HUMAN FACTORS, 1987, 29(5), 599–612

Effectiveness of Warnings

MICHAEL S. WOGALTER, SANDRA S. GODFREY, GAIL A. FONTENELLE, DAVID R. DESAULNIERS, PAMELA R. ROTHSTEIN, *and* KENNETH R. LAUGHERY,[1]
Psychology Department, Rice University, Houston, Texas

The purpose of the present work was to identify some of the factors that influence effectiveness of warnings. Two laboratory experiments designed to examine behavioral effectiveness indicated that a warning placed before procedural instructions is more likely to lead to compliance than a warning that follows instructions. Two rating experiments indicated that for greatest perceived effectiveness, environmental warnings generally require a signal word plus hazard, consequence, and instruction statements. A third rating experiment suggested that informative, nonredundant statements add to a warning's perceived effectiveness. Several field demonstration studies showed that cost of compliance and salience influence behavioral effectiveness. Implications and applications to warning design are discussed.

INTRODUCTION

Do warnings influence the behavior of people to whom they are directed? Few empirical studies have demonstrated circumstances in which warnings change behavior. Indeed McCarthy, Finnegan, Krumm-Scott, and McCarthy (1984) concluded from a review of the literature that warnings are not effective. But lack of empirical evidence does not warrant this general conclusion. Instead, a more useful approach is to identify factors that influence the effectiveness of warnings.

Several design criteria for warnings have been proposed (e.g., Cunitz, 1981; FMC, 1980; Peters, 1984; Westinghouse, 1981). The first group addresses the content of the message:

(1) *Signal word.* Warnings should have signal words appropriate to the level of hazard (e.g., "DANGER," "WARNING," "CAUTION").
(2) *Hazard statement.* Warnings should tell what the dangers are.
(3) *Consequences.* Warnings should emphasize the results of failure to heed in order to motivate people to comply.
(4) *Instructions.* Warnings should tell people what they should or should not do to avoid danger (i.e., the do's and don'ts).

The second group lists characteristics important for conveying the message:

(5) *Attention-getting.* Warnings should stand out from a noisy background (i.e., they should be conspicuous). Warnings should be present when and where they are needed.
(6) *Comprehensible.* Warnings should be understandable to the population that is exposed to the hazards.
(7) *Concise.* Warnings should be short and to the point.
(8) *Durable.* Warnings should be resistant to wear, abuse, and damaging environmental conditions.

The ultimate criterion of warning effectiveness is, of course, whether the warning actually modifies human behavior (Peters, 1984). In other words, is there a reduction of unde-

[1] Requests for reprints should be sent to Kenneth R. Laughery, Psychology Department, Rice University, Houston, TX 77251.

sirable and unsafe acts that would otherwise occur without the warning?

A study by Laner and Sell (1960) illustrates that behavior can be affected by warnings on safety posters in the workplace. The posters instructed workers to place chain slings onto a crane hook as a safety precaution when the slings were not in use. The desired behavior increased during the interval that the signs were posted. The increase in safe behavior was greatest in those low-ceilinged shops where the unsafe practice constituted the greatest hazard. This result indicates that warnings are more effective if they are directly relevant to the situation.

Conducting research such as Laner and Sell's (1960) is difficult for several reasons. Observing behavior that is the direct result of warnings is labor-intensive because critical events are generally infrequent and sporadic. Extraneous variables must be controlled in order to draw inferences about causal relationships. Laboratory studies permit greater control, but they may lack face validity. Also, laboratory findings cannot always be generalized to real-world settings. Creating hazardous situations that are ethically acceptable and at the same time believable is challenging. These difficulties illustrate why research in this area needs to be attacked on various methodological fronts.

The present study uses several methodologies, starting with a set of highly controlled laboratory experiments and ending with a set of real-world demonstrations. Laboratory Experiments 1 and 2 examine the effects of signal word presentation and the location of a warning in a set of instructions. Rating Experiments 1 and 2 examine the relative importance of the parts that make up a warning to its perceived effectiveness. Rating Experiment 3 explores the effect of redundancy on perceived effectiveness. Finally, several field demonstrations are presented illustrating that saliency and cost of compliance influence warning effectiveness.

