ERGONOMICS, 1995, VOL. 38, NO. 11, 2188–2206

Warning signal words: connoted strength and understandability by children, elders, and non-native English speakers

MICHAEL S. WOGALTER

Department of Psychology, North Carolina State University, Raleigh, NC 27695, USA

N. CLAYTON SILVER

Department of Psychology, University of Nevada, Las Vegas, NV 89154, USA

Keywords: Consumer product safety; Equipment safety; Product labelling; Warning; Hazard.

Signal words, such as DANGER, WARNING and CAUTION, are commonly used in sign and product label warnings for the purpose of conveying different levels of hazard. Previous research has focused on whether people's perceptions of connoted hazard are consistent with the levels suggested by design standards and guidelines. Most investigations have used college students to evaluate the terms; other populations who may be at greater risk have not been adequately studied. One purpose of the present research was to determine whether young children, the elderly, and non-native English speakers perceive similar connoted hazard levels from the terms as undergraduates and published guidelines. A second purpose was to assess the terms' comprehensibility using various metrics such as missing values (i.e. ratings left blank) and understandability ratings. A third purpose was to develop a list of potential signal words that probably would be understandable to members of special populations. In the first experiment, 298 fourth- to eighth-grade students and 70 undergraduates rated 43 potential signal words on how careful they would be after seeing each term. The undergraduates also rated the terms on strength and understandability. In the second experiment, 98 elders and 135 non-native English speakers rated the same set of terms. The rank ordering of the words was found to be consistent across the participant groups. In general, the younger students gave higher carefulness ratings than the undergraduates. The words that the younger children and the non-native English speakers frequently left blank were given lower understandibility ratings. Finally, a short list of terms was derived that 95% or 99% of the youngest students (fourth- and fifth-graders) and 80% of the non-native English speakers understood. Implications of hazard communication are discussed.

1. Introduction

Most standards and guidelines on warning design recommend the use of signal words on signs and labels for the purpose of quickly conveying the level of hazard involved to persons at risk. The standards usually recommend three terms, DANGER, WARNING and CAUTION, to be used as signal words connoting highest to lowest levels of hazard, respectively (ANSI 1991, FMC Corporation 1985). DANGER is intended to call attention to situations that involve immediate hazards that *will* result in *severe* personal injury or death; WARNING is intended to be used for hazards that *could* result in *severe* personal injury or death; and CAUTION is intended for hazards that *could* result in *minor* personal injury of damage (FMC Corporation 1985).

However, research in this area is equivocal on whether people actually interpret different levels of hazard from these words. Some research supports the hazard-level denotations in standards and guidelines. For example, several studies have shown that DANGER connotes a higher level of hazard than CAUTION or WARNING (Bresnahan and Bryk 1975, Chapanis 1994, Dunlap *et al.* 1986, Wogalter and Silver 1990). However, other studies have not found differences among these and other signal words) (Leonard *et al.* 1986, Ursic 1984, Wogalter *et al.* 1987, Wogalter *et al.* 1994). These latter indeterminate results call into question the utility of signal words to convey particular hazard levels as they are propounded in warning-design guidelines and standards.

Most of the research on signal words have used upper grade-level (high school or college) students as subjects. For example, Tajima *et al.* (1991) found that students at Tokyo University perceived literal translations of common signal words similarly to US college students. One exception to the widespread use of college students is a study by Dunlap *et al.* (1986), who used customer engineers and service representatives in the US, Canada, and Belgium, as well as US adults without technical training. They found that the different subject populations perceived the hazard levels of the signal words somewhat differently. This result suggests that the utility of the terms in warnings might differ depending on who is looking at them. Because many consumer products are intended for the general public, it is necessary to know hazard-related messages such as signal words are interpreted by people other than college students, and more specifically, whether other population groups understand and differentiate between them.

