two had DEADLY in all capital letters; one had mixed-case letters. Two had a black header panel with a red message panel, and the other had a red header panel with a white message panel. Also, these three headers included a skull icon.

Figure 1. Representations of ANSI Z535.2, ANSI Z535.4 , and Newly-Constructed Warning Configurations

Table 1

Specific Characteristics of the Warning Header and Message Panels in ANSI Z535.2, Z535.4, and a Proposed Set

Stimulus Number	Signal Word	Header Shape/Icon?	Header Color Print	Header Color Background	Message Panel Color Print	Message Panel Color Background
ANSI 535.2—Signs						
24	DANGER	oval shape	white	red oval with white border on black	black	white
20	WARNING	hexagon shape	black	orange hexagon on black	black	orange
66	CAUTION	No	yellow	black	black	yellow
71	NOTICE	No	white	blue	black	white
56	SAFETY FIRST	No	white	green	black	white
ANSI 535.4—Product Labels						
40	DANGER	red ! white triangle	white	red	red	white
62	WARNING	orange ! in black triangle	black	orange	black	white
43	CAUTION	yellow ! in black triangle	black	yellow	black	white
01	NOTICE	No	white	blue	black	white
Proposed Formats						
73	DANGER	No	white	red	black	white
04	WARNING	No	black	orange	black	white
16	CAUTION	No	black	yellow	black	white
75	NOTICE	No	white	blue	black	white
45	DEADLY	white skull in square border	white	black	white	red
80	Deadly	white skull in square border	white	black	white	red
38	DEADLY	white skull in square border	white	black	red	white

Procedure

The procedure was conducted in four phases.

Phase 1—Ratings within sets. Participants were asked to examine the three sets of cards, one set at a time, and to rate each card on: (1) the level of hazard conveyed; and (2) how "attention-getting" each warning appeared to be. Hazard and noticeability ratings were made on 5-point Likert-type scales. The anchors for the hazard scale were: 0 = no, 1 = low, 2 = moderate, 3 = high, and 4 = extreme. The anchors for the attention-getting scale were: 0 = not at all, 2 = moderately; 3= highly, and 4 = extremely.

Phase 2—Within-set hazard rankings. After completing the ratings, participants were asked to sort the cards in each set based on the hazard level conveyed. Participants were instructed to arrange the cards with the warning representing the greatest hazard farthest away from them and the warning for the least hazard closest to them. Card order was recorded on a response sheet.

Phase 3—Ratings of different headers for each signal word. For the second part of the study, 15 of the cards (the card with SAFETY FIRST was omitted) were recombined to form five new sets (three cards per set), based on the signal word in the headers. These sets can be seen in Table 3. Each contains the three variants for each of the following signal words: DANGER, WARNING, CAUTION, NOTICE, and DEADLY. Participants carried out rating procedures similar to those described in Phase 1.

Phase 4—Hazard rankings of different headers for each signal word. For the sets described in Phase 3, participants ranked the configurations using methods similar to those described in Phase 2.

For all procedures, presentation orders were randomly determined both for the order of the sets received, the card order within each set, and with regard to the hazard and noticeability ratings, which rating scale they used first. After participants completed the tasks, they were debriefed, thanked, and dismissed.

RESULTS

Table 2 shows the overall mean hazard ratings, the hazard ratings broken down for the students and nonstudents, the within-set hazard rankings, and the mean noticeability ratings for the three sets of stimuli (ANSI Z535.2 and Z535.4 formats

and the proposed alternate format). Higher hazard ratings and lower hazard rankings indicate greater perceived hazard. Higher noticeability ratings indicate greater perceived attention-gettingness. Table 2 shows that same basic pattern for all measures. Pearson correlation between the overall hazard ratings and noticeability ratings was .99, p < .00001. Spearman correlations between the hazard ranks and ratings was .-71, and between the hazard ranks and reported noticeability was -.75, *ps* <.01. In addition, analyses of the hazard ranks and the noticeability ratings showed virtually the same pattern of significant effects as the hazard ratings. Because of the similarity among measures, only the analyses of the hazard ratings are described below.

Separate two-way (participant group: student versus nonstudent) X (warning set) analyses of variance (ANOVAs) were conducted on the hazard ratings. Significant effects were followed by simple effects analysis and pairwise comparisons using Tukey's HSD Test. Only effects with chance probability levels of less than .05 are described.

Comparison of ratings within sets. The ANOVA on the hazard ratings of the ANSI Z535.2 warning signs was significant, $F(4, 632) = 145.17$, $MSe = .57$, $p < .00001$. DANGER received the highest ratings, followed by WARNING, CAUTION, NOTICE, and SAFETY FIRST. Comparisons showed that all differences were significant, except between WARNING and CAUTION, and between NOTICE and SAFETY FIRST (*ps* > .05). The ANOVA also showed a significant interaction, $F(4, 632) = 4.75$, $MSe = .57$, $p < .001$. The pattern means was consistent for both participant groups, except that the nonstudents rated DANGER higher than the students, and the students rated SAFETY FIRST higher than the nonstudents.