LABORATORY EXPERIMENT 1

A paradigm was developed to examine the effects of signal word and location of the warning on behavior. Subjects were asked to perform a basic chemistry demonstration. The low-level hazards associated with the task were realistic and believable, thereby providing face validity for the warnings.

Method

Subjects and design. Fifty-one Rice University undergraduates participated in the study for course credit. The study investigated effects of the location of the warning statement (before and after instructions) and two different signal word presentations ("WARNING" and "Note"). These two variables were manipulated factorially in a between-subjects design. An additional condition that was identical to the others but lacked the warning statements served as a control.

Materials. Actual chemistry laboratory equipment, such as an analog scale, flasks, beakers, and cylinders, was used to ensure credibility. Paper towels, plastic gloves, and molded paper masks were also provided. The instructions called for handling and mixing several different substances and referred to them by a number or letter label attached to the containers. The actual substances used were water, bleached white flour, corn oil, table sugar, and yellow corn flour. Green food coloring was added to the water and red coloring added to the sugar in order to disguise them. These substances were suitable because they were not actually hazardous and because they have a varied consistency and coloring similar to actual chemicals. A Mettler analytical scale was used by the experimenters to measure the pre- and posttask weights of the substances in the containers.

Procedure. Individual subjects were shown the equipment and materials, given the instructions, and told to begin. The instructions directed them to measure and mix certain substances together in a specified order. The warning message was on the first or second page, depending on whether it was at the beginning or at the end of the instructions. The format for the condition in which "WARNING" was the signal was as follows:

WARNING

(1) Skin contact may result in discoloration or irritation.
(2) Inaccurate measurement or improper mixing order may result in (a) an unusable product, (b) a foul-smelling gas or (c) a noxious gas.
Avoid skin contact with all substances.
Perform accurate measurements.
Mix substances in proper order.
Wear rubber gloves and mask.

The warning format for the note condition was the same except that the signal word was presented on the first line, the first letter of the word was capitalized, and it was separated from the first statement by a colon. The first line of the warning appeared as follows:

Note: (1) Skin contact may result in discoloration or irritation.

Before and after each session the weights of the containers were measured using the Mettler balance. Subtracting the post- from the pretask weights provided an accuracy measure for the subjects' performance on the task. While subjects performed the task, the experimenter recorded elapsed times for several events, including time to put on the mask, time to put on the gloves, time to pick up the first laboratory object, and time to complete the task. The task took an average of ten minutes to complete. After completing the task subjects were debriefed and questioned concerning their hypotheses and beliefs about the purpose of the experiment.

Results

The accuracy and time measurements were not influenced by warning location or type of signal word (F's < 1.0). However, the employment of protective equipment (use of both mask and gloves) was clearly influenced by the location and presence of the warning. Table 1 shows the proportion of subjects who complied with the warning by using protective equipment as a function of conditions. When the warning was present, a higher proportion of subjects used the mask and gloves than when there was no warning. Also, when the warning was located before the instructions, the proportion of subjects who used the mask and gloves was higher than when the warning came after the instructions. The differences among the conditions were significant, $F(4,45) = 4.45$, $p < 0.01$. Planned comparisons showed a significant effect of location (beginning versus end), $t(45) = 2.13$, $p < 0.05$, as well as of presence (warning versus no warning), $t(45) = 3.49$, $p < 0.001$. The type of signal word did not affect compliance, $t < 1.0$.

LABORATORY EXPERIMENT 2

Observation of the subjects in those conditions in which the warning message was at the end of the instructions suggested that if they turned the page before starting the task, they tended to comply with it; if they did not turn the page, they did not comply. Of 18

TABLE 1

Proportion Compliance to Warning as a Function of Format and Location (Laboratory Experiment 1)

Warning at Beginning	*Note at Beginning*	*Warning at End*	*Note at End*	*Control (no warning)*
0.90	0.70	0.50	0.50	0.10

subjects in the "warning at end" condition, 10 did not turn the page; of these 10 subjects, 9 did not put on the mask and gloves. Of the 8 subjects who did turn the page, 7 complied by using the protective equipment.

Subjects who did not turn the page may have been unaware that the instructions continued on a second page. Alternatively, subjects may have simply decided after reading the first part of the instructions not to read further but rather to begin carrying out the task immediately. A follow-up study was conducted in order to replicate the findings of the first laboratory study and at the same time to increase the number of subjects who read the complete set of instructions.

