Hazard Associations of Warning Header Components

Michael S. Wogalter, Amy B. Magurno, Ann W. Carter,
Julie A. Swindell, William J. Vigilante, and Jason G. Daurity
Department of Psychology
North Carolina State University
Raleigh, NC 27695-7801

ABSTRACT

There has been relatively little warnings research directed at systematically examining the component features comprising existing warning signs as specified in standards (ANSI Z535) and guidelines (e.g., FMC, 1985 ; Westinghouse, 1981). This research examines several elemental features found in real-world warning signs to determine their individual as well as their combined effects on people's hazard perceptions. Various colors, signal words, shapes and configurations—both individually and in combination— comprising existing warning headers as well as newly developed constituents were evaluated in a series of rating and ranking tasks. The results confirmed several existing published recommendations (e.g., the color red is perceived to connote more hazard than other solid colors), but also showed instances where people's perceptions differed from those assumed in design standards and guidelines. Some newly-developed header configurations (e.g., having a skull icon to the right of the signal world) show promise as alternatives for signaling hazardous conditions.

INTRODUCTION

The investigation of warning-related issues has received considerable attention by human factors investigators in recent years. Research suggests that certain characteristics, such as color, the presence of a signal word and icon, can be useful in gaining attention, increasing perceived hazard levels, facilitating comprehension, and motivating compliance (see Laughery, Wogalter, and Young, 1994). However, much of this work has not systematically examined the elements and combinations of elements that comprise warnings.

Previous research on warnings has employed a host of different measures. The most important is the effect warnings have on precautionary behavior. However, behavioral compliance research is very difficult to do (e.g., Wogalter, Godfrey, Fontennelle, Desaulniers, Rothstein, and Laughery, 1987). Because of this, researchers have employed intermediate measures of effectiveness such as knowledge and memory tests and evaluative ratings. Some of these measures have been shown to have utility in predicting behavioral intentions and compliance. One strong predictor of precautionary behavior is perceived hazard level (e.g., DeJoy, 1989; Laughery et al., 1994).

Currently there are disparities in warning standards and guidelines that have not been clarified by research. The present research examines the perceived hazard connoted by different components of warning headers (the upper color panel of warnings containing the signal word) that are currently recommended by various standards and guidelines (e.g., ANSI, 1991; FMC, 1985; Westinghouse, 1981). The 1991 ANSI Z535.2 standard for environmental warning signs specifies different headers for different levels of hazard. The header that is intended to warn about the most hazardous conditions employs the signal word DANGER. The word itself is supposed to be printed in white within a red oval, bordered by white and set on a black background. However, the 1991 ANSI Z535.4 standard for product labels describes another style of header for DANGER. On the left of the mostly red header is an alerting icon (an exclamation point surrounded by a triangle) followed by the word DANGER, both of which are printed in white. Having two (or more) different styles of headers is not necessarily a negative; the negative is the fact that these header standards are based on little or no systematic research. At this point, there is no demonstrated empirical evidence to indicate whether one style of warning is better than another.

Research in this domain is also needed because over time, repeated use of the same style of warning could produce habituation. As a consequence, headers which are intended to capture people's attention and convey an approximate level of hazard may not elicit the response intended. That is, frequent exposure to warning words, colors, or shapes may no longer carry the same meaning they once did. As exposure increases, new combinations of warning elements are necessary to maintain noticeability and abate habituation.

This research examines the hazard levels conveyed by specific warning sign elements. Some of the elements were derived from recommendations in standards and guidelines, and some were created specifically for the present study. Participants rated and ranked several sets of warning sign elements: colors, multiple color combinations, shapes, signal words, and multi-component headers.

