INTERNATIONAL JOURNAL OF COGNITIVE ERGONOMICS, 2(1–2), 123–143

Hazard Level Perceptions of Warning Components and Configurations

Michael S. Wogalter
North Carolina State University

Michael J. Kalsher
Rensselaer Polytechnic Institute

Linda J. Frederick
West Virginia University

Amy B. Magurno
North Carolina State University

Blair M. Brewster
Electromark Co.

ABSTRACT

The hazard level perceptions by participants of existing and alternative warning signs and labels were examined in three experiments. Participants included undergraduates, community volunteers from public shopping areas, and industrial workers who evaluated various warning elements (shapes, colors, icons, signal words), as well as different combinations of these elements on their perceived hazard level. Results confirmed portions of existing warning standards and guidelines, for example, the color red connoted greater hazard than other colors and the word *DANGER* connoted a higher level of hazard than either *WARNING* or *CAUTION.* In other instances, participants' interpretations were different from the hazard denotations promulgated in current standards. Some alternative designs such as a skull icon and the signal word *DEADLY* show promise as alternatives for signaling unsafe conditions.

1. INTRODUCTION

In recent years, the design of warnings has received considerable attention. Research indicates that warning design characteristics such as color, shape, and the presence of a signal word, pictorials and icons can enhance conspicuity, increase perceived hazard levels, facilitate comprehension, and motivate compliance (e.g., Laughery, Wogalter, & Young, 1994; Wogalter & Laughery, 1996). Such research has contributed substantially to our

Requests for reprints should be sent to Michael S. Wogalter, Psychology Department, 640 Poe Hall, Campus Box 7801, North Carolina State University, Raleigh, NC 27695–7801.

empirical knowledge on what design characteristics improve warnings. What is still needed, however, is systematic research aimed at assessing the relative effectiveness of the specific elements and combinations of elements that comprise warnings.

It is noteworthy that most of the design characteristics described in published standards and guidelines (American National Standards Institute [ANSI], 1991; FMC Corp., 1985; Westinghouse Inc., 1981) are not based on empirical research. When these standards and guidelines were compiled, specific research to benefit decision making was not available. Consequently, some of the recommendations that have been promulgated may not produce the most effective warnings. Alternative warning designs that make use of nontraditional signal words, color combinations, and configurations might better signal hazardous conditions. In addition, the availability of alternative configurations having similar hazard connotation as existing warnings may be useful in retarding habituation resulting from repeated exposure to similar warning designs.

The ANSI is a private, nonprofit organization whose purpose is to promote and facilitate voluntary consensus standards and conformity assessment systems to help U.S. businesses be competitive. ANSI standards for warnings have five parts: color (Z535.1); environmental signs (Z535.2); pictorials, icons, and symbols (Z535.3); product labels (Z535.4); and tags (Z535.5). In addition, both ANSI (1991) and FMC (1985) recommend the use of three signal words to convey decreasing levels of personal injury hazard, that is, *DANGER, WARNING,* and *CAUTION,* in that order. Although research indicates that people perceive DANGER as being higher in connoted hazard than WARNING and CAUTION, they do not readily distinguish between the latter two words (e.g., Wogalter & Silver, 1990, 1995).

Moreover, discrepancies exist in the ANSI (1991) warning standards that have not yet been clarified by research. For example, the signal word headers in the sign and label standards, ANSI Z535.2 and Z535.4, differ in appearance. For environmental warning signs, ANSI Z535.2 recommends that the word DANGER be printed in white within a red oval that has a white border and which is set on a black background. In contrast, for product labels ANSI Z535.4 defines another header style for the word DANGER that contains an alert symbol (an exclamation point surrounded by a triangle) to the left of the word DANGER, both printed in white on a red panel. Although the use of more than one style of header is not necessarily problematic, the fact that these header standards were established based on little or no systematic empirical research clearly is. In short, we do not know whether one style is superior to another or whether headers within a style convey the intended level of hazard.

Previous research on warnings has employed a host of different measures, the most important of which is the effect warnings have on precautionary behavior. Behavioral compliance research is extremely difficult to do, however, because of situational, cost, and ethical constraints (e.g., Wogalter et al., 1987). One cannot expose research participants to actual hazards in the course of carrying out a study. Therefore, behavioral compliance studies must use special procedures to make the situation appear potentially hazardous while not actually exposing the participants to any real danger. Such circumstances are often challenging to construct and laborious to conduct. Consequently, researchers have employed intermediate measures of effectiveness, such as knowledge, memory tests, and evaluative ratings. Although relatively little research has related these measures to actual precautionary behavior, there is research that shows a reasonably strong relation to a presumed precursor of behavior, and that is *behavioral intentions.* Research in other domains has shown that behavioral intentions relate to actual behavior when the situation (or scenario) in which the behavioral intention measure is evaluated is similar to the behavioral one (Ajzen & Fishbein, 1977). In the warnings literature, one particularly strong predictor of intended precautionary behavior is *perceived hazard* (e.g., DeJoy, 1989; Laughery et al., 1994; Wogalter, Brelsford, Desaulniers, & Laughery, 1991). Because perceived hazard has been found to relate to behavioral intentions such as intended carefulness and willingness to comply, it was used as the main response measure in the research presented here.

Three experiments are presented in which participants provided estimates of the extent of hazard they associated with various colors, words, and configurations for: (a) warnings constructed according to existing standards, and (b) alternative warnings designed specifically for this research.

In Experiment 1, participants provided hazard ratings for individual components and combinations of components of warnings (e.g., color, signal words, shapes, and configurations) using existing as well as several alternative designs. In Experiment 2, complete sign configurations were assessed. The participants in these two experiments included both college students and individuals recruited at public shopping areas. Experiment 3 evaluated warning stimuli using participants from an industrial worker population.

2. EXPERIMENT 1

The goals of the first experiment were to examine the hazard levels conveyed by specific warning elements and to assess whether participants' hazard perceptions are consistent with the level of hazard intended by existing warning guidelines. The participants' task was to rate and rank, individually and in combination, the level of hazard conveyed by several sets of warning sign components, including colors and color combinations, shapes, and multicomponent headers. The specific warning components investigated were either derived from existing warning standards and guidelines (e.g., ANSI, 1991; FMC Corp., 1985; Westinghouse, 1981) or they were selected based on their use in other kinds of safety signs and labels or in published research. In addition, sets of alternative formats were examined to determine their relative standing compared to existing ones. In some cases, the term *RESVRE*, a meaningless combination of letters (a nonsense term), was inserted in the position normally occupied by a signal word in order to evaluate multicomponent header configurations (e.g., color, shape) without the effect that an actual signal word might produce.