Method

The method for the second study was identical to the first, with several exceptions. A new variable was introduced, which was the presence or absence of the following statement printed at the top of the first instruction sheet: "Please read through the entire set of instructions before beginning." Subjects were randomly assigned to one of three conditions and participated individually. The first group received the warning before the instructions (n = 15). The other two groups received the warning after the instructions. One of these two groups (n = 15) received the "read through" statement and the other group did not (n = 16). Since the accuracy measures were not affected by the manipulations in the first study and the signal word manipulation had no effect, they were both dropped from the second study.

Results

The "read through" instructions were successful in that they increased the proportion of subjects who turned the page from 0.50 to 0.80. As can be seen in Table 2, they also increased compliance (use of mask and gloves); however, the difference between the "warning at the end" and the "warning at the end with 'read through' instructions" was not statistically significant. Nor was the difference between the "read through" condition and the "warning at the beginning" condition statistically significant. Compliance in the "warning at the beginning" condition was higher than in the "warning at the end" condition without "read through" instructions, $t(29) = 2.70$, $p < 0.05$, thus replicating the finding in the first study. Most of the subjects who turned the page in the "warning at the end" condition put on the mask and gloves ($\phi = 0.82$, $n = 31$, $p < 0.001$). These results indicate that the failure to attend to the warning before beginning the task was responsible for the differences in compliance, and that placing the warning at the end of a two-page set of instructions lowers the chances that it will be read or heeded.

TABLE 2

Proportion Compliance to Warning as a Function of Location and Instructions (Laboratory Experiment 2)

Warning at Beginning	*Warning at End with "Read Through" Instructions*	*Warning at End*
0.867	0.667	0.4375

RATING EXPERIMENT 1

The laboratory paradigm used in Experiment 1 allowed the manipulation of several variables in a believable setting while controlling for other variables. Another approach is to study warnings using rating measures. The low frequency of critical incidents, as previously mentioned, suggests that rating measures are a practical alternative to

the more costly and time-consuming behavioral measures (Collins, Lerner, and Pierman, 1982; Smith and Weir, 1978). In Rating Experiment 1 subjects were asked to rate the effectiveness of a sample of warning signs. The purpose was to determine the extent of the influence of various components of the warning messages on perceived effectiveness. Each warning had four components: a signal word, a hazard statement, a consequence statement, and an instruction statement.

Method

Materials and design. Seventeen warning signs depicting various hazard situations were used. There were five versions of each sign. One version was the complete sign, consisting of the following four statements: a signal word, a hazard statement, a consequence statement, and an instruction statement. The other four versions were incomplete warnings; each had one statement missing. The four incomplete signs were constructed by systematically removing one of the statements contained in the four-statement signs. Thus there were 17 four-statement signs and their 68 three-statement versions, for a total of 85 stimuli. The relative order of warning statements remained constant for the five versions of each sign. Following are two examples of complete four-statement signs:

Signal Word:	DANGER	WARNING
Hazard:	HIGH VOLTAGE WIRES	UNDERGROUND GAS LINE
Consequence:	CAN KILL	EXPLOSION AND FIRE POSSIBLE
Instruction:	STAY AWAY	NO DIGGING

The warnings were printed in 24-point Monoco bold font using an Apple Macintosh computer and Imagewriter printer. They were converted into transparencies and presented to subjects on an overhead projector. Subjects rated the warnings on an eight-point Likert-type scale. A rating of zero indicated that the warning would have no effect on people seeing the sign, and a rating of seven indicated that the presence of the warning would ensure that all people would obey the message implied by the warning. The anchor labels and corresponding numbers were explained in oral instructions to subjects and were available on the rating response sheet. Each subject rated all 85 signs. One independent variable was content, which had five levels: one four-statement sign and four three-statement signs. The hazard situation was the second variable; 17 different hazards were addressed.

Subjects and procedure. One hundred seven University of Houston undergraduate students participated in the experiment and received extra credit in psychology courses. Subjects were run in large groups, and a different random order of signs was presented in each session. Stimuli were presented at the rate of one every ten seconds. Subjects were specifically told not to make their ratings according to the level of hazard involved; rather, they should assume that the signs were placed in appropriate locations and make their ratings on the basis of sign effectiveness. Prior to the experimental trials, subjects were shown five sample signs.

