

BROADENING THE RANGE OF SIGNAL WORDS

N. Clayton Silver
Psychology Department
Appalachian State University
Boone, NC 28608

Michael S. Wogalter
Psychology Department
Rensselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

Most guidelines on warning design recommend using an appropriate signal word that connotes the degree of hazard involved. Usually three levels of signal words, DANGER, WARNING, and CAUTION are suggested for warnings that convey high to low degrees of hazard. The purposes of the present research were threefold. The first goal was to examine whether these terms differed in implied hazard level. The second goal was to determine whether an additional group of five words recommended in guidelines or used in previous research differed in connoted hazard level. The third goal was to explore the possibility of increasing the number and range of words that connote different levels of hazard. Subjects rated a list of 84 potential signal words on six questions assessing strength, severity of implied injury, likelihood of implied injury, attention-gettingness, carefulness, and understandability. The results indicated that DANGER connoted greater strength (arousal) than WARNING and CAUTION, but the results failed to show a difference between WARNING and CAUTION. Among other words tested, DEADLY was seen as having strongest arousal connotation, and NOTE the least. From the long list of 84 terms, a "short" list of 20 signal words was developed based on understandability, low variability, shortness of word, and frequency of use. It is suggested that an expanded list of signal words might alleviate potential problems of habituation from overuse of the currently recommended terms.

INTRODUCTION

Most standards and guidelines on warning design recommend the use of signal words on warnings for the purpose of calling attention to the safety sign and to convey the degree of seriousness of the hazard. Guidelines suggest that two to four levels of hazard be communicated by signal words. The American National Standards Institute guidelines (ANSI, 1972), Z35.1, recommends that the word DANGER be used to indicate hazards of immediate and grave peril and those capable of producing irreversible damage or injury. ANSI recommends the word CAUTION be used on signs to call attention to potential hazards that could result in severe but not irreversible injury or damage. Thus, DANGER is reserved for use on warnings of greater hazard level than the word CAUTION. Other guidelines advocate more than just two levels of hazard. The *Product Safety Sign and Label System* (FMC, 1985) recommends using DANGER, WARNING, and CAUTION with the selection based on the probability and degree of severity. The term DANGER is retained for immediate hazards which *will* result in severe personal injury or death. WARNING is retained for hazards or unsafe practices which *could* result in severe personal injury or death. CAUTION is retained for hazards or unsafe practices which *could* result in minor personal injury or product/property damage. A more recent draft of the ANSI (1988) guidelines, Z535.4, makes similar recommendations. The *Westinghouse Product Safety Label Handbook* (Westinghouse, 1981) extends the list by adding a fourth term, NOTICE, to indicate important but not hazard-related information.

In spite of these recommendations, only a few studies have investigated whether people actually perceive differential strength (arousal) from these signal words. That is, do people know that DANGER means a greater hazard level than WARNING and CAUTION, and that WARNING means a greater hazard level than CAUTION? The research literature is equivocal on this. Bresnahan and Bryk (1975)

found that DANGER expressed a greater level of hazard than CAUTION. However, Leonard, Matthews, and Karnes (1986) found no significant differences between DANGER, WARNING, and CAUTION on perceptions of risk. In addition, other research has also failed to find differences between experimental conditions in which signal words were manipulated (e.g., DANGER vs. CAUTION vs. no signal word in Ursic, 1985; WARNING vs. NOTE in Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein, & Laughery, 1987). In recent research, Leonard, Karnes, and Schneider (1988) used a set of symbols and the words DEADLY, DANGER, WARNING, CAUTION, BE CAREFUL, and ATTENTION. They found strong positive linear relationships between ratings in the symbol and word conditions. From this report, however, it was not clear whether there were significant differences among the words.

Because three levels of signal words, DANGER, WARNING, and CAUTION, are recommended in warning guidelines for the purpose of conveying high to low degrees of hazard, one goal of the present research was to determine whether these three terms differed in implied or connoted strength. Because other words have been either recommended or used in research, the second goal was to determine whether they also differed in implied hazard level.