METHOD

Participants

A total of 112 individuals participated: 56 were North Carolina State University (NCSU) undergraduates and 56 were shoppers at a Raleigh, North Carolina flea market. The students were given credit in introductory psychology courses. The flea market participants were given a small gift (e.g., a baseball cap, a refrigerator magnet) for taking part in the study. Half of the students were female; 16 of the 56 flea market participants were female. The mean age of the students was 20.8 years (*SD* = 4.7); the mean age of the flea market sample was 43.4 years (*SD* = 12.2; ranging from 13 to 65). Both groups had similar education levels (*M* = 14. 1 years of school, *SD* = 2.2, or in other words, 2 years of college) and ethnic/racial composition: 18% were African-American, 7% Asian, 64% Caucasian, 3% Hispanic, and 8% multi-racial and other ethnic/racial backgrounds.

Stimulus Materials

The stimuli used in this study are listed in Table 1. The five stimulus sets consisted of: (A) 8 solid color bars, (B) 6 multi-color bars, (C) 12 shape-color configurations, (D) 6 signal words, and (E) 7 nonsense word (RESVRE) headers with various shapes and colors. Example black and white representations are shown in Figure 1.

(A) The hues for the colors are based on those specified in ANSI Z535.1 (1991). (B) The multi-color bars were composed of two- and three-color combinations. (C) The shape-color configurations included component shapes (e.g., a simple triangle) as well as shape combinations (e.g., a triangle surrounding an exclamation point) and color (e.g., a red and black oval). The skull icon was included as a shape because recent warning design proposals have suggested its use for extremely severe hazards. (D) The signal words are the hazard- and safety-related terms specified in current ANSI (1991) standards. The term DEADLY was included because earlier work suggests it be used for eminent death situations. The signal words were printed in black capital letters on white backgrounds. (E) The nonsense word configurations combined the elements of shape and color to create entire headers with a meaningless word. Three had color and shape configurations as specified in ANSI Z535.2, (E-48, E-67, and E-44) and two as specified in ANSI Z535.4 (E-63 and E-22). Two other header designs included a skull icon that has been proposed for use with extreme hazards (E-52 and E-13).

All stimuli were printed on plastic cards having the dimensions 5.1 x 25.4 cm (2 x 10 in) except the shape-color configurations which were 6.4 x 11.4 cm (2.5 x 4.5 in).

Procedure

Participants were tested individually and were told that the study concerned people's impressions of various features and elements of warnings.

In the rating task, the experimenter presented the stimulus cards, one at a time. Participants rated each on the hazard level conveyed. Ratings were on a 5-point scale with the numerical and verbal anchors: (0) no hazard, (1) low hazard, (2) moderate hazard, (3) high hazard, and (4) extreme hazard. Participants recorded their answers on a response sheet.

Following the ratings, participants then ranked the stimuli within each set. Participants were instructed to order the cards of a given set by placing the card representing the most hazard further from them, and card representing least hazard closer to them. The resulting card orders were recorded by the experimenter with the highest score assigned to the furthest position. The card in the closest position was always given a score of 1. For example, in a set of six cards, scores ranged from 1 to 6; with a set of 12 cards, scores ranged from 1 to 12. Higher rank scores indicate greater perceived hazard.

Cards within each set were randomized for every presentation of the stimuli (for each participant and every test, i.e., re-randomized for the ranking test). The order of presentation of the first four stimuli sets was determined using a balanced Latin Square. The evaluations of the nonsense word headers were always given last to prevent an asymmetric carryover effect from participants' viewing of these complex multi-component header panels. Thus, the rating and ranking of the nonsense headers was performed after the rating and ranking of the other stimuli.

Following the evaluations, participants filled out a form that requested demographic information, were thanked for their assistance, and given research credit or a choice of gift.

RESULTS

Table 1 shows the mean hazard association value ratings (and standard deviations) for both groups, combined and separately, as well as the combined within-set rankings.