2.1. Method

2.1.1. Participants. A total of 112 individuals participated; half were North Carolina State University undergraduates and half were shoppers at a Raleigh, North Carolina flea market. The students received course credit for their participation; flea market participants received a small gift, such as a baseball cap or a refrigerator magnet. Half of the students were female; 16 of the 56 (29%) flea market participants were female. The mean age of students was 20.8 years (SD = 4.7); the mean age of the flea market group was 43.4 years (SD = 12.2; ranging from 13 to 65).

2.1.2. Materials. The five sets of stimuli used in this study are described in Table 1. The stimulus sets consisted of (a) eight solid color bars, (b) six multicolor bars, (c) 12 shape–color configurations, (d) six signal words, and (e) seven nonsense word (RESVRE) headers with various shapes and colors. All stimuli were printed on plastic cards having the dimensions 5.1 cm × 25.4 cm (2 in. × 10 in.) except the shape–color configurations, which were 6.4 cm × 11.4 cm (2.5 in. × 4.5 in.). The participants' evaluations of these stimuli were tracked using label codes (e.g., A–31) that contained a letter name of the set and a randomly assigned number which was printed in 12-point san serif font on the lower right side of the

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

Set-#	Stimulus	Overall *M*	Overall *SD*	*M* Undergrad	*M* Comm Vol	Rankings Within-Set
Set A	Solid colors					
A-31	Red	3.2	1.1	3.2	3.1	1.7
A-58	Yellow	2.2	1.1	2.2	2.2	3.3
A-25	Orange	2.0	1.1	2.0	1.9	3.4
A-68	Black	1.7	1.5	2.0	1.5	3.7
A-60	Purple	0.8	1.0	1.0	0.6	5.7
A-64	Green	0.8	1.0	1.1	0.4	5.8
A-70	Blue	0.7	0.9	0.8	0.6	6.0
A-46	White	0.6	0.9	1.0	0.2	6.5
Set B	Multi colors					
B-69	Black/Yellow	2.3	1.4	2.3	2.2	3.1
B-37	Black/Red/White	2.1	1.2	2.0	2.2	3.0
B-50	Red/White	2.1	1.2	1.9	2.2	3.2
B-49	Black/Orange	2.0	1.1	2.0	1.9	3.4
B-79	Black/White/Red	1.9	1.2	1.7	2.1	3.3
B-65	Black/White	1.4	1.2	1.2	1.6	5.0
Set C	Shape and color configurations					
C-34	White skull in black square	3.8	0.6	3.9	3.7	1.3
C-83	Red oval in black rectangle	2.6	1.1	2.4	2.8	3.7
C-80	Black/yellow diagonal stripes	2.6	1.0	2.6	2.6	4.2
C-47	White ! in black triangle	2.3	1.1	2.4	2.1	4.7
C-84	Orange elongated hexagon in black rectangle	2.0	1.2	1.8	2.2	4.8
C-6	Black/white diagonal stripes	1.7	1.0	1.5	1.9	6.2
C-21	Black triangle	1.4	1.0	1.3	1.5	7.9
C-51	Black elongated hexagon in black rectangle	1.1	1.1	0.9	1.4	8.2
C-12	Black oval in black rectangle	1.1	1.1	0.8	1.4	8.8
C-81	Black capsule (lozenge shape) in black rectangle	1.0	1.1	0.7	1.4	9.0
C-23	Black square	1.0	1.0	0.8	1.1	9.5
C-55	Black circle	0.9	1.0	0.8	1.1	9.5
Set D	Signal words					
D-32	DEADLY	3.8	0.6	4.0	3.6	1.2
D-53	DANGER	3.4	0.6	3.4	3.5	2.0
D-76	WARNING	2.6	0.9	2.6	2.6	3.2
D-35	CAUTION	2.3	0.8	2.5	2.0	3.9
D-11	SAFETY FIRST	1.4	1.1	1.1	1.6	5.3
D-39	NOTICE	1.2	0.8	1.2	1.2	5.4
Set E	Nonsense word RESVRE headers					
E-52	White print & skull on red background	3.7	0.6	3.9	3.6	1.7
E-13	White print & skull on black background	3.6	0.8	3.7	3.5	2.1
E-63	White print & triangle / ! on red background	2.7	1.1	3.0	2.5	3.7
E-48	White print in red oval on black background	2.5	0.9	2.3	2.6	4.6
E-22	White print & triangle / ! on yellow background	2.4	0.9	2.6	2.2	4.5
E-67	Black print in orange elongated hexagon on black background	2.1	1.0	2.0	2.2	5.4
E-44	Yellow print on black background	2.0	0.1	1.9	2.1	5.9

Note. Undergrad = undergraduates; Comm Vol= community volunteers.

cards. Response sheets were provided to participants with numbered blanks. Sample black and white representations of the stimuli are presented in Figure 1.

The colors in Set A are based on ANSI Z535.1 (ANSI, 1991) specifications. The multicolor bars (Set B) were comprised of two- and three-color combinations. The 12 shape–color configurations (Set C) included component shapes (e.g., a simple triangle) as well as shape combinations (e.g., a triangle surrounding with an exclamation point) and color (e.g., a red and black oval). The skull icon (C–34) was included because proposals to the ANSI committee have suggested it be used for signs and labels that warn of extreme hazards (Brewster, 1995). The signal words (Set D) are the safety terms specified in current ANSI (1991) standards. DANGER, WARNING, and CAUTION are terms intended to convey high to low levels of potential personal hazard, respectively. *NOTICE* is intended to be used to communicate company policy that relates directly or indirectly to the safety of personnel or protection of property. *SAFETY FIRST* is intended to indicate general instructions related to safe work practices, procedures, or location of safety equipment. NOTICE and SAFETY FIRST are not considered by ANSI (1991) to be part of the three-tier personal injury hazard classification system (i.e., DANGER, WARNING, and CAUTION). Although not an ANSI term, *DEADLY* was included because research (e.g., Wogalter & Silver, 1990, 1995) indicates that it might connote substantially greater hazard than ANSI's highest level term (DANGER). The signal words were printed in black capital letters on white backgrounds. The seven nonsense word configurations (Set E) combined shapes and colors to create entire headers with a meaningless signal word (RESVRE). 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 (E–52 and E–13).

2.1.3. Procedure. In the *rating* task, the experimenter presented each stimulus card, one at a time, to participants who judged each card for the level of hazard it connoted. Ratings were on a 5-point scale with the following 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.