Results

The mean effectiveness scores reported in Table 3 indicate that removal of any of the statements from the signs reduced perceived effectiveness. The effect of the content variable (i.e., removal of statements) was significant, $F(4,424) = 20.03$, $p < 0.001$. This effect was also significant using individual signs as the random variable (collapsing across subjects), $F(4,64) = 11.83$, $p < 0.001$.

Comparisons of the three-statement means

TABLE 3

Mean Perceived Effectiveness Rating as a Function of Warning Signs and Their Content

	All Four Statements	*Minus Signal Word*	*Minus Hazard Statement*	*Minus Consequence Statement*	*Minus Instruction Statement*
Rating Experiment 1 (8-point scale)	5.04	4.77	4.47	4.72	4.50
Rating Experiment 2 (11-point scale)	7.28	7.12	6.65	6.83	6.54

against the four-statement mean (collapsing across subjects and signs) showed that removal of any of the statements led to a significant decrease in perceived effectiveness (all p's < 0.01). Removal of either the hazard statement or the instruction statement resulted in the greatest decrease in effectiveness. Since these two statements provide specific information about the hazard and how to avoid it, their greater importance to rated effectiveness seems reasonable.

The ANOVA showed a significant effect for the 17 different hazard situations (signs), $F(16,1696) = 82.15$, $p < 0.001$. There was also a significant hazard × content interaction, $F(64,6784) = 5.93$, $p < 0.001$. Comparisons—adjusting for alpha using Bonferoni's multiple comparison test (minimum significant difference = 0.52)—showed that none of the four three-statement signs was perceived as significantly more effective than its corresponding four-statement sign. The relationship between hazard and content is explored further in the second rating experiment.

RATING EXPERIMENT 2

Rating Experiment 2 served to replicate and extend the results of the first rating experiment.

Method

Materials and design. This study had a larger sample of stimulus signs, different experimental instructions, and a different rating scale for measuring warning effectiveness. An 11-point scale was used, with end points of zero and ten. Prior to this experiment, 12 judges (psychology department faculty and graduate students) ranked 48 four-statement signs (17 signs from Rating Experiment 1 and 31 "new" signs) according to level of hazard. Twenty-five signs were subsequently selected; the selection criteria were that they exhibited low variability in the judges' ratings and were distributed across a wide range of hazards. 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 Rice University and the University of Houston were told to assume that each warning sign was placed in an appropriate location and to make ratings based on the percentage of people who would obey the warning sign if they saw it. For example, if 100 people see a particular sign but only 20 would obey it, then the sign should receive a rating of two, signifying 20%. If, on the other hand, 70 people out of 100 would obey a particular sign, then it

should be given a rating of seven, for 70%. Each subject rated all 125 warning stimuli.

Results

As can be seen in Table 3, the effectiveness ratings for the content variable show the same pattern of means as the first rating study. Again, the different hazards (signs) showed different levels of perceived effectiveness, $F(24,1920) = 52.31$, $p < 0.001$. This effect is difficult to interpret partly because it interacted with the content variable, $F(96,7680) = 9.95$, $p < 0.001$.

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 a wet slippery floor. In order to explore this relationship the hazard rankings obtained from judges in the stimuli selection phase (discussed earlier) were paired with the mean perceived effectiveness ratings for the four-statement signs from both studies. The correlations between the hazardousness measure and the effectiveness ratings for the first and second studies were significant, $r = 0.80$, $p < 0.001$, and $r = 0.62$, $p < 0.001$, respectively. These results suggest that the greater the hazard, the greater the perceived effectiveness of the sign. Perhaps people are more motivated to comply with warnings that concern very serious hazards; precautionary behavior may be greater in hazardous situations and thus informative signs dealing with these situations are considered more effective than signs dealing with less serious hazards.

In both Rating Experiments 1 and 2, approximately two-thirds of the three-statement signs received ratings that did not differ significantly from those of their corresponding four-statement signs. The following is an example in which removal of the consequence statement resulted in a higher rating of effectiveness:

CAUTION
WET FLOOR
YOU COULD FALL
WATCH YOUR STEP

Such instances appeared to be due to the removal of obvious, implied, and redundant information.