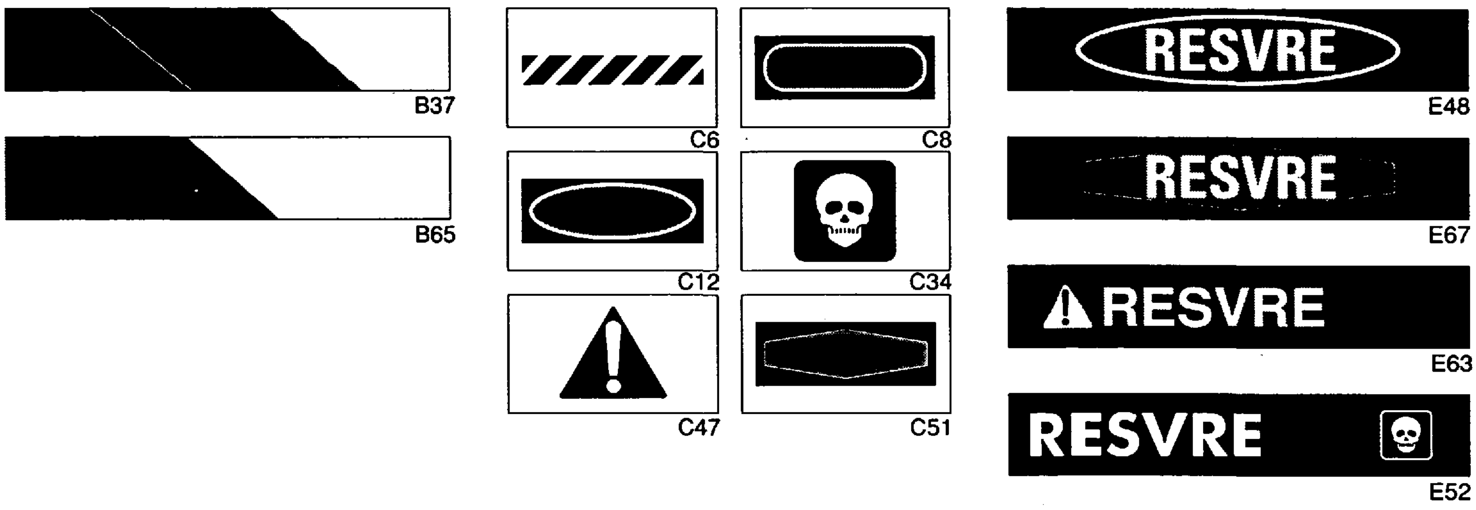
Correlations

Spearman Rho correlations of the overall ratings and rankings within each set were .89 for the multi-colors, .96 for the nonsense word configurations, and 1.0 for the other sets (solid colors, shape-color configurations, and signal words). Because the ratings and rankings showed nearly the identical pattern, only the analyses involving the ratings are described.

Examination of Table 1 shows that the NCSU students and the flea market shoppers corresponded reasonably closely. A Pearson correlation across all of the mean ratings of the two participant groups (using the 39 paired means in Table 1) is .92. Within each set, the correlations are .97, .83, .96, .96, and .92 for the solid colors, multi-colors, shape-color configurations, signal words, and nonsense word configurations, respectively.

Ratings were submitted to separate two-way mixed-model analyses of variance (ANOVAs) that included the two participant groups (NCSU vs. flea market) as the between-subjects factor and the components within each stimuli set as the within-subjects factor. These analyses were followed by paired-comparisons using Tukey's Honestly Significant

Figure 1. Examples of Stimulus Materials Tested



Difference (HSD) Test and simple effects analysis for significant effects with chance probabilities of less than .05.

Solid Color Stimuli

The ANOVA on the solid colors (set A) showed a significant main effect of color, $F\ (7, 770) = 112.81$, $MSe = .842$, $p < .0001$. Comparisons showed that Red was assigned significantly higher hazard ratings than all other colors. Yellow was next and was significantly greater than all other colors. Orange was significantly greater than all of the remaining colors except for Black. Black was significantly greater than all of the other colors. Purple, Green, Blue, and White were lowest and did not significantly differ. The ANOVA also showed a significant effect of participant group, $F\ (1, 110) = 7.82$, $MSe = 3.43$, $p < .01$. The students gave higher hazard ratings ($M = 1.67$) for these stimuli than the flea market shoppers ($M = 1.32$). In addition, the ANOVA also showed a small, but significant interaction, $F\ (7, 770) = 2.44$, $MSe = .842$, $p < .02$. Simple effects analysis showed that the ratings of both groups were consistent except that the students gave significantly higher ratings for black, green, and white than the flea market participants.