Participants also gave within-set hazard rankings of the stimuli. They arranged the cards of each set by placing the one representing the greatest hazard farthest from them and the one representing the least hazard closest to them and ordering the others between the two extremes. The experimenter recorded the resulting card arrangement with a rank order of 1 being assigned to the farthest position and the card closest to the participant being given the highest rank score of the set.

Half of the participants did the rating task first and half did the ranking task first. Cards within each set were randomized for every set of stimuli presented in both the rating and ranking tasks. A balanced Latin Square was used to determine the order of presentation of the first four stimulus sets. The rating and ranking evaluations of the nonsense word headers (Set E) was always performed last in the procedure to avoid possible carryover effects from participants' viewing these complex multicomponent headers prior to the simpler configurations. Upon completing the study, participants were debriefed, thanked, and dismissed.

2.2. Results

Table 1 shows the mean hazard perception ratings (and standard deviations) for both groups of participants, combined and separately, as well as the combined within-set rankings. Higher hazard ratings and lower hazard rankings indicate greater perceived hazard.

Spearman Rho correlations of the overall ratings and rankings within each set were –.89 for the multicolors, –.96 for the nonsense word configurations, and –1.0 for the other sets (solid

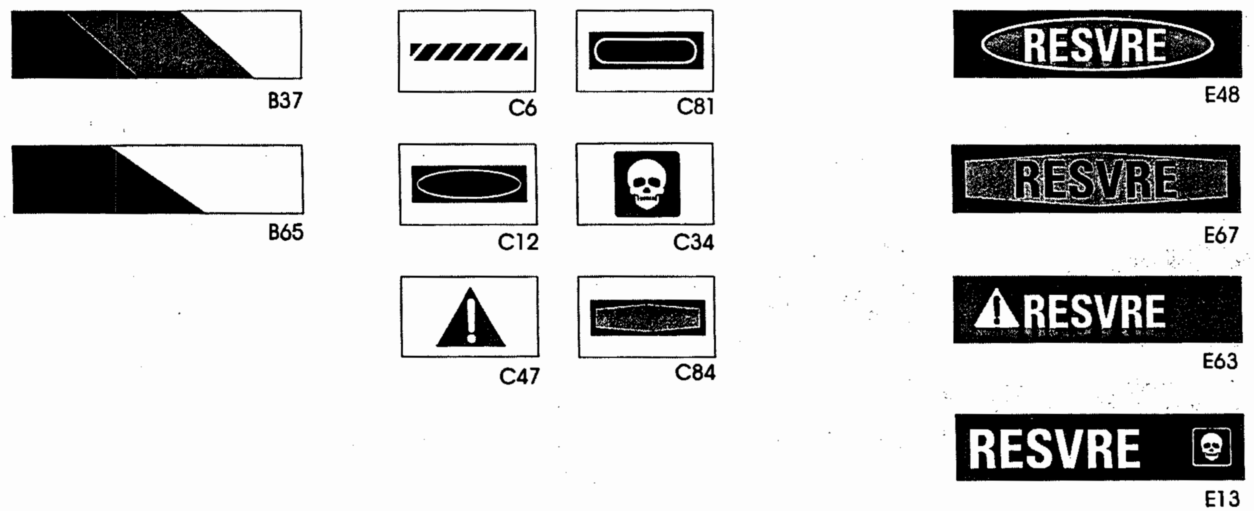

FIGURE 1: Examples of stimulus materials tested. Descriptions are provided in Table1.

colors, shape–color configurations, and signal words). Because the ratings and rankings showed nearly identical patterns, only the analyses involving the ratings are described below.

Examination of Table 1 shows that the undergraduates' and the community volunteers' responses are very similar. A Pearson correlation across all of the mean ratings of the two participant groups (using the 39 paired means in Table 1) was .92. Within each set, the correlations were .97, .83, .96, .96, and .92 for the solid colors, multicolors, shape–color configurations, signal words, and nonsense word configurations, respectively.

Ratings were next submitted to separate two-way mixed-model analyses of variance (ANOVAs) that included the two participant groups (undergraduates vs. community volunteers) as the between-subject factor and the components within each of the five stimulus sets as the within-subjects factor. These analyses were followed by paired comparisons using Tukey's Honestly Significant Difference (HSD) Test and simple effects analysis for significant effects with p levels less than .05.

2.2.1. 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 remaining colors. Orange was significantly greater than all of the remaining colors except for black. Black was significantly greater than all of the remaining colors. Purple, green, blue, and white were lowest and did not differ significantly. There was a significant effect of participant group, $F(1, 110) = 7.82$, $MSe = 3.43$, $p < .01$. The undergraduates gave higher hazard ratings ($M = 1.67$) than the community volunteers ($M = 1.32$). In addition, the ANOVA also showed a significant interaction, $F(7, 770) = 2.44$, $MSe = .842$, $p < .02$. Simple effects analysis showed that the ratings by both groups were consistent except that the undergraduates rated black, green, and white significantly higher than the community volunteers.

2.2.2. Multicolor stimuli. The ANOVA on the multicolor 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.

2.2.3. Shape–color configurations. Stimuli showed a significant main effect of the shape–color configurations (Set C), $F(11, 1210) = 121.49$, $MSe = .723$, $p < .0001$. Comparisons showed that the skull icon was rated significantly higher than all other shapes. Next were the Red Oval in Black Rectangle and the Black–Yellow Diagonal Stripes, both of which were significantly higher than all remaining configurations. This was followed in turn by the White Exclamation Point in Black Triangle, the Orange Elongated Hexagon in Black Rectangle, 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 differ significantly. The ANOVA also showed a significant main effect of participant group, $F(1, 110) = 4.75$, $MSe = 4.47$, $p < .05$, with the community volunteers giving higher hazard ratings ($M = 1.91$) than the undergraduates ($M = 1.66$). Moreover, the ANOVA revealed a significant interaction, $F(11, 1210) = 3.87$, $MSe = .723$, $p < .0001$. Simple effects analysis indicated that both groups gave consistent ratings to the stimuli except that the community volunteers gave significantly higher ratings than the undergraduates to the Red Oval in Black Rectangle, the Orange Elongated Hexagon in Black Rectangle, the Black Oval in Black Rectangle, and the Black Capsule (lozenge shape).

2.2.4. 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 undergraduates rated DEADLY and CAUTION significantly higher than the community volunteers did, whereas the opposite was true for SAFETY FIRST. In addition, the undergraduates did not differentiate between the terms WARNING and CAUTION, or between SAFETY FIRST and NOTICE ($ps > .05$), whereas the community volunteers did ($ps < .05$).