RATING EXPERIMENT 3

Rating Experiments 1 and 2 raised the question of whether the difference between the perceived effectiveness of the three- and four-statement signs is related to the redundancy of the deleted statement. Rating Experiment 3 addresses the following question: Will the removal of unique, informative warning statements decrease the signs' perceived effectiveness?

Method

Materials and design. The three-statement warnings (signs) used in Rating Experiment 1 and their respective "missing statements" were presented simultaneously (i.e., side by side). The following is an example:

WARNING
WATER CONTAMINATED
ILLNESS MAY RESULT

DO NOT DRINK

Subjects rated the degree to which the information in the single statement was already included in the three-statement sign. This rating provided a measure of the degree to which subjects perceived the missing statement to be redundant.

An eight-point Likert-type scale (with end points of zero and seven) was provided for rating each sign and statement pair. Anchors that describe amounts of redundancy were below each point on the scale: the ordered verbal

labels were "none," "a limited amount of," "some," "fairly much," "a lot of," "an extreme amount of," "almost entirely," and "all" (Bass, Cascio, and O'Conner, 1974).

Subjects and procedure. Sixty-six undergraduates from introductory psychology classes at the University of Houston participated and received extra course credit. A different random order of signs was used for each of the four sessions. Subjects were told to read each sign and the accompanying statement carefully and to try not to confuse similar variations of signs. Subjects were instructed on the rating scale: they were told that a rating of zero means that none of the information in the missing statement is contained in the sign whereas seven means that all of the information contained in the missing statement is contained in the sign. The intermediate scale points were explained as indicating varying degrees of redundancy. Subjects were instructed to base their ratings on their knowledge of their environment as well as on the information being presented to them at that moment. Stimuli were presented at the rate of one every 10 s. Three practice trials were given to acquaint subjects with the stimuli.

Results

Table 4 shows the mean redundancy ratings for each type of missing warning statement (e.g., signal word, consequence, etc.), which were obtained by collapsing across the 17 hazard situations. The four types of content statements produced a significant effect on redundancy ratings, $F(3,195) = 94.91$, $p < 0.005$.

Signal words were rated the most redundant and hazard statements the least. Tukey's Honestly Significant Difference Test showed that the mean redundancy rating for the hazard statement was significantly lower than the other statements (HSD = 0.37), indicating that it is the most informative part of the warning.

Data from Rating Experiment 1 were used to obtain a measure of the change in perceived effectiveness due to the deletion of a statement. These scores—the difference between the effectiveness ratings of the four-statement signs and each of the three-statement versions—reflect the decrement in perceived effectiveness that occurs when statements are deleted. Table 4 shows that the pattern of decrement and redundancy means is in the opposite direction. Moreover, a small but significant relationship was found between the 68 redundancy ratings and their corresponding effectiveness decrements, $r = -0.33$, $p < 0.005$. Apparently deleting the more informative warning statements produces a corresponding decrease in the perceived effectiveness of a sign. Conversely, deleting redundant warning statements produces a small decrease in perceived effectiveness. According to the means in Table 4, the hazard statement provides more information than the other parts of the warning; removing the hazard statement produces the greatest loss of perceived effectiveness. For example, consider the hazard statement "HIGH VOLTAGE." This statement allows the reader to infer appropriate courses of action based on past experience. In contrast, the instruction statement "DO NOT TOUCH" gives the reader no information about the hazard or the appropriateness of alternative actions.

Ratings are only a first step in evaluating effectiveness of warning signs. Ultimately the sign should be tested in real-world settings to obtain measures of behavioral compliance.

FIELD DEMONSTRATION STUDIES

The purpose of the studies reported in this section is to demonstrate the effectiveness of

TABLE 4

Mean Perceived Redundancy (Rating Experiment 3) and Mean Decrement in Effectiveness (from Rating Experiment 1) Resulting from Removal of Statement

	Signal Word	*Consequence Statement*	*Instruction Statement*	*Hazard Statement*
Redundancy	5.85	5.56	5.46	3.70
Effectiveness decrement	0.26	0.33	0.53	0.57

warnings in the context of various real-world settings and to attempt to identify some of the attributes of warnings that influence their effectiveness.