Multi-color Stimuli

The ANOVA on the multi-color stimuli (set B) showed only a significant main effect of stimuli, $F(5, 275) = 10.78$, $MSe = .914$, $p < .0001$. Comparisons showed that the Black/White combination was rated significantly lower than all other stimuli in this set. No other difference was significant.

Shape-Color Configurations

The ANOVA on the shape-color configurations (set C) showed a significant main effect of stimuli, $F(11, 1210) = 121.49$, $MSe = .723$, $p < .0001$. Comparisons showed the Skull shape was rated significantly higher than all other shapes. Next were the Red/Black Oval and the Black/Yellow Diagonal Stripes, both of which were significantly higher than all other configurations. This was followed in turn by the Triangle/Exclamation Point, the Orange/Black Elongated Hexagon, the Black/White Diagonal Stripes, and the Black Triangle—each of which was significantly different from each other and all other configurations. The five remaining lower-rated shapes did not significantly differ. The ANOVA also showed a significant main effect of group, $F\ (1, 110) = 4.75$, $MSe = 4.47$, $p < .05$. The flea market participants gave higher hazard ratings ($M = 1.91$) to these stimuli than the students ($M = 1.66$). The ANOVA also produced a significant interaction, $F\ (11, 1210) = 3.87$, $MSe = .723$, $p < .0001$. Simple effects analysis showed that both groups gave consistent ratings to the stimuli except that the flea market participants gave significantly higher ratings than the students to the red/black Oval, the black Elongated Hexagon, the black Oval, and the black Lozenge (capsule shape).

Signal Words

The ANOVA on the signal words (set D) showed a significant main effect of stimuli, $F(5, 550) = 226.49$, $MSe = .554$, $p < .0001$, and a significant interaction, $F(5, 550) = 6.18$, $MSe = .554$, $p < .0001$. The highest to lowest mean ratings were DEADLY, DANGER, WARNING, CAUTION, SAFETY FIRST, and NOTICE. All were significantly different from each other. Simple effects analysis showed that the students rated DEADLY and CAUTION significantly higher than the flea market participants did, whereas the opposite was true for SAFETY FIRST. In addition, the students did not differentiate between the terms WARNING and CAUTION, or between SAFETY FIRST and NOTICE ($ps > .05$), whereas the flea market participants did ($ps < .05$).

Table 1. Mean Hazard Perception Ratings (Overall and by Participant Group), Standard Deviations, and Within-Set Rankings

