PERCEIVED EFFECTIVENESS OF ENVIRONMENTAL WARNINGS

Michael S. Wogalter, David R. Desaulniers, Sandra S. Godfrey

Department of Psychology, Rice University, Houston, TX 77001

ABSTRACT

This study examined perceived effectiveness of warning signs for various hazard situations. Four-statement signs contained a signal word, a hazard statement, a consequence statement, and an instruction statement. Four additional three-statement signs, each with a different statement systematically removed from the four-statement sign, were used, for a total of 5 signs for each hazard situation. The results of Experiments 1 and 2 indicated that removing content statements reduced perceived effectiveness. Hazard and instruction statements were the most important statements, showing the greatest decrease in effectiveness when deleted. Signs for the most hazardous situations were perceived as the most effective warnings. Experiment 3 examined redundancy of statements in a sign. The results suggested that the deletion of redundant statements, particularly signal words, had less influence on effectiveness. The hazard statement showed the lowest redundancy consistent with it producing the greatest effectiveness decrement when deleted.

INTRODUCTION

As our society becomes more technologically advanced we must deal with increasingly more complex systems. Used appropriately these systems are capable of dramatically improving the quality of our lives. However, misuse of these systems has the potential to produce bodily injury or death.

The first line of defense against injury from hazardous situations is to "design out" the hazard and the second is to provide the necessary safety barriers. However, technology and/or costs may preclude designing out hazards, and adequate barriers are not always possible. Furthermore, proliferation in the variety and complexity of new technological systems decreases the likelihood that users will recognize the potential hazards. Even if users recognize that a hazard exists, they may misjudge the probablity of sustaining injury (Slovic, Fischhoff, and Lichtenstein, 1979) or the severity of the consequences. Hence the user must be directly informed of the exact nature and magnitude of the hazard, its potential consequences, and the appropriate and prohibited behavior in the hazard area. Warnings, although considered a last line of defense in the effort to protect the user from potential hazards, have become increasingly important. In recent years, personal injury and product liability litigation has presented issues relevant to the Human Factors professional. One issue of increasing concern has been the effectiveness of warnings.

It is important to note regarding effectiveness is that "the ultimate criterion is whether the warning has actually modified human behavior" (Peters, 1984). In essence behavioral effectiveness asks, has there been a reduction of undesirable and unsafe acts that would otherwise have occurred without the warning? A serious obstacle to conducting behavioral investigations is the absence of an appropriate methodology. One difficulty is the construction of a realistic and believable scenario. Obtaining behavioral measures through observation is also difficult due to the relatively low frequency of critical incidents. Controlled experimentation on effectiveness generates the problem of maintaining ecological validity without presenting the subject with a truly hazardous situation. Also there are problems of ethics in conducting studies concerning hazardous situations. Thus, although the ultimate criterion in measuring warning effectiveness is appropriate safety behavior, such research is very difficult to do in many situations.

In recent years, there has been increased use of rating measures to assess warnings. Subjective suitability should be considered in the choice of warnings (Smith & Weir, 1978). Rating studies have opened up the evaulation of warnings at costs considerably less than the construction of behavioral effectiveness studies. For example, Collins, Lerner, and Pierman (1982) examined understandability and preference of safety symbols for various referent situations. Understandability was assessed with respect to correct identification of meaning and confusions. Preference was assessed for potential symbols of referent hazards. This report contains a variety of useable pictorial warnings. Of particular interest in the present context is the finding that preferences generally corresponded to understandability.

The present experiments involved the ratings of effectiveness of verbal environmental warning signs. This contrasts with recent research on warnings that examine pictorial/symbolic modes of communicating warnings. Potentionally, these symbols can provide safety messages to foreign speakers, illiterates, and children. However, symbols may not adequately indicate the details of a safety message; for example, a symbol indicating a respiratory protection device might also need to indicate specific kind to be used. Word signs may thus be a necessary part of complex warning messages. In the present research, no abstract pictorial messages were used. We assumed adequate comprehension since the verbal messages were written in their simplest terms. Beliefs about sign effectiveness, rather than understandablility, were of prime interest.