The ANOVA on the ANSI Z535.4 product label warnings was significant, $F(3, 474) = 225.43$, $MSe = .41$, $p < .00001$. Comparisons showed that DANGER was rated higher than all other stimuli in the set, followed by WARNING, CAUTION, and NOTICE. All comparisons were significant except between WARNING and CAUTION. The ANOVA also showed a main effect of group, $F(1, 158) = 5.37$, $MSe = 1.17$, $p < .05$, and an interaction, $F(3, 474) = 2.93$, $MSe = .41$, $p < .05$. In general, the nonstudents gave higher ratings than the students. The means were consistent between groups except the nonstudents rated DANGER and WARNING higher than the students did.

The ANOVA on the newly-constructed (proposed) warnings showed a main effect of stimuli, $F(6, 948) = 218.77$, $MSe = .44$, $p < .00001$. Comparisons showed that all three DEADLY variants were rated significantly higher than the other warnings in this set. DANGER was rated next highest, followed by WARNING and CAUTION, and lastly by NOTICE. Comparisons showed that all were significantly different from each other, except between (a) DEADLY (all caps) and Deadly (mixed case)—both with the black header/red message panels, (b) the mixed-case Deadly and DEADLY with the red header/white message panel, and (c) WARNING and CAUTION. There was also a significant interaction, $F(6, 948) = 2.25$, $MSe = .44$, $p < .05$. The nonstudents rated the red header/white message DEADLY, and DANGER significantly higher than the students did.

Comparison of Configurations with the Same Signal Word

Table 3 shows the measures derived from the ratings and rankings in Phases 3 and 4. The five sets were composed of different format groupings (Z535.2, Z535.4, and proposed) for the same signal word. As with the earlier analyses, the hazard ratings and rankings, and the noticeability ratings showed nearly the same pattern of results. Therefore, only the analyses on the hazard ratings are presented.

The 2 (participant group) x 3 (variants) ANOVA for DANGER showed only a significant effect of group, $F(1, 157) = 16.01$, $MSe = 1.17$, $p < .0001$. Nonstudents gave higher ratings to these terms than students.

The ANOVA on WARNING showed a significant effect of variant, $F(2, 314) = 4.86.01$, $MSe = .36$, $p < .01$. Comparisons showed that the ANSI Z535.2 version was rated significantly higher than the proposed version.

The ANOVA on CAUTION showed a significant effect of variant, $F(2, 314) = 36.80$, $MSe = .28$, $p < .00001$. Comparisons showed that the ANSI Z535.2 CAUTION was

Table 2

Mean Hazard Ratings, Within-set Rankings, and Noticeability Ratings for ANSI Z535.2, ANSI Z535.4, and Proposed Formats [1]

#	Signal Word	Overall Hazard Rating	Student Hazard Rating	Nonstudent Hazard Rating	Overall Hazard Rank	Overall Noticeability Rating
ANSI Z535.2 Sign Format						
24	DANGER	3.2	2.9	3.3	1.4	3.1
20	WARNING	2.7	2.5	2.7	2.4	2.7
66	CAUTION	2.4	2.5	2.4	2.4	2.8
71	NOTICE	1.2	1.2	1.2	4.2	1.4
56	SAFETY FIRST	1.1	1.4	1.0	4.6	1.4
ANSI Z535.4 Consumer Product Label Format						
40	DANGER	3.4	3.1	3.5	1.1	3.4
62	WARNING	2.5	2.2	2.6	2.5	2.6
43	CAUTION	2.3	2.2	2.3	2.5	2.6
01	NOTICE	1.1	1.1	1.1	3.9	1.3
Newly Constructed/ Proposed Formats						
73	DANGER	3.1	2.8	3.2	4.0	3.1
04	WARNING	2.4	2.2	2.5	5.3	2.4
16	CAUTION	2.1	2.2	2.1	5.3	2.2
75	NOTICE	1.4	1.3	1.4	6.9	1.7
45	DEADLY	3.8	3.8	3.9	1.4	3.8
80	Deadly	3.7	3.7	3.7	2.0	3.6
38	DEADLY	3.6	3.3	3.6	3.1	3.3

[1] *Higher ratings and Lower ranking indicate greater perceived hazard. Rankings were performed within sets.*

Table 3

Mean Hazard Ratings, Within-set Rankings, and Noticeability Ratings Across ANSI Z535.2, ANSI Z535.4, and Proposed Formats[1]