The populations of greatest concern would be those that contain individuals who have the highest risk of potential injury. One such population is children. They lack the experience and education of adults and, as a consequence, children are disproportionately involved in accidents. It is estimated that over three million children in the USA are treated in hospital emergency rooms for injuries sustained when using consumer products (US Consumer Product Safety Commission 1980). The best method of protecting children is to remove the hazard altogether from their environment. However, households with children usually contain many hazardous products that are not always kept secure. Many of these products contain printed-language warnings that are usually directed to adult populations. However, the instructions and warnings on products are often so complex and technical that many adults question these materials' (and their own) adequacy. Certainly, print warnings are a poor method of directly controlling children's use of hazardous products, because it can not be expected that they will understand their semantic content. Yet, children and others lacking proficiency with the English language may be capable of understanding sections of labels if the warnings contain terms that these individuals are likely to know. Thus, one way to communicate that a product is dangerous is to use specific words that individuals with lower-level language abilities are likely to interpret as intended.

Relative to language-proficient adults, language-deficient individuals may not readily discriminate small differences among most of the signal words. For example, children may not differentiate among the words in the same way that adults with good language skills might. A selected subset of terms may be useful in facilitating accurate hazard level discriminations (where accuracy is based on evaluations made by language-proficient individuals).

Besides young children, there are other at-risk populations who might fail to understand and differentiate among signal words. These include elderly and non-native

English speakers. Elderly individuals are more likely to have visual or cognitive impairments (Owsley *et al.* 1991, Schlag 1993) than younger individuals, and are a disproportionately-represented population in serious accidents (Owsley *et al.* 1991). This also applies to individuals who are recent immigrants or visitors to an English-speaking country and who are non-native speakers of English who would be expected to be less proficient in the English language and to have exposure to signal words than native speakers. Therefore, the non-native English speakers would be less likely to discriminate among the signal words in the same way that (or as well as) native speakers of the language would. Hence, it would be desirable to use a set of signal words in warnings that these special populations are likely to understand and discriminate.

If individuals do not understand the level of hazard implied by a signal word, then they may not exhibit the appropriate care and injury may result. If an injury does occur under these circumstances, then litigation may ensue against the manufacturer for not communicating the hazard in a prudent fashion (e.g. Bituminous Casualty Corp. *v.* Black & Decker Mfg. Co. 1974). In the USA, the California Court of Appeals, Fifth Appellate District, ruled that if non-English speaking people are likely to use a product, then the manufacturer may need to provide foreign language warnings (Ramirez *v.* Plough Inc. 1992). Although it would be difficult to provide warnings that would encompass every language and nationality, perhaps certain English signal words that are familiar to children and non-native English speakers, like STOP and DON'T would be appropriate. Moreover, if a signal word is chosen that fails to communicate or minimizes the extent of the hazard (such as using the word NOTE for radioactive substances), then the manufacturer could also be found to be liable (Lopez *v.* Aro 1979).

Thus, one purpose of the present research was to determine the perceived hazard levels of a set of signal words using three special populations: grade-school children, the elderly, and non-native English speakers. The data from these groups were then compared to (a) native-English speaking college students, and (b) definitions in warning-design guidelines and standards.

Another consideration with regard to the utility of signal words (as opposed to the hazard level connotation) is comprehension. Owing to their lower English-language skills, children, elderly, and non-native speakers are less likely to understand certain terms that might be readily understandable by more highly-skilled English-language users. Thus, it is important to determine the extent to which the signal words are understood by various populations. In earlier research, Wogalter and Silver (1990) evaluated potential signal words using comprehension measures such as rated understandability, rating variability, and frequency of occurrence in the English language. Another indication of comprehension is the number of persons who are unable to give ratings to the terms because they do not understand them. Words that are frequently left blank are less likely to be useful in communicating hazards. In the present research, this missing-values measure is evaluated and compared to the understandability ratings of college students, as well as other objective measures such as language frequency. If the missing-values measure is valid, it could be used to evaluate the appropriateness of terms as signal words for the general population, and be used to derive a list of words for warnings that are likely to be understood by less able English readers.