Because the guidelines limit the number of words to signal hazards, overuse is possible. That is, the terms might appear so often that people habituate to them. Thus, with continued exposure, the signal words may no longer attract attention or signal anything (Cowan, 1988). For example, consider the industrial worker who is frequently exposed to the term DANGER in the course of using job-related equipment. Suppose that new, more hazardous equipment is introduced into the workplace. According to the standards, the term DANGER should be used as the signal word to convey maximal hazard. This term, because of prior

overexposure, may no longer attract attention or adequately communicate the dangerousness of the new equipment. Accordingly, another purpose of the present research was to explore the possibility of enlarging the set of signal words. The goal was to develop a list of words that semantically connote a range of hazard. An objective, criterion-based selection procedure was planned a priori. Using ratings from a large set of potential words, a shorter list of words would be retained using criteria taken from guidelines on warning design including measures of comprehensibility, interpretability, and salience.

METHOD

Subjects

Twenty-eight University of Richmond undergraduates from an introductory psychology course participated to satisfy a research requirement. Another group of 34 undergraduates participated in a subsequent study for the purpose of obtaining rating reliabilities.

Materials and Stimuli

A list of potential signal words were selected from *Roget's Thesaurus* (Morehead, 1982) and a synonym dictionary (Devlin, 1982). From this initial selection, 111 terms representing a wide range of implied strength were obtained. Only single terms were included in the first list (e.g., the term CAREFUL was used instead of BE CAREFUL, cf. Leonard et al., 1988). At this point, some of the words were deleted from further consideration because they appeared to be inappropriate signal words for warnings (e.g., the terms, WOUND, CALAMITOUS, IMPAIR). This determination was made by three judges (the two

authors and an undergraduate research assistant). Words could only be deleted at this stage if all three judges agreed to its unsuitability as a signal word. The retained list of 84 terms, shown in Table 1, were arranged into four randomly-determined orders. Subjects rated the terms on six questions on 9-point Likert-type scales anchored with 0 denoting absence of quantity to 8 indicating maximum quantity. The questions and anchors were:

1. "What is the *STRENGTH* of this term?" The numerical and verbal anchors for this questions were: (0) not at all strong, (2) slightly strong, (4) strong, (6) very strong, and (8) extremely strong.
2. "What is the *SEVERITY* of injury implied by this term?" The numerical and verbal anchors for this questions were: (0) not severe, (2) slightly severe, (4) severe, (6) very severe, and (8) extremely severe.
3. "What is the *LIKELIHOOD* of injury implied by this term?" The numerical and verbal anchors for this questions were: (0) never, (2) unlikely, (4) likely, (6) very likely, and (8) extremely likely.
4. "How *ATTENTION-GETTING* is this term?" The numerical and verbal anchors for this questions were: (0) not at all attention-getting, (2) slightly attention-getting, (4) attention-getting, (6) very attention-getting, and (8) extremely attention-getting.
5. "How *CAREFUL* would you be after seeing this term?" The numerical and verbal anchors for this questions were: (0) not at all careful, (2) slightly careful, (4) careful, (6) very careful, and (8) extremely careful.
6. "How *UNDERSTANDABLE* is this term? In making your rating please consider whether the term would be understood by ALL people in the general population (including young children, visiting foreigners, etc.)?" The numerical and verbal anchors for this questions were: (0) not at all understandable, (2) somewhat understandable, (4) understandable, (6) very understandable, and (8) extremely understandable.

Table 1. Eighty-Four Words Rated by Subjects.