2.2.5. 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 headers with the skull icon (E–52 and E–13) received significantly higher ratings than all others but did not differ between themselves. The White Print and Triangle–Exclamation Point on a Red Background (E–63) was next highest, receiving significantly higher ratings than the remaining headers. The White Print in Red Oval on Black Background (E–48) was significantly higher than the other headers except the White Print and Triangle–Exclamation Point on a Yellow Background (E–22). This stimulus in turn was significantly higher than the two lowest rated headers (Black Print in an Orange Elongated Hexagon on Black Background, E–67, and Yellow Print on Black Background, E–44), which did not differ between themselves.

2.3. Discussion

Experiment 1 examined several basic components and combinations of components specified in standards and guidelines on warning design. This study also examined several newly developed configurations proposed as alternatives to existing designs. Some of the findings confirmed published design recommendations. For example, red was perceived as connoting greater hazard than any other solid color. However, other findings were less supportive. For example, ANSI standards designate the signal word WARNING for situations with greater hazard than the signal word CAUTION, but these results indicate this difference existed only for the community volunteers and not for the undergraduates.

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 & Silver, 1990, 1995). The term DEADLY could be a highly effective signal for extremely dangerous situations because of its preexisting high-hazard connotation. Its use in signs is more likely than on consumer product labels because manufacturers might want to avoid using the word DEADLY on their products due to the possibility of lost sales. Exceptions to this might be certain types of products such as toxic pesticides and solvents or high voltage electrical components that are sold at local hardware stores. In addition, the term DEADLY is likely not to be overused, so its relatively infrequent appearance may better attract viewers' attention than the more familiar and more commonly used term, DANGER.

Although the finding that red is associated with the highest level of hazard supports ANSI (1991) Z535, other color-related results only partially supported the standard. Yellow, for example, was found to connote greater hazard than orange, whereas the standards assign orange to a hazard level higher than yellow. In the multicolor 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 other colored stimuli in this study indicate that the presence of hues, other than black and white, 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 extreme 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 participants' familiarity with current warnings, certain shapes were associated with somewhat higher levels of hazard than others. The Black–White Diagonal Stripes and the Black Triangle were rated as moderately hazardous. The triangle combined with the exclamation point was perceived to connote a hazard level slightly above that. The addition of color increased perceived hazard for some shapes. For example, increases due to color were found with the diagonal stripes, the oval, and the elongated hexagon. This suggests that shape and color appear to have an additive or interactive effect when combined.

The evaluations of the more complex configurations, involving various shapes and colors 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. Also, the configuration resembling the highest level ANSI Z535.4 product label header (comprised of a Triangle–Exclamation point and red color) surpassed the configuration resembling the highest Z535.2 sign header (comprised of the red–black oval).

It is worth noting that several analyses produced significant interactions, revealing differences in hazard perceptions between the student and community volunteers for some of the stimuli tested, but not others. Most of these disparities were magnitude differences, not changes in the relative ordering of the terms. However, the finding of interactions has at least two important implications. First, they confirm the importance of ensuring a broad representation among study participants. Second, they suggest that it is important to select warning components so that the levels of hazard that they convey have a shared meaning among broad segments of the population in which they will be deployed.

Finally, the overall mean ratings for the various types of stimuli highlight the fact 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; and (c) the signal words DEADLY and DANGER. These elements used separately or in combination could be employed to enhance the connoted hazard of a warning sign or label.

The primary goal of Experiment 1 was to examine the relative effectiveness of specific components of warnings or combinations of these components. The goal of the next experiment is to evaluate the effectiveness of intact sign and label configurations.

3. EXPERIMENT 2

The objective of Experiment 2 was to compare people's hazard level perceptions of three sets of complete warning styles or formats. Thus, instead of a predominate focus on individual elements as in Experiment 1, this experiment examined combinations of elements that are more representative of actual signs and labels except that the message text was absent. In addition to the hazard ratings, the stimuli were evaluated on noticeability. Two of the sets were constructed according to existing standards for environmental signs (ANSI Z535.2) and for product labels (ANSI Z535.4); the third consisted of a newly developed set of designs proposed as an alternative to the existing standards.

3.1. Method

3.1.1. Participants. Two groups of volunteers served as participants. One group consisted of 36 undergraduates from Rensselaer Polytechnic Institute, 16 men and 20 women, ranging in age from 18 to 23 years with a mean age of 20.1. The other group was comprised of 124 community volunteers from shopping malls in upstate New York and western Massachusetts. The nonstudents, 65 men and 59 women, ranged in age from 18 to 84 years with a mean age of 42.6.

3.1.2. Materials. Gray scale representations of the warning stimuli are shown in Figure 2. Table 2 lists some of the specific characteristics of the warning stimuli. Stimuli consisted of 16 cards displaying 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 cm × 17.8 cm; 5 in. × 7 in.) and were constructed to resemble existing warning designs specified in current standards (ANSI Z535.2 and Z535.4, 1991) or they were newly constructed alternative designs. The cards differed on one or more of the following dimensions: signal word; color of foreground figures and print and the background header and message panels; capitalization (e.g., DANGER vs. Danger; all caps vs. small caps); letter size and boldness; left vs. center justification; and the presence of shape borders or icons such as a triangle–exclamation point or a skull.

The five cards in the first set were consistent with ANSI Z535.2 standards for environmental sign warnings having the following 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 alternate warning headers. Four of the cards had one of the following signal words DANGER, WARNING, CAUTION, and NOTICE; the other three used variants of the term DEADLY. The alternate headers differed from their counterparts in the first two sets in several ways: (a) 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, that is, small caps; (b) there was somewhat more color space, for example, more red in the DANGER header; (c) all text was left justified; and (d) no shape or icon was used in the headers for the four conventional signal words. The alternate set also contained three signs with the term DEADLY in the header. Previous research (e.g., Wogalter & Silver, 1990, 1995) as well as the results of Experiment 1 have shown that DEADLY consistently receives higher hazard ratings than DANGER, the highest level conventional term. Two had DEADLY in all capital letters (in modern small caps style with slightly smaller letters after the first letter); 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. These three headers also included a skull icon.