In summary, the purposes of the present research are threefold. The first is to determine whether signal words are interpreted by grade-school children, the elderly, and non-native English speakers in the same manner as college students. It is hypothesized that the interpretations of these groups will reflect the denoted definitions

prescribed in warning design guidelines and standards. The second purpose is to examine the understandability of signal words using several metrics that might be used as part of a set of criteria for selecting suitable signal words for a broad range of populations. These include: (a) low of frequency of being left blank (not rated) by respondents (i.e. fewer missing values), (b) high ratings of understandability by college students, (c) high frequency of appearance in written language (as enumerated in an analysis of English-language literature), (d) low rating variability (i.e. consistency), and (e) shorter word length. The third purpose is to develop a concise list of terms that most individuals who are less skilled with the English language are likely to know and understand.

2. Experiment 1

In the first experiment, signal word evaluations were made by elementary and middle-school children (fourth- to eighth-grades), which were compared to the judgements of college students.

2.1. *Method*

2.1.1. *Participants*: A total of 298 elementary and middle school students from Allegheny County, North Carolina (56 fourth-, 60 fifth-, 50 sixth-, 68 seventh-, and 64 eighth-grade) participated in the experiment. In addition, 70 Appalachian State University (ASU) undergraduates from introductory psychology courses completed an expanded version of the younger students' questionnaire.

2.1.2. *Stimuli and procedure*: Forty-three words were selected from a list of 84 terms studied by Wogalter and Silver (1990). Selection from the original list was based on two criteria: (a) word length, and (b) understandability. With respect to the first criterion, words having more than nine letters were excluded based on the practical consideration that there is often limited label or sign space for warnings. A shorter signal word can have larger characters, which is likely to provide a better visibility at a distance and capture attention. Moreover, holding character size constant, a shorter signal word would provide more space for the other warning components (e.g. hazard, consequence, and instruction statements) than a longer word. With respect to the second criterion, words having mean understandability ratings above 4·0 (rated 'understandable' and above on the scale) from Wogalter and Silver (1990) were included. The purpose of this criterion was to limit the number of evaluations that participants would make and to include the most understandable terms on this list.

Participants were given the list of terms and were told to examine the entire set of words before starting the ratings. Grade-school students were given a separate sheet that contained the question: 'How *careful* would you be after seeing this term?' A 9-point rating scale was provided with the even-numbered anchors having the verbal labels: (0) not at all careful, (2) somewhat careful, (4) careful, (6) very careful, and (8) extremely careful. Participants were told that the verbal labels were to help them make their ratings and that they could use any whole number from 0 to 8. The instructions emphasized that they should not rate words that they did not understand (i.e. leaving these words blank).

College subjects rated the terms on the same carefulness question and two other questions (strength and understandability) on 9-point scales: 'What is the *strength* of this term?' with the anchors: (0) not all strong, (2) somewhat strong, (4) strong, (6) very

strong, and (8) extremely strong; and 'How *understandable* is this term?' with the anchors: (0) not at all understandable, (2) somewhat understandable, (4) understandable, (6) very understandable, and (8) extremely understandable. For the latter question, the college students were to consider in their judgements whether all people in the general population (including young children, visiting foreigners, etc.) would understand the terms.

All participants received one of two random word orders. In addition, the college students were assigned randomly to one of the six possible question orders, and rated all 43 words on one question before making evaluations on the next question.

2.2. Results

Several analyses described below used repeated-measures designs to compare terms. In all analyses described in this and the next experiment, casewise deletion was used. That is, participant scores with missing data relevant to that analysis were not included. Analyses specifically examining the missing data are described later.