The present research has been directed toward the identification of the features of the warning message which influence the perception of warning effectiveness. There are numerous publications which provide guidelines for the development of warning signs. Two of these are the *Product Safety Sign and Label System* (FMC, 1980) and the *Westinghouse Product Safety Label Handbook* (Westinghouse, 1981). There are published guidelines for the development of psychologically effective warnings (e.g. Cunitz, 1981; Peters, 1984). Although these publications differ in a number of details, they agree for the most part on the fundamental elements necessary for an effective warning. The most frequently noted characteristics are that warnings should attract attention, provide information about the level of the hazard, provide a direct statement of the hazard, motivate behavior by stating the consequences, tell people how to avoid being hurt, and provide this information in a clear and concise fashion.

These basic guidelines served as a basis for developing warnings in this research effort. Our warnings concerned environmental hazards as opposed to product warnings. Each warning consisted of four statements: a signal word appropriate to the referent hazard, a statement of the hazard, a statement of the consequences, and an instruction ast to appropriate behavior within the hazard area. The following signs are examples.

Signal word:	DANGER	WARNING
Hazard:	HIGH VOLTAGE WIRES	WATER CONTAMINATED
Consequence:	CAN KILL	ILLNESS MAY RESULT
Instruction:	STAY AWAY	DO NOT DRINK

Three experiments involved the manipulation of amount and type of information available in the warning sign. The complete four-statement signs were systematically compared to signs with some of the content removed. This was accomplished by 4 three-statement variations of each sign. Each of the 4 was the original sign without one of the component statements.

As the focus concerns those aspects of warnings that influence perception of effectiveness, this research does not involve direct behavioral measures of warning effectiveness. In Experiments 1 and 2, subjects rated a series of systematically manipulated warnings on effectiveness. Specifically, they were asked to imagine warnings placed in appropriate environments and to give a ratings indicating whether people would obey the signs. Experiment 3 examined redundancy of statements in warnings.

EXPERIMENT 1

Method

Materials and Design. Seventeen warning signs depicting various hazard situations were used. There were five variations of each sign. One variation was a complete sign consisting of the following four statements: a signal word, a hazard statement, a consequence statement, and an instruction statement.. The other four variations each had one statement missing. The four "incomplete signs" were constructed by systematically removing each statement contained in the four-statement signs.

Subjects rated the warnings an eight-point Likert-type scale indicating that a rating of zero should mean that the warning would have no effect on people seeing the sign and a rating of seven should indicate that the presence of the warning would ensure that most people would obey. This warning effectiveness rating was the dependent measure. Each subject rated all 85 signs. Warnings were made using the Apple Macintosh 24-pt. Monoco bold font, and they were presented subjects on an overhead projector. The content of the sign (five levels: one four-statement and four three-statement variations) was one independent variable; the hazard situation addressed by the seventeen different signs was the second independent variable.

Subjects and Procedure. One hundred-seven University of Houston undergraduate students participated for extra credit in psychology courses. Subjects were run in groups in four sessions. A different random order of signs was presented in each session. Stimuli were presented at the rate of one every 10 seconds. Subjects were specifically told not to make their ratings according to the level of hazard involved. Rather, it was emphasized that they should make their ratings on the basis of sign effectiveness given the signs were placed in appropriate locations. Prior to the experimental trials, subjects examined five sample signs.

Results

The mean effectiveness scores reported in Table 1 indicate that removal of any of the statements from the signs reliably reduces perceived effectiveness. Removal of either the hazard statement or the instruction statement results in the greatest drop in effectiveness. This result is not surprising given the fact that these two statements provide specific information about the hazard and how to avoid it. A somewhat smaller drop in perceived effectiveness occurs with the removal of the consequence statement or the signal word.

TABLE 1
Perceived Effectiveness as a Function of Warning Signs and their Content (Exp. 1).