ACIDIC	FORBIDDEN	POISON
ADMONITION	HALT	POISONOUS
ALARM	HARMFUL	PRECARIOUS
ALERT	HAZARD	PRECAUTION
ATTENTION	HAZARDOUS	PREVENT
BEWARE	HEED	PROHIBIT
CAREFUL	HOT	PROHIBITED
CAUSTIC	IMPERATIVE	QUARANTINED
CAUTION	IMPORTANT	REFRAIN
CEASE	INFECTIOUS	RELEVANT
COMPULSORY	INFORMATION	REMINDER
CONSEQUENTIAL	INJURIOUS	REQUIRED
CONTAMINATION	JEOPARDIZE	REQUISITE
CORRODE	LETHAL	RISKY
CORROSIVE	MANDATORY	SERIOUS
CRITICAL	MEMORANDUM	SEVERE
CRUCIAL	NECESSARY	SIGNIFICANT
DANGER	NEEDED	STOP
DANGEROUS	NEVER	SUGGESTION
DEADLY	NO	SUSPEND
DESTRUCTIVE	NOTE	TOXIC
DIRECTIONS	NOTICE	UNPREDICTABLE
DISASTROUS	NOTIFICATION	UNPROTECTED
DISCONTINUE	NOXIIOUS	UNSAFE
DONT	OBLIGATORY	UN SOUND
ESSENTIAL	PARAMOUNT	URGENT
EXPLOSIVE	PERILOUS	VITAL
FATAL	PERTINENT	WARNING

Procedure

Each subject received a different random order of the six questions and one of the four random orders of the terms. Subjects were first told to read the entire list of terms to familiarize themselves with the variety of words listed. They were told to work on one question at a time and to rate all of the words on that particular question before beginning the next question. Subjects were told that even though some of the values on the scales had verbal anchors, they could use any whole number from 0 to 8.

RESULTS

Intercorrelations

The first set of analyses examined the intercorrelations of signal words means (collapsed across subjects) for the six questions. Table 2 shows that the responses to five of the six questions are highly related (r 's ranged from .90 to .96), except understandability. Clearly, these questions are measuring the same thing--what might be termed the arousal quality of the words. Correlations with understandability, while positive, were considerably lower.

To check the reliability of these data, another group of 34 subjects rated the same list of words on the questions of understandability, strength, and carefulness. The reliability of the ratings with the first group of subjects were .933, .904, and .930, respectively.

Table 2. Intercorrelations among questions.

	Strength	Severity	Likelihood	Attention	Careful
Severity	.934				
Likelihood	.915	.961			
Attention	.961	.934	.904		
Careful	.940	.963	.964	.932	
Understand	.342	.249	.318	.385	.321

The word rating data were also examined with respect to objective characteristics of the words: frequency of use in the English language (Thorndike & Lorge, 1944; Francis & Kucera, 1982) and the number of letters and syllables in the words. Understandability was positively correlated with frequency of use ($r = .517, p < .0001$ with the Thorndike-Lorge count; $r = .429, p < .0001$ with the Francis-Kucera count) and negatively related to the number of letters ($r = -.491, p < .0001$) and syllables ($r = -.464, p < .0001$). Thus, the words rated more understandable, are used more frequently and have fewer letters and syllables. However, the correlations of these variables with the other five questions were much smaller: with the Thorndike-Lorge count, the r 's ranged from $-.044$ to $.026$; with the Francis-Kucera count, the r 's ranged from $-.021$ to $.053$; with the number of letters, the r 's ranged from $-.300$ to $-.176$; and with the number of syllables, the r 's ranged from $-.265$ to $-.176$.

Analysis of three signal words

The correlations indicated that five of the questions (strength, severity of injury, likelihood of injury, attention-getting, and carefulness) were measuring the same thing, which we have termed *arousal* (however, any of the five question labels could be substituted). To further examine the overall arousal quality, a repeated measures analysis of variance (ANOVA) using these five questions (collapsed over 84 words) indicated that there was no significant difference among the mean ratings of these questions, $F(4,108) = 1.84, p > .05$. With this evidence, and in order to gain a more powerful, general, or overall measure of arousal, the data was collapsed across the five questions creating arousal mean scores. The arousal scores were used in some of the following analyses to simplify the presentation of results.

One of our primary interests was to determine whether DANGER, WARNING, and CAUTION connoted a differential range of hazard. The top row of Table 3 shows that the term DANGER had a higher arousal mean than WARNING and CAUTION. A one-way repeated-measures ANOVA indicated a significant effect of signal word, $F(2,54) = 11.94, p < .0001$. Subsequent Newman-Keuls range tests showed that DANGER produced significantly higher arousal than either WARNING or CAUTION (p 's $< .05$). WARNING and CAUTION did not differ ($p > .05$). Table 3 also shows the means each of the five arousal questions considered separately. The pattern of question means were similar to the overall arousal mean. The only

Table 3. Mean ratings for Overall Arousal and for each Question as a Function of Signal Word.