2.2.1. *Analysis of the carefulness ratings*: Carefulness ratings were collapsed across subjects for each grade level separately (including the ASU undergraduates) to form mean scores for each word. These means are shown in table 1. These scores were then correlated with the mean arousal strength ratings derived from University of Richmond (UR) undergraduates in an earlier study (Wogalter and Silver 1990). In general, the correlations showed an increase with grade levels ($r = 0{\cdot}78$, 0·86, 0·93, 0·90, 0·94 and 0·94 for the fourth- to eighth-graders and ASU students, respectively, $ps < 0{\cdot}0001$). Using a procedure by Rao (1970) for simultaneously evaluating multiple independent correlations, a significant difference among grade levels was noted, $\chi^2(5, N = 43) = 15{\cdot}52$, $p < 0{\cdot}0005$. Subsequent multiple range tests (Levy 1976, Silver and Burkey 1991) indicated that the correlation of the fourth-grade students' ratings were significantly lower than the correlations of the seventh-grade, eighth-grade, and ASU college student' ratings ($p < 0{\cdot}05$). Moreover, ASU students' strength ratings were highly correlated with their carefulness ratings, $r = 0{\cdot}98$, $p < 0{\cdot}0001$, and with the overall arousal strength ratings of the UR students of Wogalter and Silver (1990), $r = 0{\cdot}93$, $p < 0{\cdot}0001$.

A 6 (grade levels: 4 to 8 plus college students) $\times$ 8 (signal words: NOTE, ATTENTION, NOTICE, CAREFUL, CAUTION, WARNING, DANGER, DEADLY) analysis of variance (ANOVA) was performed using carefulness as the dependent variable. The eight signal words included the three most common terms (DANGER, WARNING and CAUTION) plus five others that have been investigated in previous signal word research (Leonard *et al.* 1988, Leonard *et al.* 1989). These terms are underlined in table 1. The three common terms are shown in bold print. The ANOVA showed a significant main effect of grade, $F(5, 333) = 4{\cdot}40$, $p < 0{\cdot}01$. Subsequent Newman–Keuls tests showed that: (a) both the fourth- and fifth-graders gave significantly higher ratings than the sixth-graders and college students, and (b) the seventh-graders gave significantly higher ratings than sixth-graders ($ps < 0{\cdot}05$).

There was also a significant main effect of signal word, $F(7, 2331) = 180{\cdot}95$, $p < 0{\cdot}0001$, with DEADLY, DANGER, WARNING, CAUTION, CAREFUL, NOTICE, ATTENTION and NOTE rated from the greatest to least. This order was consistent among all groups, W (Kendall's Coefficient of Concordance) = 0·97, $p < 0{\cdot}01$. Subsequent Newman–Keuls tests showed significant differences among all

Table 1. Carefulness means of signal words by student grade level and overall. Also included are the overall standard deviations, missing value percentages, and understandability ratings.

	Carefulness by grade level						Overall carefulness			Understandability
Word*	4th	5th	6th	7th	8th	ASU	Mean	SD	% missing	ASU college students
NOTE	4·80	4·20	2·80	4·00	3·65	3·37	3·80	2·57	4·1	4·01
REMINDER	4·96	4·07	3·18	4·09	4·07	3·53	3·97	2·50	3·3	4·36
NEEDED	4·94	3·66	3·30	4·37	4·41	4·09	4·14	2·54	1·9	4·81
REQUIRED	4·79	4·31	3·85	4·29	4·36	4·27	4·31	2·48	6·0	4·30
NECESSARY	5·38	4·26	3·66	4·71	4·36	4·10	4·41	2·61	2·4	4·66
ATTENTION	5·16	5·05	3·96	4·64	4·65	4·30	4·62	2·37	3·8	5·04
NOTICE	5·25	5·51	3·86	5·12	4·44	4·01	4·70	2·41	1·1	4·41
PREVENT	5·84	4·96	4·06	5·33	5·07	4·40	4·93	2·29	3·8	4·24
HALT	4·72	5·25	4·22	5·52	4·93	5·61	5·12	2·35	10·1	4·81
NO	5·77	5·50	4·20	4·74	4·91	5·60	5·13	2·62	0·8	7·24
IMPORTANT	6·31	5·61	4·76	5·25	4·90	5·06	5·30	2·38	1·6	5·53
DON'T	6·30	5·95	4·26	5·19	5·13	5·24	5·36	2·48	0·5	6·09
CAREFUL	5·84	5·88	4·94	5·65	5·20	4·76	5·37	2·35	0·3	5·49
PROHIBIT	6·24	5·98	4·89	5·36	5·04	5·30	5·43	2·14	10·3	4·46
RISKY	6·65	5·86	5·16	5·46	5·16	5·14	5·55	2·21	1·1	4·69
URGENT	6·21	5·06	4·94	5·76	5·55	5·73	5·55	2·31	4·9	4·99
ALARM	6·55	5·80	5·38	5·60	5·64	5·01	5·63	2·26	1·1	5·31
NEVER	5·71	6·45	4·98	5·79	5·53	5·93	5·75	2·39	0·8	6·07
STOP	6·09	6·14	4·80	5·69	5·33	6·43	5·78	2·50	0·3	7·09
ALERT	6·75	6·17	5·22	5·84	5·53	5·33	5·79	2·23	1·6	5·26