Set-#	Stimulus	Overall Mean Rating	Overall SD	NCSU	Flea Market	Mean Within-Set Rank
Set A	**Solid color bars**					
A-31	Red	3.15	1.1	3.23	3.07	7.31
A-58	Yellow	2.22	1.1	2.21	2.23	5.68
A-25	Orange	1.96	1.1	2.02	1.89	5.61
A-68	Black	1.74	1.5	2.00	1.48	5.34
A-60	Purple	0.81	1.0	1.00	0.63	3.34
A-64	Green	0.75	1.0	1.07	0.43	3.23
A-70	Blue	0.73	0.9	0.84	0.63	2.99
A-46	White	0.61	0.9	0.98	0.23	2.50
Set B	**Multi-color bars**					
B-69	Black/Yellow	2.25	1.4	2.30	2.20	3.87
B-37	Black/Red/White	2.11	1.2	1.98	2.23	4.04
B-50	Red/White	2.06	1.2	1.91	2.21	3.82
B-49	Black/Orange	1.95	1.1	1.96	1.93	3.58
B-79	Black/White/Red	1.90	1.2	1.71	2.09	3.71
B-65	Black/White	1.39	1.2	1.18	1.61	1.97
Set C	**Shape and color configurations**					
C-34	White Skull in Black Square	3.79	0.6	3.91	3.66	11.67
C-83	Red Oval in Black Rectangle	2.60	1.1	2.39	2.80	9.31
C-80	Black and Yellow Diagonal Stripes	2.57	1.0	2.55	2.59	8.80
C-47	White Exclamation Point in Black Triangle	2.25	1.1	2.43	2.07	8.28
C-84	Orange Elongated Hexagon in Black Rectangle	2.01	1.2	1.84	2.18	8.18
C-6	Black and White Diagonal Stripes	1.70	1.0	1.52	1.88	6.79
C-21	Black Triangle	1.40	1.0	1.30	1.50	5.09
C-51	Black Elongated Hexagon in Black Rectangle	1.14	1.1	0.93	1.36	4.81
C-12	Black Oval in Black Rectangle	1.11	1.1	0.79	1.43	4.17
C-81	Black Lozenge in Black Rectangle	1.00	1.1	0.66	1.36	3.99
C-23	Black Square	0.96	1.0	0.84	1.07	3.46
C-55	Black Circle	0.92	1.0	0.77	1.05	3.46
Set D	**Signal Words**					
D-32	DEADLY	3.81	0.6	4.00	3.63	5.77
D-53	DANGER	3.41	0.6	3.38	3.45	4.99
D-76	WARNING	2.63	0.9	2.61	2.64	3.79
D-35	CAUTION	2.26	0.8	2.48	2.04	3.12
D-11	SAFETY FIRST	1.38	1.1	1.13	1.64	1.68
D-39	NOTICE	1.18	0.8	1.16	1.20	1.65
Set E	**Nonsense word RESVRE headers**					
E-52	White Print & Skull on Red Background	3.74	0.6	3.88	3.61	6.29
E-13	White Print & Skull on Black Background	3.61	0.8	3.70	3.52	5.88
E-63	White Print & Triangle/Exclamation Point on Red Bckgrnd	2.71	1.1	2.95	2.46	4.28
E-48	White Print in Red Oval on Black Background	2.47	0.9	2.30	2.64	3.39
E-22	White Print & Triangle/Exclamation Point on Yellow Bckgrnd	2.41	0.9	2.59	2.23	3.47
E-67	Black Print in Orange Elongated Hexagon on Black Bckgrnd	2.11	1.0	2.04	2.18	2.59
E-44	Yellow Print on Black Background	1.99	0.1	1.86	2.13	2.09

Nonsense Word Headers

The ANOVA on the nonsense word header configurations (set E) showed a main significant effect of stimuli, $F(6, 660) = 76.82$, $MSe = .703$, $p < .0001$. The two Skull headers received significantly higher ratings than all others but not between themselves. The Red/White Triangle/Exclamation Point header was next highest, receiving significantly higher ratings than the remaining headers. The Red/Black Oval header was significantly higher than the other headers except the Yellow/Black Triangle/Exclamation Point. The Yellow/Black Triangle/Exclamation Point was significantly higher than the two lowest-rated headers (Orange/Black Elongated Hexagon and Yellow/Black) which did not differ.

DISCUSSION

This research examined several basic elements and combinations of elements specified in standards and guidelines on warning design. The study also examined several newly-developed configurations proposed as alternatives to existing designs. Some of the findings confirm published design recommendations (e.g., red is perceived as connoting the most hazard), whereas other findings are less supportive. For example, the standards designate the signal word WARNING for situations having greater hazard than the signal word CAUTION, but the present results indicate this difference existed only for the flea market participants and not for the students. Group differences may influence the perception of hazard attached to these terms. The flea market participants were older, more likely to have jobs in hazardous settings, and probably have been exposed to more warnings.