3.1.3. Procedure. The experiment was conducted in four distinct parts that were counterbalanced in their presentation order across participants. In one part, participants were asked to examine the three sets of cards, one set at a time, and to rate each card on two dimensions: (a) the level of hazard conveyed and (b) how *attention getting* or *noticeable* each warning was. Hazard and noticeability ratings were made on 5-point scales. The numerical and verbal anchors for the hazard scale were: 0 = *no hazard*, 1 = *low hazard*, 2 = *moderate hazard*, 3 = *high hazard*, and 4 = *extreme hazard*. The anchors for the attention getting scale were: 0 = *not at all attention getting*, 1 = *somewhat attention getting*, 2 = *moderately attention getting*, 3 = *highly attention getting*, and 4 = *extremely attention getting*.

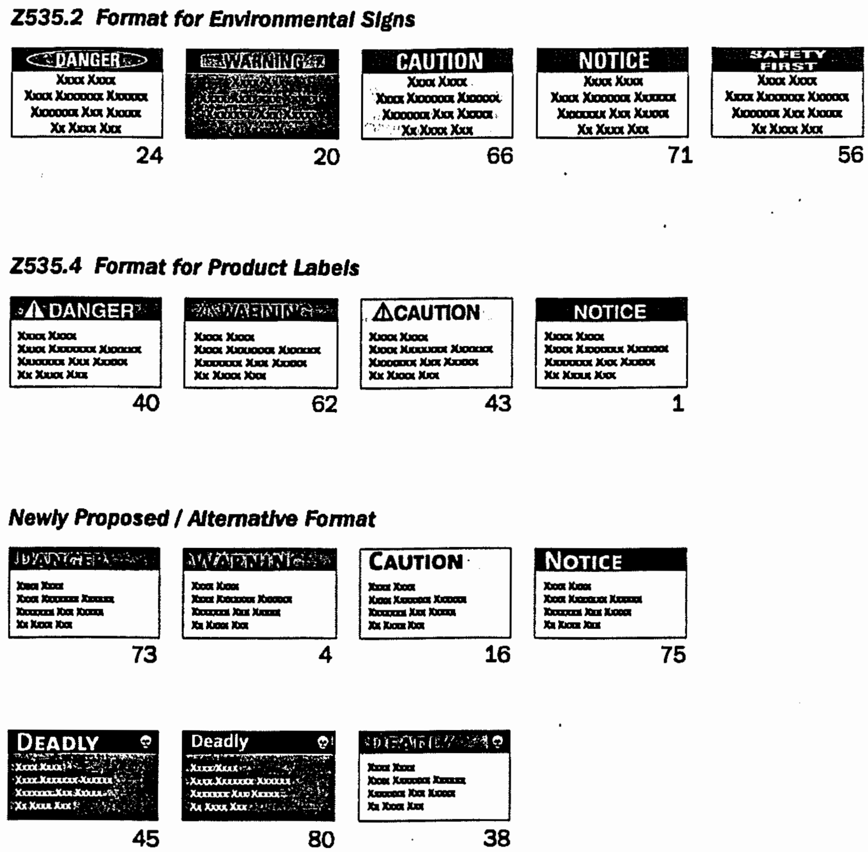

FIGURE 2: Representations of ANSI Z535.2, ANSI Z535.4, and newly constructed warning configurations. See descriptions in Table 2.

A second part involved within-set hazard rankings. Participants were asked to sort the cards in each set according to the hazard level conveyed. They 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. The experimenter recorded the card order on a separate response sheet.

A third part involved having participants rate the three different formats for each signal word. Fifteen of the 16 cards (the card with SAFETY FIRST was omitted) were recombined to form five new sets (3 cards per set), based on the signal word in their headers. These sets can be seen in Table 4. Each contains the three variants for each of the following signal words: DANGER, WARNING, CAUTION, NOTICE, and DEADLY. Participants were asked to examine the five sets of cards, one set at a time, and to rate each card using the procedure and rating scales described in Phase 1.

A fourth part consisted of hazard rankings of different formats for each signal word. Participants rank ordered the three formats of each signal word.

For all procedures, presentation orders were randomly determined both for the order of the sets received and the card order within each set. Half of the participants did the hazard evaluations first followed by the noticeability evaluations, and half did them in the reverse order. Order was also counterbalanced for the tasks of rating and ranking the stimuli. After participants completed the tasks, they were debriefed, thanked, and dismissed.

3.2. Results

Table 3 shows the overall mean hazard ratings, separate mean hazard ratings for the undergraduates and community volunteers, overall within-set hazard rankings, and overall noticeability ratings for the three sets of stimuli, that is, ANSI Z535.2, Z535.4, and the alternate format. Higher hazard ratings and lower hazard rankings indicate greater perceived hazard. Higher noticeability ratings indicate greater perceived attention gettingness. Table 3 shows that same basic pattern of results for all measures. A Pearson correlation between the overall hazard ratings and noticeability ratings was .99, $p < .00001$. The Spearman Rho correlation between the hazard ratings and hazard rankings was –.71 and between the hazard rankings and the noticeability ratings was –.75, $ps < .01$. Analyses involving the hazard rankings and the noticeability ratings showed virtually the

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

			Header Color		*Message Panel Color*	
#	*Signal Word*	*Header Shape or Icon*	*Print*	*Background*	*Print*	*Background*
ANSI Z535.2: 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	none	yellow	black	black	yellow
71	NOTICE	none	white	blue	black	white
56	SAFETY FIRST	none	white	green	black	white
ANSI Z535.4: Product labels						
40	DANGER	red ! in 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	none	white	blue	black	white
Proposed formats						
73	DANGER	none	white	red	black	white
04	WARNING	none	black	orange	black	white
16	CAUTION	none	black	yellow	black	white
75	NOTICE	none	white	blue	black	white
45	DEADLY	white skull in square	white	black	white	red
80	Deadly	white skull in square	white	black	white	red
38	DEADLY	white skull in square	white	black	red	white

TABLE 3
Mean Hazard Ratings, Within-set Rankings, and Noticeability Ratings for ANSI Z535.2, ANSI Z535.4, and Proposed Formats (Experiment 2)

# Signal word	*Hazard Rating* Overall	Undergrad	Comm Vol	*Hazard Ranking Overall*	*Rating Noticeability*
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: 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
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

Note. Undergrad = undergraduates; Comm Vol = community volunteers.

same pattern of significant effects as the hazard ratings. Because of the similarity among these measures, only the analyses of the hazard ratings are described.

Separate two-way Participant Group (undergraduate versus community volunteer) × Signal Word ANOVAs were conducted on the hazard ratings within each format. Significant effects were followed by simple effects analysis and pairwise comparisons using Tukey's HSD Test. Only effects having p values of less than .05 are described.