HOT	6·02	5·98	5·12	5·96	5·50	6·21	5·83	2·42	0·0	6·54
VITAL	6·29	5·90	5·72	5·80	5·96	5·60	5·86	2·27	9·8	4·37
FORBIDDEN	5·94	6·20	5·72	6·06	5·68	5·81	5·90	2·20	5·2	5·03
CRUCIAL	6·29	6·46	5·89	5·77	5·90	5·50	5·92	2·19	13·0	4·44
INJURIOUS	6·04	6·48	5·98	6·23	5·95	5·37	5·98	2·12	11·1	3·74
UNSAFE	6·75	6·20	5·86	6·13	5·64	5·46	5·98	2·18	0·5	5·59
CAUTION	**6·37**	**6·90**	**5·80**	**6·25**	**5·88**	**5·22**	**6·05**	**2·12**	**1·6**	**5·12**
BEWARE	6·82	6·50	5·34	6·31	5·97	5·77	6·13	2·09	0·5	5·71
SEVERE	6·12	5·85	6·00	6·42	6·45	6·23	6·20	2·25	6·8	5·06
WARNING	**6·62**	**6·42**	**5·86**	**6·26**	**5·88**	**6·13**	**6·20**	**1·97**	**0·0**	**5·60**
HAZARD	6·60	6·54	6·10	6·21	6·59	5·67	6·27	2·04	1·9	5·36
HARMFUL	6·82	6·67	6·10	6·28	6·48	5·94	6·37	2·02	0·0	5·69
SERIOUS	7·17	6·64	6·20	6·45	6·27	5·73	6·38	1·92	1·1	5·54
CRITICAL	6·63	6·56	6·53	6·58	6·46	6·03	6·44	2·04	7·6	4·60
LETHAL	6·02	6·38	6·27	6·44	6·74	7·41	6·61	2·12	10·6	5·00
DANGER	**7·38**	**6·88**	**6·40**	**6·57**	**6·44**	**6·49**	**6·68**	**1·80**	**0·0**	**6·67**
HAZARDOUS	7·00	7·07	6·80	6·94	6·51	6·24	6·74	1·87	3·3	5·46
DANGEROUS	7·29	7·08	6·34	6·90	6·52	6·64	6·79	1·70	0·5	6·31
FATAL	6·78	7·04	6·63	6·69	7·00	7·36	6·94	1·82	5·7	5·57
TOXIC	6·76	6·72	6·88	6·86	7·21	7·17	6·95	1·79	3·5	4·81
POISON	7·70	7·30	7·12	7·19	7·12	7·00	7·23	1·67	0·0	6·44
EXPLOSIVE	7·54	7·48	7·14	7·54	7·09	7·01	7·29	1·54	1·4	5·44
DEADLY	7·89	7·27	7·44	7·65	7·62	7·30	7·53	1·30	0·3	6·24
Mean	6·26	6·01	5·34	5·86	5·69	5·59				
n	56	60	50	68	64	70				

* Separate analyses of the bold and underlined words are described in the text. Words are ordered by the overall carefulness mean.

pairwise comparisons ($p < 0{\cdot}0001$), except between WARNING and CAUTION and between ATTENTION and NOTICE.