#	Signal Word	Overall Hazard Rating	Student Hazard Rating	Nonstudent Hazard Rating	Overall Hazard Rank	Overall Noticeability Rating
24	DANGER (Z535.2)	3.2	2.9	3.3	1.5	3.2
40	DANGER (Z535.4)	3.2	2.7	3.3	2.0	3.1
73	DANGER (Proposed)	3.1	2.8	3.1	2.5	2.9
20	WARNING (Z535.2)	2.9	2.7	2.9	1.4	2.9
62	WARNING (Z535.4)	2.7	2.6	2.7	2.0	2.7
04	WARNING (Proposed)	2.6	2.5	2.7	2.6	2.6
66	CAUTION (Z535.2)	2.8	3.0	2.7	1.4	3.1
43	CAUTION (Z535.4)	2.5	2.5	2.5	1.8	2.6
16	CAUTION (Proposed)	2.3	2.2	2.4	2.7	2.4
71	NOTICE (Z535.2)	1.6	1.9	1.5	1.3	1.8
01	NOTICE (Z535.4)	1.2	1.2	1.2	2.8	1.4
75	NOTICE (Proposed)	1.4	1.6	1.4	1.9	1.6
45	DEADLY (Proposed)	3.9	3.8	3.9	1.2	3.8
80	Deadly (Proposed)	3.7	3.4	3.7	2.1	3.6
38	DEADLY (Proposed)	3.6	3.1	3.7	2.7	3.4

[1] *Higher ratings and Lower ranking indicate greater perceived hazard. Rankings were performed within sets.*

rated significantly higher than the ANSI Z535.4 version, which in turn was rated higher than the proposed version. The ANOVA also showed a significant interaction, $F(2, 314) = 5.43$, $MSe = .28$, $p < .01$, which was due to a crossover pattern for the two variants by groups. The students rated the Z535.2 version higher than the nonstudents, whereas, the nonstudents rated the proposed version higher than the students, but neither difference is individually significant.

The ANOVA on NOTICE produced a significant effect of variant, $F(2, 314) = 22.86$, $MSe = .25$, $p < .00001$. Comparisons show that the ANSI Z535.2 variant was rated higher than the proposed variant, which in turn received significantly higher hazard ratings than the ANSI Z535.4 variant. The ANOVA also showed a significant interaction, $F(2, 314) = 5.06$, $MSe = .25$ $p < .01$. Simple effects analysis showed that students rated the Z535.2 variant higher than the nonstudents did.

Finally, the ANOVA on the DEADLY ratings showed a significant effect of variant, $F(2, 314) = 27.16$, $MSe = .18$, $p < .00001$. The all caps version of DEADLY with the black header/red message background received significantly higher ratings than the mixed-case Deadly version of this header, which was in turn rated significantly higher than the red header/white message version. The ANOVA also showed an effect of group, $F(1, 157) = 11.69$, $MSe = .86$, $p < .001$, and an interaction, $F(2, 314) = 8.04$, $MSe = .18$, $p < .001$. The nonstudents gave higher overall hazard ratings than the students did. Simple effects analysis showed that the nonstudents rated all three very highly, whereas, the students gave significantly higher ratings to the DEADLY with black header/red message variant than the other two variants.

DISCUSSION

The two population groups produced remarkably similar results given number of stimuli and the relatively subtle differences between some of the configurations. Moreover, the results were reasonably consistent irrespective of the use of hazard ratings or rankings (or noticeability ratings).

Some of the findings confirm the specifications of the existing ANSI standards (e.g., DANGER received higher hazard ratings than WARNING or CAUTION), whereas others do not. For example, the standards specify that WARNING be used for greater level hazards than CAUTION; however, the results do not fully confirm this. While the nonstudents appeared to differentiate between WARNING and CAUTION, the students did not. Most research to date suggests little or no differentiation between these two terms or their associated colors (Chapanis, 1994; Wogalter and Silver, 1989, in press). While there might be a statistically significant difference in a given study or participant group, the practical difference appears inadequate.

Direct comparison between the different formats suggests that the warning sign Z535.2 configurations are perceived more hazardous than either the warning label standard Z535.4 or the proposed format. This does not mean that either of the latter two systems are inferior to the sign system, because the main issue is whether people discriminate separable hazard levels from the terms and configurations within each set. All three systems are adequate in this regard, except between WARNING and CAUTION. The most successful result for the proposed format was the DEADLY format that included a skull icon (particularly the black header/red message panel). The consistently high ratings for this configuration suggests that it could be useful in signs whose intent is to keep people at a distance from extreme hazards. Its use on products labels is probably limited (e.g., on toxic pesticides and solvents; high-voltage electrical components).

Overall, the results suggest the need for additional systematic testing of warning configurations to determine people's impressions of them, whether they understand the meanings intended, and their level of effectiveness in eliciting appropriate compliance behavior. These and other studies could facilitate the development of more effective warnings.

REFERENCES

ANSI (1991). *Warning Signs, Labels, Tags, Colors, and Symbol Standards, Z535.1-5.* Washington, DC: National Electrical Manufacturers Association.

Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. *Ergonomics, 37,* 265-275.

FMC. (1985). *Product Safety Sign and Label System.* Santa Clara, CA: FMC Corporation.

Laughery, K.R., Wogalter, M.S., and Young, S.L. (Eds.) (1994). *Human Factors Perspectives on Warnings.* Santa Monica, CA: Human Factors and Ergonomics Society.

Westinghouse (1981). *Product Safety Label Handbook.* Trafford, PA: Westinghouse Printing Division.

Wogalter, M.S., and Silver, N.C. (1990). Arousal strength of signal words. *Forensic Reports, 3,* 407-420.

Wogalter, M.S., and Silver, N.C. (1995). Warning signal words: Connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics,* in press.