3.2.1. Comparison of ratings within each format. The ANOVA on the hazard ratings of the ANSI Z535.2 set showed a significant main effect of signal word, $F(4, 632) = 145.17$, $MSe = .57$, $p < .00001$. 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. The ANOVA also showed a significant interaction, $F(4, 632) = 4.75$, $MSe = .57$, $p < .001$. Both participant groups were consistent except that the community volunteers rated DANGER higher than the undergraduates, and the undergraduates rated SAFETY FIRST higher than the community volunteers.

The ANOVA on the ANSI Z535.4 product labels showed a significant main effect of signal word, $F(3, 474) = 225.43$, $MSe = .41$, $p < .00001$. 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 participant 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, community volunteers gave higher ratings than undergraduates but this difference was due mainly to the community volunteers rating DANGER and WARNING higher than the undergraduates did.

The ANOVA on the alternative and proposed warnings showed a significant main effect of signal word, $F(6, 948) = 218.77$, $MSe = .44$, $p < .00001$. 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 warnings in this set were significantly different from each other, except between (a) DEADLY (small caps) and Deadly (mixed case)—both with the black header–red message panels; (b) the mixed case Deadly with the black header–red message panel 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 community volunteers rated DEADLY with the red header–white message panel and DANGER significantly higher than the undergraduates did.

3.2.2. Comparison of configurations with the same signal word. Table 4 shows the measures derived from the ratings and rankings from the third and fourth parts, in which participants directly compared the three formats (Z535.2, Z535.4, and alternative) having the same signal word. As with previous 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 ANOVA for DANGER showed only a significant effect of participant group, $F(1, 157) = 16.01$, $MSe = 1.17$, $p < .0001$. Community volunteers gave higher ratings to these terms than undergraduates.

TABLE 4
Mean Hazard Ratings, Within-set Rankings, and Noticeability Ratings Across ANSI Z535.2, ANSI Z535.4, and Proposed Formats (Experiment 2)

#	*Signal word*	*Hazard Rating* *Overall*	*Undergrad*	*Comm Vol*	*Hazard Ranking Overall*	*Rating Noticeability*
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

Note. These stimuli are described in Table 3.

The ANOVA for WARNING showed a significant effect of format, $F(2, 314) = 4.86$, $MSe = .36$, $p < .01$. Comparisons showed that the ANSI Z535.2 version was rated significantly higher than the alternative version..

The ANOVA for CAUTION showed a significant effect of format, $F(2, 314) = 36.80$, $MSe = .28$, $p < .00001$. Comparisons showed that the ANSI Z535.2 CAUTION was rated significantly higher than the ANSI Z535.4 version, which in turn was rated significantly higher than the alternative version. The ANOVA also showed a significant interaction, $F(2, 314) = 5.43$, $MSe = .28$, $p < .01$. The means show a crossover pattern. The undergraduates rated the Z535.2 version higher than the community volunteers, whereas the community volunteers rated the alternative version higher than the undergraduates, but neither difference was statistically significant.

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

Finally, the ANOVA on the DEADLY ratings showed a significant effect of format, $F(2, 314) = 27.16$, $MSe = .18$, $p < .00001$. The small caps version of DEADLY with the black header–red message background received significantly higher ratings than the mixed case version of Deadly, which in turn was rated significantly higher than the red header–white message version. The ANOVA also showed a main effect of participant group, $F(1, 157) = 11.69$, $MSe = .86$, $p < .001$, and an interaction, $F(2, 314) = 8.04$, $MSe = .18$, $p < .001$. The community volunteers gave higher overall ratings than did the undergraduates. Comparisons among the means showed that the community volunteers rated all three variants highly, whereas the undergraduates gave significantly higher ratings to DEADLY with black header–red message than to the other two variants.

3.3. Discussion

The two population groups produced remarkably similar results given the number of stimuli evaluated. Moreover, the results were reasonably consistent irrespective of the use of hazard ratings or rankings (or noticeability ratings).

Direct comparison between the different formats suggests that the warning sign Z535.2 configurations are perceived to convey greater hazard levels than either the warning label standard Z535.4 or the alternative 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 for the rather small difference noted between WARNING and CAUTION.

The alternate format signs with the highest hazard ratings had the word DEADLY with a skull icon. The consistently high ratings for this configuration, plus the findings for these two components in Experiment 1, suggest that it could be useful in warnings in which the intent is to convey extreme hazard.

4. EXPERIMENT 3

Several aspects of current standards and guidelines were confirmed in the two previous experiments, but others were not. It is noteworthy that participants (undergraduates and community volunteers) in both Experiments 1 and 2 did not readily differentiate between

WARNING and CAUTION. Although this finding confirms most research on connoted hazard of signal words, it is inconsistent with these terms' definitions in current ANSI standards, which denotes WARNING as having a higher level hazard than CAUTION. One possible reason for this result is that most research on signal words uses lay persons, and it may be that lay persons do not differentiate between the terms because of lack of substantial exposure to properly labeled products and sign warnings bearing these terms. If this is true, then one might expect an industrial worker population—a population that is more likely to come into contact with signage bearing these terms in the course of their work—to understand the meanings of the signs and labels and their component parts. The industrial workers may be more likely to see the terms in the context of properly assigned low (CAUTION) versus medium (WARNING) level hazards. In other words, because of exposure and possible training, industrial workers' hazard perceptions may be more consistent with ANSI recommendations. Consequently, Experiment 3 was conducted to address this question directly and to attempt to replicate some of the other findings. To do so, this study re-examined some of the warning features evaluated in Experiments 1 and 2 using industrial workers.

4.1. Method

4.1.1. Participants. Seventy-five individuals, employees at three large industrial manufacturing and assembly plants in Janesville, Wisconsin, participated. Although each of the plants manufactured different goods, they had a common employer. All had employee safety programs, including a formalized lock-out/tag-out program, posted workplace safety signs, and chemical labeling. Eighty percent (80%) were female; mean age = 40, $SD = 9.0$.

4.1.2. Materials. Materials were assembled in three-ring binders. Each page contained different sets of items, in one of two random orders. The pages were full color photocopies ordered so that participants viewed the more elemental features (e.g., colors, shapes, and signal words) first, followed by multifeature configurations (e.g., entire headers, complete signs) to avoid any bias from their seeing assembled configurations before the simpler features. The specific stimuli in the survey are described in Table 5.