	DANGER	WARNING	CAUTION
Arousal mean	6.09	5.31	5.26
Strength	5.89	5.39	5.32
Severity of Injury	6.14	5.18	4.79
Likelihood of Injury	6.04	5.07	5.50
Attention-getting	6.00	5.39	5.32
Carefulness	6.36	5.50	5.39
Understanding	6.86	6.46	6.14

exception to this was CAUTION receiving greater ratings for likelihood of injury than WARNING. The ANOVAs were significant for severity of injury, $F(2,54) = 13.01, p < .0001$, likelihood of injury, $F(2,54) = 6.59, p < .003$, and carefulness, $F(2,54) = 5.89, p < .005$. Subsequent Newman-Keuls range tests for these effects showed that the only significant differences were between DANGER and WARNING or CAUTION (p 's $< .05$). The reversal of the CAUTION and WARNING means for likelihood of injury was not significant.

The understandability means are shown on the bottom row of Table 3. The repeated-measures ANOVA indicated a significant effect of word, $F(2,54) = 5.78, p < .006$. Subsequent Newman-Keuls range tests showed only one reliable difference: DANGER received significantly greater understandability ratings than CAUTION ($p < .05$).

Analysis of other potential words

Another interest of the present research was whether other terms recommended or used in previous research would differ in their arousal quality. In these analyses, the words ATTENTION, CAREFUL, DEADLY, NOTE, and NOTICE were added to the data set that included DANGER, WARNING, and CAUTION. The overall arousal level means are shown in the first column of Table 4. The repeated measures ANOVA indicated a significant effect of word, $F(7,189) = 83.02, p < .0001$. Subsequent Newman-Keuls range tests showed that all differences among the overall arousal means were significant (p 's $< .05$) except between ATTENTION and CAREFUL, and as shown earlier, between WARNING and CAUTION.

The understandability of the eight words were also examined and the means are shown in the second column of Table 4. The ANOVA showed a significant effect, $F(7,189)$

= 5.84, $p < .0001$. Subsequent Newman-Keuls range tests showed that NOTICE was significantly less understandable than all of the other words except for NOTE. NOTE was also significantly less understandable than WARNING, CAREFUL, and DANGER (p 's $< .05$). There were no other reliable differences between words on the dimension of understandability.

Table 4. Overall Arousal and Understandability Means for 8 Words.

	Arousal	Understandability
DEADLY	7.34	6.07
DANGER	6.09	6.86
WARNING	5.31	6.46
CAUTION	5.26	6.14
CAREFUL	3.81	6.50
ATTENTION	3.45	6.29
NOTICE	2.80	5.25
NOTE	2.12	5.50

Selection of signal words

Efforts were also directed at the development of a "short" list of signal words that covered a range of implied strength. An objective set of criteria was used in the selection process that first began with the list of 84 words. Because most guidelines emphasize comprehensibility, we deleted terms that received mean understandability ratings in the bottom one-third (33%) of the list. Another aspect of comprehensibility is interpretation. It is desirable to have terms that have a consistent meaning; that is, words that people interpret in the same way. Therefore, terms were deleted for which the summed variance of the five individual arousal questions exceeded an apparent breakpoint in the data (sumvar = 20.0). Also, deleted were terms for which Thorndike and Lorge (1944) and Francis and Kucera (1982) showed no occurrence or very infrequent occurrence (less than 2 per million words) in the English language. Lastly, word size was considered in the selection process. Because the size of a signal word affects its salience (i.e., larger type is more discriminable) and label or sign space is often limited, terms were deleted having greater than seven letters. Using these criteria, a "short" list of 20 signal words was formed from the original list of 84. This list is shown in Table 5 ordered on arousal-quality. This table also shows the individual question means and standard deviations.