There was also a significant grade level × signal word interaction, $F(35, 2331) = 2{\cdot}01$, $p < 0{\cdot}001$. Simple effects analysis (Satterthwaite 1946) indicated that every grade level produced a significant effect of word ($ps < 0{\cdot}0001$). The means showed the same consistent word order across all grade levels as described above for the main effect, with two exceptions: (a) the fifth graders gave CAUTION the second highest mean rating, but the Newman–Keuls test showed that this term was not significantly different from either DANGER or WARNING ($p > 0{\cdot}05$); and (b) the sixth-graders, eighth-graders, and college students gave NOTICE the second lowest mean rating, but it was not significantly different from ATTENTION ($p > 0{\cdot}05$). For the ASU students, DANGER and WARNING had significantly higher carefulness ratings than CAUTION, $ps < 0{\cdot}05$. Simple effects analysis also showed significant differences among grade levels for NOTE, NOTICE, CAREFUL and CAUTION ($ps < 0{\cdot}05$), but not for the other words ($p > 0{\cdot}05$). In general, younger students gave higher carefulness ratings than the older students, except that (a) NOTE and NOTICE received significantly lower ratings by the sixth-graders than by the seventh-graders ($p < 0{\cdot}05$), and (b) CAUTION received significantly higher ratings by the fifth-graders than by the fourth-graders ($p < 0{\cdot}05$).

2.2.2 *Analysis of understandability*: The correlation of the understandability ratings between the college students of the present study (ASU undergraduates) with those from the Wogalter and Silver (1990) study (UR undergraduates) was 0·82, $p < 0{\cdot}0001$. This result indicates that there is good reliability between the two studies and the undergraduates of different universities. Of greater importance, however, is whether this subjective evaluation of understandability reflects actual understandability. As mentioned earlier, participants were told that if they did not understand some of the terms, then they should not rate them. Thus, one objective measure of the terms' understandability is the number of ratings that were actually provided, or in other words, the inverse of the number of ratings that were left blank. Missing data for the undergraduates was extremely infrequent, so the greatest percentage of missing data shown in table 1 is attributable to the grade-school students. The correlations of the grade-school students' missing data with the ASU students' understandability ratings were $-0{\cdot}59$, $-0{\cdot}64$, $-0{\cdot}50$, $-0{\cdot}52$, and $-0{\cdot}66$, for the fourth- to the eight-graders respectively ($p < 0{\cdot}0001$), and the correlation of the overall percentage missing to the understandability ratings was $-0{\cdot}64$ ($p < 0{\cdot}0001$). That is, the more understandable the word, the fewer missing ratings.

Other objective indications of understandability were also examined: frequency of occurrence in the English language, variability in the ratings, and letter length. First, frequency of occurrence in the English language was thought to be related to understandability because frequent words are generally learned before less frequent words. The undergraduates' understandability rating had correlations of $rs = 0{\cdot}46$ and 0·53 to the Thorndike and Lorge (1944) and Francis and Kucera (1982) word frequency counts, respectively ($ps < 0{\cdot}0001$). However, the missing-value counts for each of the separate grade levels were not strongly correlated to these frequency counts (rs ranging from $-0{\cdot}16$ to $-0{\cdot}38$). The overall missing-value percentages were significantly related to the Thorndike–Lorge counts ($r = -0{\cdot}36$, $p < 0{\cdot}05$), but not to the Francis–Kucera counts ($r = -0{\cdot}22$, $p > 0{\cdot}05$).