4.1.3. Procedure. Permission was requested from and granted by management of the industrial plants to test workers during break and lunch times. Participants rated the stimuli on a 5-point scale of perceived hazard with the following anchors: 0 = no hazard, 1 = *low hazard,* 2 = *moderate hazard,* 3 = *high hazard,* and 4 = *extreme hazard.* Participants were allowed as much time as they wanted to complete the ratings. Afterwards they were offered a small gift, for example, a coffee mug or calendar, in exchange for their assistance.

4.2. Results

Table 5 shows the means and standard deviations for the stimuli within each set. ANOVAs were conducted on each set of stimuli. All of them were significant ($ps < .05$). These analyses were followed by paired comparisons using Tukey's HSD Test. The following description summarizes some of the most important comparisons among the stimuli in each set.

Table 5 shows the mean hazard ratings and standard deviations of warning sign and label elements, including solid colors, multicolors, shapes, and signal words. The color red was significantly higher than all other colors, followed by yellow, black, and orange. The multicolor combinations only differed in extreme cases, for example, red–white versus black–white. The skull icon had the highest connoted hazard, followed by the diagonal stripes, and then the two exclamation point configurations. The other configurations and shapes had little or no connoted

TABLE 5
Mean Hazard Ratings of Signal Elements and Configurations (Experiment 3)

Signal Elements and Configurations	*M*	*SD*
Solid and multicolor bars		
Red	3.5	1.0
Yellow	2.1	1.1
Black	1.8	1.6
Orange	1.5	1.3
Blue	0.9	1.0
Purple	0.8	1.0
Green	0.7	0.9
White	0.3	0.7
Multi-color bars		
Red/white	2.4	1.2
Black/yellow	2.3	1.1
Black/red/white	2.3	1.1
Black/white/red	2.1	1.2
Black/orange	1.5	1.2
Black/white	1.2	1.1
Shapes		
Skull	3.9	0.7
Diagonal stripes	2.3	1.2
Triangle/exclamation	1.6	1.0
Triangle	1.4	1.0
Oval in rectangle	0.9	0.9
Square	0.9	1.1
Capsule (lozenge shape) in rectangle	0.9	1.0
Circle	0.7	1.1
Signal words		
DEADLY	3.9	0.6
DANGER	3.1	1.0
WARNING	2.4	1.2
CAUTION	2.0	0.9
NOTICE	0.9	1.0
SAFETY FIRST	1.2	1.2
Header configurations with nonsense word RESVRE		
White on red with skull	3.9	0.4
White on black with skull	3.6	0.9
White on red with triangle/exclamation	2.7	1.4
Black on red elongated hexagon in black	2.4	1.1
White on red oval in black	2.2	1.1
White on black oval in red	1.7	1.1
Black on yellow with triangle / !	1.6	1.0
Yellow on black	1.6	0.8
Black on orange elongated hexagon in black	1.5	1.1
Header configurations with word DANGER		
White on black with skull	3.8	0.5
White on red with skull	3.8	0.6
White on red with triangle / !	3.1	1.0
Black on yellow with triangle / !	2.8	1.0
Black on red elongated hexagon in black	2.8	1.0
White on red oval in black	2.8	0.8
White on black oval in red	2.6	0.9
Yellow on black	2.6	1.0
Black on orange elongated hexagon in black	2.6	1.3

continued

TABLE 5
Mean Hazard Ratings of Signal Elements and Configurations (Experiment 3; Continued)

Signal Elements and Configurations	*M*	*SD*
Entire-sign configurations from Z535.2 Z535.4 and alternative sets[a]		
ANSI Z535.2 environment signs		
DANGER white on red oval in black	3.5	0.7
WARNING black on orange hexagon in black	2.0	1.3
CAUTION yellow on black	2.0	1.0
NOTICE white on blue	0.9	1.0
SAFETY FIRST white on green	1.0	1.1
ANSI Z535.4 consumer product labels		
DANGER white on red with triangle / !	3.4	0.9
WARNING black on orange with triangle / !	1.8	1.1
CAUTION black on yellow with triangle / !	1.6	0.9
NOTICE white on blue with triangle / !	0.7	1.0
Proposed alternative system		
DEADLY white on black with skull (red message panel with white print)	3.8	0.7
DEADLY white on black with skull	3.8	0.6
DANGER white on red	3.1	0.8
CAUTION yellow on black	1.8	0.9
NOTICE white on blue	0.9	1.1

[a]All text message panels of the entire sign configurations (the X's) were printed in black on a white background, except (a) the Z535.2 WARNING and CAUTION which had black print on orange and yellow backgrounds, respectively, and (b) one of the proposed DEADLY configurations, which had white print on a red background.

hazard, and did not differ significantly. The term DANGER was rated significantly higher than WARNING and CAUTION, but there was no difference between the latter two terms. DEADLY was perceived to have the highest connoted hazard. Table 5 also shows the ratings of various header configurations, one set with the nonsense signal word RESVRE and the other set with the signal word DANGER. The nonsense and actual signal word headers showed a similar pattern except that ones with RESVRE tended to be rated lower than those with DANGER. Panels containing the skull icon were rated significantly higher than the other configurations.

Lastly, Table 5 shows the mean hazard ratings for entire-sign configurations from the Z535.2, Z535.4 and alternative sets in which the omitted message is replaced by *X*s. Of note is that for both sets of ANSI stimuli, DANGER in a red panel was rated highest. WARNING and CAUTION, with their respective orange and yellow panels, did not differ. In the alternative set, the two configurations using DEADLY with the skull were both rated higher than DANGER.

4.3. Discussion

The first two experiments obtained ratings from lay persons (university undergraduates and community volunteers recruited at a flea market and a shopping mall). This experiment (Experiment 3) surveyed industrial workers, a group of people who are more likely to have been exposed to both kinds of ANSI-type signs and labels in the course of their work and in safety training sessions, to determine whether their hazard perceptions for the stimuli are more in accord with the ANSI specifications than the lay participants in the other two experiments. The results showed that the ratings by the industrial workers were generally consistent with the perceptions of the undergraduates and community volunteers in the first two experiments in that similar parts of the ANSI standards were confirmed and other parts were not across all three experiments. For example, consistent with the first two experiments and in accord with the ANSI standards, red attained the highest hazard ratings and DANGER was rated higher than WARNING or CAUTION. Also, the industrial workers produced hazard ratings that were consistent with the other participant groups in the first two experiments but whose patterns are inconsistent with the current ANSI standards. For example, WARNING did not have a higher hazard connotation than CAUTION; orange was not rated higher than yellow; and certain shapes had no recognizable hazard association.