DISCUSSION

The results showed that DANGER implied a more serious hazard than the word CAUTION, confirming Bresnahan and Bryk's (1975) finding. DANGER also implied greater hazard than WARNING. However, the present results failed to show that WARNING implied greater hazard than CAUTION. These two words basically mean the same thing; that is, people do not distinguish the terms with regard to hazard level. This result suggests that people do not discriminate between these two terms in the way they are purported in current standards and guidelines. One possible reason for this finding is that lay people process the meaning of the terms WARNING and CAUTION using similar lexical information. By lexical

definition, warning means "something serving to warn, caution, or admonish" and caution means "a warning or admonition" (New Webster's Dictionary of the English Language, 1975). Apparently, people have acquired a similar connotation for these two words from personal experiences rather than from information received directly from warning signs.

The present research also examined the arousal quality of an additional set of signal words that were chosen for analysis because they had been used in previous research or recommended in guidelines. It was also shown that the word DEADLY possessed a significantly greater arousal quality than the other signal words. NOTE and NOTICE implied the least hazard. All of the words showed a range of significantly different hazard levels, except for between ATTENTION and CAREFUL and between WARNING and CAUTION. The ordering of words corroborate the rankings of Leonard et al. (1988), although the rank order of ATTENTION and CAREFUL (or BE CAREFUL, cf. Leonard et al., 1988) was reversed.

Because the guidelines recommend a limited number of words to signal many hazards, their overuse is a problem. The arousal quality of the words might be reduced with continued exposure: They may no longer attract attention when reexposed in the same and other situations. For example, the term DANGER may not communicate extreme levels of dangerousness. One way to alleviate possible habituation effects might be to use a word with greater novelty but having similar or appropriate arousal properties, thus augmenting the signal word's attention-gettingness. A list of 20 words that semantically connote a range of hazard was developed based on measures of comprehension, interpretation, and salience. This list might be used as a starting point for subsequent warnings research and development, and might find use in assisting in the determination of the appropriate signal word for a given hazard. It should be emphasized, however, that several other important factors must be kept in mind when selecting the appropriate signal word: the hazard involved (the severity of injury and likelihood of injury), the suitability of the word (POISON makes no sense on an electrical hazard), person factors (target group population, knowledge and familiarity), and environmental factors (circumstances and surroundings).

There are a number of limitations to consider when interpreting the results of the present study. First, the signal words were presented out of context. Subjects just rated a series of words, making it difficult to know how the arousal quality of the words would be affected when placed in actual warnings in real-world settings. For example, is arousal enhanced or diminished when placed in warnings with appropriate or inappropriate levels of hazard? That is, do the words interact with other variables (e.g., symbols, different types of hazards)? Second, college freshmen and sophomores rated the words. Thus, it is difficult to know whether other groups of people interpret the words in the same way. The results did concur with the differential hazard interpretation of DANGER and CAUTION found by Bresnahan and Bryk (1975) who used industrial workers. The words, however, need to be evaluated using other target populations to determine their interpretation and understandability. For example, how would persons who do not have strong command of the English language interpret these words? Third, the list of 20 words that we

Table 5. Twenty Objectively-Selected Words Sorted on Arousal.