Broken Copy Machine and Telephone

Method. The first demonstration was conducted in a room with two copy machines in the library of a large medical center. Baseline data collection sessions were conducted before and after the experimental sessions for both the copy machine and telephone studies. One copy machine was kept busy by the experimenters during the one-hour sessions. During the experimental session a 3 × 5 in (7.6 × 12.7 cm) warning sign was placed on the second machine with the following message:

CAUTION
MACHINE DOES NOT WORK
MAY CAUSE DELAY
USE ANOTHER MACHINE

The number of people who used the machine and the number of people who appeared intent on using it but did not (i.e., entered the room carrying written material, approached the machine on which the sign was posted, looked at the sign, and left) was recorded.

The second field demonstration was conducted in the lobby of a large medical office building where three pay telephones are located near a sitting area. The experimenters kept two of the telephones busy during the 45-minute sessions. During the experimental session, a 3 × 5 in warning sign was placed on the receiver of the middle telephone with the following message:

CAUTION
TELEPHONE IS OUT OF ORDER
MONEY WILL BE LOST
USE OTHER TELEPHONE

The number of people who used the telephone and the number of those who obviously intended to use it but did not (i.e., approached, looked at the sign, and walked away) was recorded.

Results. Six people used the copy machine during the first session and eight during the third session. During these baseline periods no one with the apparent intent to use the machine left without using it. During the experimental session four people used the machine, and 11 people approached the machine but did not use it. Thus of the people who intended to use the copy machine, 100% did use it when there was no warning on it. But when the warning was posted, only 27% of the people who intended to use the machine did so. This difference was significant, $\chi^2 = 7.54$, $p < 0.01$.

Eight people used the free telephone during the first session and five during the third session. During these baseline sessions no one approached the telephone with the apparent in-

tent to use it and left without doing so (i.e., 100% of the people who had intended to use the telephone used it). During the middle, experimental session no one used the telephone (0%). Seven people approached the telephone with the warning sign on it, looked at the sign, and walked away without using it. The difference between the baseline and the experimental sessions was significant, $\chi^2 = 9.79$, $p < 0.01$.

In both of these demonstration studies, the effectiveness of the warning signs was quite apparent. During the baseline periods everyone who approached the copy machine and telephone used them. When the warnings were posted, only 27% of the people used the copy machine, and no one used the telephone. In short, behavior was changed by the presence of the warnings.

Broken Door

Method. This study illustrates cost of compliance, a variable that moderates warning effectiveness. It was predicted that warnings that do not require very much time or trouble for compliance will be more effective than warnings that require considerably more time and trouble. Behavior was examined in three different sessions, on the campus of a large university. The first warning situation evoked a high cost for compliance. The cost of compliance was moderate for the second session, and it was low for the third session.

For the high cost of compliance condition a warning was posted at about eye level on the left-hand door of a set of glass double doors in a classroom building with the following message:

WARNING
BROKEN DOOR
COULD INJURE
USE ANOTHER EXIT

Complying with this warning would require a considerable detour, as the double doors used for this session are at the end of a hallway leading from the lobby and the nearest alternative exit is in the lobby about 200 feet (61 m) away. Below the sign, in the lower right corner of the door, a piece of brown paper approximately two feet square was taped to the door, giving the appearance that the glass was broken. Baseline and experimental data collection were each conducted in 40-minute sessions.

The moderate and low-cost sessions were conducted in the lobby of a different classroom building, which has two sets of double doors on the same wall about 50 feet (15 m) apart. For the moderate condition a warning was posted on the left door of one of the sets of double doors with the following message:

WARNING
BROKEN DOOR
COULD INJURE
USE ANOTHER EXIT

On this warning an arrow directed users to the other set of double doors about 50 feet away, which created a moderate cost of compliance warning situation. The brown paper was taped to the door in the same manner as in the first session. The number of people who exited through each of the doors was recorded while the sign was posted. Baseline data collection and experimental observation each lasted for a period of one hour.