	CONTENT			
ALL	MINUS SIGNAL	MINUS HAZARD	MINUS CONSEQ	MINUS INSTRU
5.04	4.77	4.47	4.72	4.50

A two-way repeated measures analysis of variance shows the effect of the content variable (removal of statements) to be significant, $F(4,424) = 20.03$, $p < .001$. This effect is also significant using individual signs as the random variable (collapsing across subjects), $F(4,64) =$

11.83, $p < .001$. Comparisons of the three-statement content means against the four-statement mean (collapsing across subjects and signs) shows that removal of any of the statements leads to a significant decrease in perceived effectiveness (all p's $< .01$). There is a significant main effect of the different hazard situations (signs), $F(16,1696)=82.15$, $p < .001$. This is not surprising given the number of referent hazard situations used; further discussion of this effect is deferred until Experiment 2.

Although the means for the four-statement signs are, in general, rated higher than all of the three-statement versions, the significant sign by content interaction, $F(64, 6784)= 5.93$, $p <.001$, indicats that the content effect may not hold in all cases. For a few signs, deleting a statement results in an increased rating of perceived effectiveness. With the large number of cell means involved in this experiment, correction for experiment- wise error rate is essential. Comparisons adjusting for alpha using Bonferoni's multiple-comparison test (minimum signficant difference = .52) shows that in no instance are any of the three-statement signs perceived significantly more effective than their corresponding four-statement signs. Of the 68 three-statement signs, 23 were rated significantly lower in effectiveness than the respective four-statement versions (9 lacking the instruction statement, 8 lacking the hazard statement, 5 lacking the consequence statement, and 1 lacking the signal word). There are 45 three-statement signs that did not differ reliably from the four-statement version.

EXPERIMENT 2

This experiment was a replication of Experiment 1 with a larger sample of stimulus signs. Different experimental instructions and rating scale were used to specify in operational terms a definition of warning effectiveness.

Method

Materials and Design. Subjects were told to rate "what percentage of people would be likely to obey the sign" rather than"how effective you think the sign would be." The ratings utilized an 11-point scale with the labels 0% to 100%. Below each point were the numbers 0 to 10 which corresponded directly with the percentages (e.g., 3 corresponded to 30%). Prior to this experiment, twelve "expert"judges (psychology faculty and graduate students) ranked 48 (17 from Experiment 1 and 31 "new") four-statement signs according to level of hazard. Twenty-five signs that had low variablility and were distributed across a wide range of hazards were used in this experiment. As in the previous experiment, there were five variations of each sign: one four-statement warning sign and four three-statement signs.

Subjects and Procedure. Eighty-one subjects from University of Houston and Rice University psychology courses were told to assume that each warning sign was placed in an appropriate location. They were told to make ratings based on the percentage of people, who after seeing the warning sign, would obey it. Subjects were told, for example, that if 100 people see a particular sign but only 20 obey it, then this sign should receive a rating of 2 for 20%. If on the other hand, they thought that 70 people out of 100 would obey a particular sign, then it should be given a rating of seven for 70%. Each subject rated 125 warnings.

Results

The results are shown in Table 2. In general, the removal of any statement from the signs reduces perceived effectiveness. These means show basically the same pattern as Experiment 1. Removal of either the hazard statement or the instruction statement leads to the greatest drop in effectiveness. A somewhat smaller drop occurs with the removal of the consequence statement or the signal word.

TABLE 2
Perceived Effectiveness as a Function of Warning Signs and their Content (Exp. 2 using an 11-point scale).

	CONTENT			
ALL	MINUS SIGNAL	MINUS HAZARD	MINUS CONSEQ	MINUS INSTRU
7.28	7.12	6.65	6.83	6.54

A two-way repeated measures ANOVA shows the effect of the content variable was significant, $F(4, 320)= 23.72$, $p < .001$. This effect is also significant using individual signs as the random variable (collapsing across subjects), $F(4,96)= 8.34$, $p < .001$. Comparisons of the three-statement means against the four-statement mean shows that the individual removal of any of the statements leads to a significant decrease in perceived effectiness (all p's$<.001$). A significant sign content interaction, $F(96, 7680)=9.95$, $p < .001$) indicates that the general content effect does not hold in all cases. Comparisons adjusting for alpha using Bonferoni's multiple-comparison test (minimum signficant difference = .79) showed that there was only one three-statement sign significantly more effective than its corresponding four-statement version. Of the 100 three-statement signs, 30 were rated significantly lower in effectiveness than the respective four-statement versions (11 lacking the instructions statement, 9 lacking the hazard statement, 8 lacking the consequence statement, and 2 lacking the signal word).