	Arousal		Strength		Severity		Likelihood		Attention		Carefulness		Understand	
	mean		mean	STD	mean	STD	mean	STD	mean	STD	mean	STD	mean	STD
NOTE	2.12		2.46	2.03	1.61	1.97	1.64	1.45	2.57	1.93	2.32	2.07	5.50	2.12
NOTICE	2.80		2.89	1.59	2.39	1.99	2.75	1.90	2.96	2.03	3.00	1.80	5.25	1.65
NEEDED	3.16		3.29	2.03	2.86	1.69	2.82	1.79	3.25	1.55	3.57	2.40	5.86	1.74
PREVENT	3.67		4.21	1.59	3.25	1.97	3.36	2.38	3.71	1.70	3.82	1.89	4.86	1.92
CAREFUL	3.81		3.96	2.19	3.32	1.96	3.54	1.99	3.57	1.79	4.68	1.96	6.50	1.26
ALERT	4.47		4.89	1.97	3.93	1.65	4.36	1.97	4.82	1.98	4.36	2.13	5.61	1.55
ALARM	5.19		5.61	1.71	4.64	1.99	5.04	1.69	5.32	1.59	5.36	1.75	5.96	1.64
CAUTION	5.26		5.32	1.85	4.79	1.75	5.50	1.67	5.32	1.76	5.39	2.06	6.14	1.43
HARMFUL	5.26		4.86	1.82	5.50	1.75	5.68	1.66	5.07	1.74	5.21	1.83	5.93	1.58
WARNING	5.31		5.39	1.99	5.18	1.83	5.07	1.58	5.39	1.83	5.50	1.35	6.46	1.67
BEWARE	5.41		5.32	1.95	5.46	1.43	5.36	1.70	5.50	1.75	5.39	2.01	6.57	1.91
URGENT	5.41		6.00	1.83	4.93	1.82	4.21	2.35	5.86	1.63	6.04	1.43	5.43	1.55
SERIOUS	5.51		5.46	1.44	6.04	1.62	5.50	1.75	4.79	1.81	5.79	1.93	6.07	1.58
SEVERE	5.55		5.89	1.59	5.68	1.39	5.75	1.92	4.93	1.74	5.50	1.79	4.25	1.69
VITAL	5.80		6.36	1.39	5.68	1.76	5.43	2.28	5.86	1.82	5.68	1.81	4.18	1.59
HAZARD	5.84		6.04	1.60	5.54	1.86	5.68	1.83	5.79	1.66	6.18	1.39	5.21	2.02
DANGER	6.09		5.89	2.02	6.14	1.69	6.04	1.83	6.00	2.09	6.36	1.61	6.86	1.56
POISON	6.74		6.54	1.37	6.79	1.87	6.75	1.69	6.64	1.81	7.00	1.63	6.93	1.46
FATAL	7.20		7.54	0.69	7.39	1.57	7.07	1.30	7.04	1.35	6.96	1.97	5.75	2.01
DEADLY	7.34		7.11	1.59	7.68	0.86	7.29	1.46	7.11	1.47	7.54	1.07	6.07	1.98

present is just that -- a list having certain criteria. Using other criteria, other lists could be formed. It is noted, however, that in lists using different criteria (e.g., number of syllables and different cut-off points), we arrived at lists that were similar to the one presented here.

ACKNOWLEDGEMENTS

The authors would like to thank Megan E. McGuire for her assistance in this research and Stephen L. Young for help on the final copy.

REFERENCES

ANSI (1972). *American National Standard Specification for Accident Prevention Signs: Z35.1*. New York: American National Standards Institute.
 ANSI (1988). *American National Standard on Product Safety Signs: Z535.4 - Draft*. New York: American National Standards Institute.
 Bresnahan, T. F., & Bryk, J. (1975). The hazard association values of accident-prevention signs. *Professional Safety*, January, 17-25.
 Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104, 163-191.

Devlin, J. (1982). *A Dictionary of Synonyms and Antonyms*. New York: Warner Books.
 FMC (1985). *Product Safety Sign and Label System*. Santa Clara, CA: FMC Corporation.
 Francis, W. N., & Kucera, M. (1982). *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.
 Leonard, S. D., Karnes, E. W., & Schneider, T. (1988). Scale values for warning symbols and words. In F. Aghazadeh (Ed.) *Trends in Ergonomics/Human Factors V*. Amsterdam: Elsevier. 669-674.
 Leonard, S. D., Mathews, D., & Karnes, E. W. (1986). How does the population interpret warning signals? *Proceedings of the 30th Annual Meeting of the Human Factors Society*, Santa Monica, CA: Human Factors Society, 116-120.
 Morehead, A. H. (1982). *Roget's College Thesaurus*. New York: World Publishing.
New Webster's Dictionary of the English Language. (1975). New York: Consolidated Book Publishers.
 Thorndike, E. L., & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College Press, Columbia University.
 Ursic, M. (1984). The impact of safety warnings on perception and memory. *Human Factors*, 26, 677-682.
 Westinghouse (1981). *Westinghouse Product Safety Label Handbook*. Trafford, PA: Westinghouse Printing Division.
 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.