The low cost of compliance session was conducted at the same door as the moderate condition. The procedure was the same, except that the warning directed users to the single, right-hand door immediately adjacent to the left-hand door on which the warning was

posted. Thus very little effort was required for compliance. The message was as follows:

WARNING
BROKEN DOOR
COULD INJURE
USE OTHER DOOR

An effort was made to control for the variables of time and location. All conditions were conducted at the same time of day. The low-cost session was conducted at the same door as the moderate-cost session; however, in order to achieve a high cost of compliance, it was necessary to conduct the high-cost session at a different door.

Results. The high-cost warning received no compliance at all. Twenty people exited through the door while it was posted; no one obeyed the sign by turning around and returning to the lobby to use another exit. Twenty-five people exited through the door when no sign was posted.

Only three people (5.9%) obeyed the moderate-cost warning that was posted during the second session. Forty-eight people (94.1%) ignored it and exited through the doors instead of using the other exit 50 feet away, as the warning directed. When no warning was posted, 88 people (100%) exited through the doors.

During the lost-cost session 60 people (93.8%) obeyed the sign. Only four people (6.2%) used the door on the left instead of the door on the right. When no warning was posted, 25 people used the door on the left and 20 people used the door on the right, indicating that in general people do not have a tendency to use either the right or the left door. This fairly even distribution of traffic contrasts with behavior in the low-cost condition, in which almost everyone used the right-hand door as directed by the warning message. The low cost of compliance warning was effective in that it produced a change in behavior, $\chi^2 = 32.90$, $p < 0.001$. The high and moderate cost of compliance warnings did not produce any such significant change.

Contaminated Water Fountain

The last field study illustrates another factor that influences the effectiveness of a warning: salience. The more noticeable a warning is, the more effective it will be. Baseline and two experimental observations were made at a water fountain in a university gymnasium in late afternoon. Each session lasted 30 minutes. In the first experimental session an "unenhanced" warning sign was placed on the wall behind the water fountain that read as follows:

WARNING
WATER CONTAMINATED
DO NOT DRINK

The sign was unobscured, at about eye level, and about five by nine inches (12.7 × 22.8 cm) with half-inch (1.3 cm) print. During the second experimental session an "enhanced" sign was posted in the same location at the same water fountain. Several attention-getting characteristics were employed, as illustrated in Figure 1. The size was larger, about 9 × 14 inches (21.6 × 35.5 cm). "Warning" was printed in one-inch (2.5 cm) black letters on an orange background across the top, and the sign also contained a pictorial of a head and torso of a person in profile with the mouth open and the digestive tract revealed (FMC, 1980). The remainder of the message was under the figure in half-inch letters. The "unenhanced" sign was placed on the fountain itself during this session; however, access to the fountain was not blocked.

Figure 1. *Enhanced warning on water fountain. Note that the top panel that serves as a background to the word* WARNING was orange.

Results. During the 30-minute baseline period 22 people drank from the fountain. No one approached the fountain with the apparent intent to drink from it and left without drinking. During the unenhanced warning session, 34 people approached the fountain and 30 people (88%) drank from it. Four people (12%) walked toward the fountain and touched it or otherwise indicated that they wanted to drink, but never did. The effect of this sign was not significant, $\chi^2 = 1.30$, $p > 0.20$.

During the enhanced warning session only five people drank from the fountain. Ten persons stopped at the foutain, bent over it, or touched it but never drank the water. The behavior of 67% of the participants in this session was affected by the enhanced warning. That result was significantly different from that of the baseline session, $\chi^2 = 20.10$, $p < 0.001$.

During the unenhanced warning session many people simply did not seem to notice the sign; the enhanced sign was demonstrated to be a much more salient warning. This finding illustrates the importance of designing warnings that are highly visible, noticeable, and persuasive.

GENERAL DISCUSSION

These studies have provided support for the previously mentioned criteria for warnings. Manipulation of the location of the warning in the chemistry demonstration experiment illustrated that a warning must attract attention. The warning at the beginning of the instructions was more effective (90% compliance) than the warning at the end of the instructions (50% compliance). Obviously, a warning must be seen and read in order to be effective. The presence of a warning on a label or its inclusion in a set of instructions does not guarantee that it will be encountered. Warning labels frequently follow instructions on many products or are located opposite the instructions on a side panel; they may even be "buried" inside owners' manuals. Even when consumers are specifically told to read all instructions before using a product, some may only read the information necessary to perform the task and ignore the warning. This was the case for 33% of the subjects in the "read through" condition in the second chemistry demonstration study.