Various hazard situations show different levels of perceived effectiveness, $F(24, 1920)=52.31$, $p <.001$). This effect is difficult to interpret partly because it interacts with the content variable. However, examination of the four-statement signs revealed that those dealing with high-level hazards such as severe electric shock were rated as more effective than warnings of mild hazards such as wet slippery floor. In order to explore this relationship further, signs were rank ordered by "expert" judges as to the degree of hazard and these rankings were then paired with the mean perceived effectiveness ratings for the four-statement signs from both Experiments 1 and 2. Correlations between the two measures were significant,

r= .80, p < .001 and r = .62, p < .001 respectively. These results strongly indicate that the greater the hazard depicted by the sign, the greater the perceived effectiveness of the sign. Perhaps people look for warnings around greater hazards, and perhaps the greater the hazard level the more motivation to comply.

Those instances where three-statement signs were rated equal to (or in a very few cases, somewhat greater than) the corresponding four-statement signs appeared to be due to the removal of obvious, implied, and redundant information. In some cases, the four-statement sign sounds somewhat silly because the information is so obvious. For this reason, some of the four-statement signs might be perceived somewhat effectiveness.

EXPERIMENT 3

The question asked here is whether the change in perceived effectiveness reported in Experiments 1 & 2 can be attributed to redundancy of the "deleted" statement in comparing three-statement to four-statement signs. It is predicted that the more unique or nonredundant the information provided by a statement, the greater the effect of that statement's removal will be.

Method

Materials and Design. The stimuli for this experiment were taken from the signs used in Experiment 1 The earlier set of signs was used because this experiment was started prior to Experiment 2. Specifically, sixty-eight three-statement signs (all four versions of each sign) were paired with the statements that were deleted from the original four-statement versions. Three-statement signs were presented simultaneously with related "deleted" statements. Subjects rated the degree to which the information in the single statement was already included in the information given in the three-statement sign, providing a measure of the degree to which subjects perceived the "missing" statement to be redundant. The term "redundancy" was not used in the instructions.

An 8-point Likert-type scale (with endpoints of zero and seven) was used to rate each sign/statement. Below each point on the scale were anchors describing degrees of amount (Bass, Cascio, and O'Conner, 1974). Each point of the scale was numbered and defined so that at the extremes zero represented NONE and seven represented ALL.

Subjects and Procedure. Sixty-six undergraduates from introductory psychology classes at the University of Houston, participated for extra credit. Seven sessions were run with each session using a different sign order. Subjects were told that they would be viewing several variations of a warning sign and that they should read each sign and the accompanying statement carefully and not confuse similar variations of signs. They were instructed to make their ratings on the basis of their world knowledge as well as the information being presented to them at that moment. Stimuli were presented at rate of one every 10 seconds. Three practice trials were given to acquaint subjects with the stimuli.

Results

Mean ratings of redundancy for each elemental statement with respect to the remaining three-statement warnings are shown along the first row in Table 3. These scores were obtained by collapsing across hazard situations to obtain the mean redundancy for content. Reordering the redundancy means shows the following descending order: signal word, consequence statement, instruction statement, and hazard statement. The signal word is the most redundant statement (relative to the other three statements of a sign), and the hazard statement is the least redundant. A within-subjects analysis of variance showed a significant main effect of redundancy for content type, $F(3,195)= 94.91$, $p < .005$). Tukey's Honestly Significant Difference test showed mean redundancy for the hazard statement was significantly lower than all other statement types (HSD=.37).

TABLE 3
Perceived redundancy of statement (in bold print from Exp. 3) and effectiveness differences without statement (in reg. print from Exp.1) as a function of content.

	CONTENT			
	Signal	Hazard	Conseq	Instruc
Redundancy	**5.85**	**3.70**	**5.56**	**5.46**
Exp 1 differences	-0.26	-0.57	-0.33	-0.53

Using the perceived effectivenesss data from Experiment 1, a measure of the change in perceived effectiveness due to the deletion of a statement was obtained by subtracting effectiveness ratings of the four-statement signs from the effectiveness rating of each of its four related three-statement signs. These effectiveness decrement scores (collapsing across hazard situations) are shown in the bottom row of Table 3. The pattern of mean redundancy ratings is the same but in the opposite order of the perceived effectiveness decrements. The deletion of more redundant information is associated with smaller decrements in perceived effectiveness. In other words, the greater the relative amount of information in a statement, the greater the negative effect of deleting that statement. The mean ratings for the hazard statement shows the lowest redundancy consistant with it producing a large effectiveness decrement when deleted.