The term DEADLY was rated significantly higher than all the other signal words. This confirms previous research that has evaluated alternative signal words (e.g., Wogalter and Silver, 1990, in press). Because of its high level of perceived hazard level and its infrequent use in warnings, it should be effective in signaling extremely hazardous situations. Its newness should attract attention of viewers who are overly familiar with the term DANGER.

Although the finding that red is associated with the highest level of hazard is supportive of ANSI Z535, other color-related results only partially support the current standard. Yellow was found to connote greater hazard than orange, whereas the standards assign orange to greater level hazards than yellow. In the multi-color stimulus set, the black/yellow combination elicited the highest hazard ratings (although it was not significantly different from the other stimuli except the black/white card). The ratings of these stimuli and the color-shape combinations indicate that the presence of color increases people's evaluation of hazard.

The skull icon connoted the highest hazard level of all shapes in the study. This result supports its possible use in signaling life-threatening hazards. With the exception of the skull, the black/white component shapes were generally associated with relatively low levels of hazard. However, due to their inherent nature or to familiarity with current warnings, certain shapes were associated with somewhat higher hazard levels than others. The black/white diagonal stripes and the triangle were rated as moderately hazardous. The triangle combined with the exclamation point was perceived slightly above that level. Also, the addition of color increased perceived hazard for some shapes (i.e., increases were found with the diagonal stripes, the oval, the elongated hexagon, and the lozenge). Thus, shape and color have an additive or interactive effect when combined.

The results of the more complex configurations, using shape, color, and a nonsense word as components, also show that a newly-developed header with a skull configuration was perceived as having significantly greater connoted hazard than any of the header configurations presently specified in standards. For example, ANSI Z535.2 designates the red/black oval configuration to be the highest level hazard. In this study, the ANSI Z535.4 product label version (comprised of a triangle/exclamation mark, red color, and signal word) surpassed Z535.2 configuration.

The overall mean ratings for the various types of stimuli demonstrate that only a few elements reached an average rating above 3.0 (high hazard). These were: (a) the color red, used alone or with the skull in the nonsense word header configuration; (b) the skull, used alone or in the nonsense word header configuration in either the red or black background; (c) and the signal words DEADLY and DANGER. We would expect that substitution of these two signal words for the nonsense word in the header configurations would raise perceived hazard to higher levels.

The present results provide empirical evidence that changes to the standards regarding some of the components and combinations of components could improve their effectiveness in conveying levels of hazard. For example, it is clear that people do not consistently differentiate between the signal words WARNING and CAUTION (or the colors orange and yellow) so it does not make sense to use these elements to indicate different hazard levels.

Lastly, these results point to the need for testing warnings to determine people's impressions of them and whether they differentiate or understand the meanings intended. Research could also determine the kinds of stimuli that cover the entire range of hazard levels, not just some parts of the hazard dimension. These and other studies could assist in the development of a more effective warning system.

ACKNOWLEDGMENTS

The authors would like to thank Blair Brewster of Electromark Co., Wolcott, New York, for encouragement and financially support. Also, thanks to Barbara L. Glover, La Tondra A. Murray, Michael A. Ettefagh, and James A. Keen for assisting in the data collection.

REFERENCES

ANSI (1991). *Sign, Color and Symbol Standards. Z535.1-5*. Washington, DC: NEMA.

DeJoy, D.M. (1989). Consumer Product Warnings: Review and Analysis of Effectiveness Research," *Proceedings of the Human Factors Society, 33*, 936-940.

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., Godfrey, S.S., Fontenelle, G.A., Desaulniers, D.R., Rothstein, P.R., and Laughery, K.R. (1987). Effectiveness of warnings. *Human Factors, 29*, 599-612.

Wogalter, M. S., and Silver, N. C. (1990). Arousal strength of signal words. *Forensic Reports, 3*, 407-420. 88.

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