Another example of the importance of the attention criterion is the field study in which

the salience of the water fountain warning was manipulated. Only the enhanced sign was effective in preventing people from drinking from the fountain. Some of the same methods that were used to enhance the water fountain warning could be used to improve other warnings. The appropriate use of color, size, and a pictorial can increase the impact of the message.

In addition to attracting attention, the ideal warning clearly and concisely describes the hazard and gives instructions for avoiding it. This criterion is supported by the rating studies in which deletion of the hazard and instruction statements led to a greater reduction of perceived effectiveness than did deletion of other parts of the warning. This result does not mean that those statements are more important than the signal word. The usual attention-attracting role of the signal word was reduced in the rating studies: attention of the participants was directed to the warnings by the experimenter, and the signal words were not enhanced. In a field setting the signal word would be more important than it was in the rating studies.

Conciseness or brevity in a warning is important. In some warning situations display space (e.g., product labels) and time to read and comprehend (e.g., highway signs) may be at a premium. Also, people may be more likely to read a short, concise message on a label than a long, wordy one. The rating studies illustrated that deletion of some information is possible without a substantial loss of effectiveness. When brevity is necessary, removing redundant information should be the best place to start.

Even if a warning is salient, informative, and concise, it will not necessarily be effective. People may see a warning, read it, and yet not comply with it. The criterion of motivation must be considered. Cost of compliance—one factor that affects motivation—is examined in the field studies. When the warning directed people to a nearby convenient exit, most of them obeyed it. However, the warning on the doors was not obeyed when a convenient alternative exit was not available. This result suggests that the less time and effort required for compliance by a warning, the more impact it will have on behavior. Cost of compliance can be reduced in many ways. For example, if a warning on a product calls for the use of rubber gloves, the gloves should be provided. Such a provision would decrease the cost of compliance, which might, in turn, increase motivation to comply.

The present research has used a number of methods to examine criteria for effective warnings. The laboratory studies using the chemistry demonstration developed a useful paradigm, providing both a controlled environment and a warning situation that is credible to participants but not hazardous. The rating studies tested a large sample of warnings with a wide range of hazard level. Precedents for using self-report as a methodology have been established in several areas of research in social and engineering psychology (e.g., Ajzen, 1982; Borg, 1978). Indeed, the sign that was posted on the telephone in the field study received a relatively high effectiveness rating (8.1 on the 11-point scale) and was also effective in the field. The field studies provide the strongest test of the ultimate criterion of warning effectiveness, because they examine behavioral change in a real-world setting; the warnings used were obeyed by many people ranging from medical personnel to college students. Guidelines for warning design should be based on studies such as these that represent a wide range of methodology and a variety of appropriate population samples.

ACKNOWLEDGMENTS

The authors would like to thank Alfred T. Crumpton, Robin K. Petry, and David L. Dean for their help in this research, and also to acknowledge the helpful comments on an earlier draft of this paper by John W. Brelsford, Jr.

REFERENCES

Ajzen, I. (1982). On behaving in accordance with one's attitudes. In M. P. Zanna, E. T. Higgins, and C. P. Herman (Eds.), *Consistency in social behavior. The Ontario symposium, Vol. 2* (pp. 131-146). Hillsdale, NJ: Erlbaum.

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

Borg, G. (1978). Subjective aspects of physical work. *Ergonomics, 21*, 215-220.

Collins, B. L., Lerner, N. D., and Pierman, B. C. (1982). *Symbols for industrial safety* (National Bureau of Standards Technical Report NBSIR 82-2485). Washington, DC: U.S. Department of Commerce.

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

FMC Corporation. (1980). *Product safety signs and label system* (3rd ed.). Santa Clara, CA: Author.

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

McCarthy, R. L., Finnegan, J. P., Krumm-Scott, S., and McCarthy, G. E. (1984). Product information presentation, user behavior, and safety. In *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 81-85). Santa Monica, CA: Human Factors Society.

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

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

Westinghouse Corporation (1981). *Westinghouse product safety label handbook*. Trafford, PA: Westinghouse Printing Division.