A correlation of the 68 redundancy ratings with the corresponding perceived effectiveness decrement scores (from Experiment 1) yields a small but significant relationship, r = .332, p <.005. This indicates that deleting more redundant statement produces smaller decrements in perceived effectiveness. Separate correlations were calculated for signs of each of the four content statements in order to examine relationships across types of content deletions. A significant correlation was found between signal word difference scores and the redundancy ratings of signal words, r =.520 , p <.05. This result is somewhat surprising given the special nature of signal words. Signal words were expected to receive consistently high redundancy ratings because their purpose is to attract

attention rather than to transmit specific information. Since signal words reflect the degree of hazard in warning situations, it was expected that restriction of range would reduce the likelihood of obtaining significant results. The correlation indicates that as redundancy of signal words increase, there is a smaller decrement in perceived effectiveness when the signal word is deleted. Correlations for the three other types of content were not significant.

DISCUSSION

The results of these experiments recommend that environmental warnings, in general, contain four types of content statements: signal word, hazard statement, consequence statement, and instruction statement. The hazard and instruction statements are the most important in environmental warnings; deletion of these statements leads to the greatest reduction of effectiveness. The consequence statement and signal word were judged somewhat less important, but only by degree; the removal of any one of the four contents produces a decrement in perceived effectiveness. Although the deletion of statements did not, in general, increase perceived effectiveness, there are many instances with no significant decrement for three-statement signs. Hence, one may be able to delete some information without a substantial loss of effectiveness; this could be useful when brevity is necessary. Examples include cases such as highway signs, where there is insufficient time to read a long message, or where there are constraints on available display space. The procedures used in the present research allowed plenty of time and close scrutiny of the warning. In the real world this frequently is not the case.

The results suggested that redundancy is related to effectiveness, but redundancy does not describe the whole story--it does not explain much effectiveness variance. A more specific examination of redundancy in future may yield positive factors related to the improvement of warnings under degraded conditions.

Ratings are only a first step in evaluating effectiveness of warning signs. Ultimately, if the conditions allow, the sign should be tested in real-world settings to obtain measures of behavioral compliance (e.g. studying accident rates). New methodologies need to be developed to test behavioral effectiveness. We should not, however, wait and do nothing to improve warning signs. Warnings should be tested by the methodologies available. Armchair philosophizing and theorizing, even by "experts," may be inappropriate despite the economy of designing warnings in the office (Miller, 1978). We need to go beyond the general impressions and demand controlled testing of warning elements and formats.

References

Bass, B. M., Cascio, W. F., & O'Conner, E. J. (1974). Magnitude estimations of expressions of frequency & amount. *Journal of Applied Psychology, 59*, 313-320.

Collins, B. L., Lerner, N. D., & Pierman, B. C. (1982). Symbols for industrial safety. Technical Report NBSIR 82-2485.

Cunitz, R. J. (1981). Psychologically effective warnings. *Hazard Prevention, 17*, 5-7.

FMC (1980)., *Product Safety Signs and Label System*, FMC: Santa Clara, California, 3rd Ed.

Laner, S. & Sell, R. G. (1960). An experiment on the effect of specially designed safety posters. *Occupational Psychology, 34*, 153-169.

Miller, J. A. (1978). *Labeling-The state of the art.* Cambridge, Massachusetts: Marketing Science Institute.

Peters, G. A. (1984). A challenge to the safety profession. *Professional Safety, 29*, 46-50.

Slovic, P., Fishoff, B. & Lichtenstein, S. (1979). Rating the risks, *Environment, 21*, 14-39.

Smith, G. and Weir, R. (1978). Laboratory visibility studies of directional symbols used for traffic control signal. *Ergonomics, 21*, 247-252.

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