In addition, this experiment also confirms some of the results found in the two earlier experiments regarding the potential utility of alternative warning components. Certain alternative features appear to convey extreme levels of hazard better than those currently assigned to serve this function (e.g, DANGER). For example, DEADLY and the use of a skull icon appear useful in this regard. In sum, the pattern of results found in this experiment essentially mirrors the findings of Experiments 1 and 2, indicating that the evaluations of these stimuli are replicable and generalizable.

5. GENERAL DISCUSSION

This research examined the perceived hazard of various elements and configurations that comprise warnings signs and labels specified in current standards and guidelines. Previous research has mainly addressed parts of the standards and guidelines (e.g., signal words and colors) but has not addressed a broad range of elements and combinations of elements that form entire signs and labels. Some of the findings in this set of experiments confirm the specifications of the existing ANSI (1991) Z535 standards and the Westinghouse (1981) and FMC Corp. (1985) guidelines (e.g., DANGER received higher hazard ratings than WARNING or CAUTION), whereas others do not. As indicated previously, the standards specify that WARNING be used for greater hazard levels than CAUTION; however, the results do not support this distinction. Indeed, only the community volunteers in Experiment 1 appeared to differentiate between WARNING and CAUTION, whereas the other participant groups did not. Most research to date suggests little or no differentiation between these two terms or their associated colors (Chapanis, 1994; Wogalter & Silver, 1990, 1995). In short, the practical difference between these terms appears to be negligible. This conclusion is bolstered by the fact that similar results were obtained across a diverse sample of participants, including undergraduates, community volunteers, and industrial workers. Thus, rather than having three distinct levels of hazard conveyed by the three injury-related signal words, as was intended, there really are only two levels that people actually distinguish—DANGER versus WARNING and CAUTION.

The results show that there are alternative elements and configurations that produce higher hazard perception ratings than those currently recommended in warning design standards. In particular, the word DEADLY and the skull icon were two elements that produced higher ratings than any of their counterparts. These elements might be considered for warnings intended to signal extreme or life threatening danger.

It is worth noting that these findings were submitted to the ANSI Z535 revision committee to help improve the current standard system. A forthcoming revision of the standards, however, does not include the recommendations provided by these results, nor other research showing discrepancies in people's understanding of signal words (WARNING and CAUTION) and colors (orange and yellow). A common argument put forth by advocates of the current signal word hierarchy is that industrial workers are familiar with these terms. Clearly, it is beyond the scope of this research to address this possibility directly. However, it is worth noting that the data collected from industrial workers mirrored the data found with the general, untrained population. Although people might not readily differentiate between WARNING and CAUTION, we do not believe that one of these two terms or the colors orange and yellow should be dropped from use. Rather it should be acknowledged that they are synonymous and that they can be used interchangeably. This, of course, means that the number of levels of hazard is actually two rather than three. If three distinct level of hazard are desired, the term DEADLY, and possibly others, should be considered. Other research has scaled additional terms that also could be considered (e.g., Wogalter & Silver, 1990, 1995). It is hoped that the development of future standards will consider these data and other studies of this kind. Clearly, the design of these materials, which have an important role in preventing accidents and injuries, should be based on empirical data, not armchair thinking or tradition, as has been the case heretofore.

Overall, these results suggest the need for systematic testing of warning configurations to determine people's impressions of them, whether they understand the meanings intended, and whether they are effective in eliciting appropriate compliance behavior. Additional work is needed to determine the kinds of stimulus configuration that can cover a range of hazard levels, not just part of the hazard dimension. These results provide empirical evidence that current standards could be improved. For example, it is clear that people do not consistently differentiate between the signal words WARNING and CAUTION or between the colors orange and yellow. Therefore, it does not make sense to use these elements to indicate different hazard levels unless some type of effective training is given. Unfortunately, providing training and education that would reach all members of a target population is probably not feasible. Moreover, the costs of doing so would no doubt be prohibitive. It makes sense, therefore, to select the best terms, ones that already have existing, distinct, and understandable levels of connoted hazard, so that specific training is unnecessary or minimized.

ACKNOWLEDGMENTS

We would like to thank Beth Miller of Lab Safety Supply (Janesville, Wisconsin) for financial assistance and help in arranging data collection at nearby industrial manufacturing plants. We would also like to thank Ann B. Carter, Julie A. Swindell, Jason G. Daurity, William J. Vigilante, Jr., and Marilyn E. Spunar for their assistance in data collection. Portions of this article were presented at the 39th Annual Meeting of the Human Factors and Ergonomics Society in Nashville, Tennessee (Kalsher, Wogalter, Brewster, & Spunar, 1995; Wogalter et al., 1995) and at the American Psychological Association meeting in Toronto, Canada (Wogalter & Brewster, 1996).

REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888–918.

American National Standards Institute. (1991). *Warning colors, signs, symbols, labels, and tag standards* (Z535.1–5). Washington, DC: National Electrical Manufacturers Association.

Brewster, B. M. (1995). *White paper on safety sign components.* Wolcott, NY: Electromark Company.

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

DeJoy, D. M. (1989). Consumer product warnings: Review and analysis of effectiveness research. *Proceedings of the Human Factors Society, 33,* 936–940.

FMC Corporation. (1985). *Product safety sign and label system.* Santa Clara, CA: Author.

Kalsher, M. J., Wogalter, M. S., Brewster, B. M., & Spunar, M. E. (1995). Hazard level perceptions of current and proposed warning sign and label panels. *Proceedings of the Human Factors and Ergonomics Society, 39,* 351–355.

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

Westinghouse, Inc. (1981). *Product safety label handbook.* Trafford, PA: Westinghouse Printing Division.

Wogalter, M. S., Brelsford, J. W., Desaulniers, D. R., & Laughery, K. R. (1991). Consumer product warnings: The role of hazard perception. *Journal of Safety Research, 22,* 71–82.

Wogalter, M. S., & Brewster, B. M. (August, 1996). *Connoted hazard of warning components and configurations by industrial workers.* Paper presented at the American Psychological Association meeting, Toronto, Canada.

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

Wogalter, M. S., & Laughery, K. R. (1996). WARNING! Sign and label effectiveness. *Current Directions in Psychological Science, 5*(2), 33–36.

Wogalter, M. S., Magurno, A. B., Carter, A. W., Swindell, J. A., Vigilante, W. J., & Daurity, J. G. (1995). Hazard association values of warning sign header components. *Proceedings of the Human Factors and Ergonomics Society, 39,* 979–983.

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

Wogalter, M. S., & Silver, N. C. (1995). Warning signal words: Connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics, 38,* 2188–2206.