Second, variability in understandability ratings are another indication of how well

the terms are understood. If a signal word is consistently rated to be a certain level of hazard, then the word can be considered to reliably convey a single meaning across participants. To the extent that the ratings vary greatly, it indicates that the term is not well understood, or at least, does not have a consistent concept across the group of raters. One measure of this variability is the overall standard deviation of the carefulness ratings. However, correlations of this overall measure with the other measures of variability were moderate (Cohen 1969) with relation to the word frequency measures (rs of 0·38 and 0·39 with the Thorndike–Lorge and Francis-Kucera counts, respectively, $p < 0·05$). Finally, the relationships of word size (letter counts) to the other understandability metrics were examined. Letter length was significantly related to the understandability ratings ($r = -0·39$), and the Thorndike–Lorge ($r = -0·34$) and Francis–Kucera ($r = -0·41$) counts, $p < 0·05$, but not to the missing-value and variability measures. Furthermore, there were no significant linear relationships between the understandability ratings of the ASU students and their carefulness or strength ratings (ps $> 0·05$), as would be expected if comprehension and the hazard-related ratings are different evaluative dimensions.

Additional analysis examined the proportion of missing values for all 43 terms as a function of grade level. College students were excluded from these analyses because they produced very few missing values. In general, the percentage of missing values decreased with increasing grade level (Ms = 9, 5, 2, 2 and 3%, for the fourth- to eighth-grades, respectively). As shown in table 2, words such as CRUCIAL, HALT, and PROHIBIT were missed by over one-quarter of the fourth graders. However, terms such as HOT, DON'T, and STOP were missed by none of the fourth graders. A one-way ANOVA on these data showed a significant effect of grade level, $F(4, 168) = 34·42$, $p < 0·0001$. Subsequent Newman–Keuls tests showed that the fourth-graders had more missing ratings than the fifth-graders ($p < 0·05$), and that both the fourth- and fifth-grade students had more missing ratings than the students of the higher grade levels (ps $< 0·05$).

2.3. *Discussion*

Younger students' evaluations of the signal words produced a similar ordering to that of the college students in this study and those tested by Wogalter and Silver (1990). One exception was that NOTICE and ATTENTION were reversed in the overall carefulness ratings between the two studies, but neither study showed these two words to differ significantly. Another difference was that the ASU college students rated WARNING significantly higher on connoted carefulness than CAUTION. This result has neither been found in previous empirical research (Dunlap *et al.* 1986, Wogalter and Silver 1990), nor does it correspond with the ratings of the grade-school students in the present research. However, this result does support the difference between WARNING and CAUTION as asserted in standards and guidelines (ANSI 1991, FMC Corporation 1985). Why these college students inferred the difference that other populations did not is unclear, and may be due to various reasons. For example, in growing up in this fairly remote mountainous region of the United States, ASU students may have learned hazard level associations with the words either from exposure to signs under certain conditions or through some sort of education and experience unique to this region.

The results also showed that the younger students generally gave higher carefulness ratings than the older students. The finding that children make more conservative estimates of danger is consistent with previous research (Howarth and Lightburn 1981,

Table 2. Percentages of missing ratings as a function of student grade level.

	Missing rates by grade level				
Word	4th	5th	6th	7th	8th
NOTE	8·9	8·3	2·0	0·0	6·2
REMINDER	10·7	3·3	0·0	1·5	4·7
NEEDED	8·9	3·3	0·0	0·0	0·0
REQUIRED	14·3	8·3	6·0	4·4	4·7
NECESSARY	5·4	5·0	0·0	0·0	4·7
ATTENTION	8·9	8·3	2·0	1·5	3·1
NOTICE	1·8	1·7	0·0	1·5	1·6
PREVENT	10·7	6·7	0·0	1·5	4·7
HALT	28·6	13·3	8·0	2·9	10·9
NO	5·4	0·0	0·0	0·0	0·0
IMPORTANT	3·6	5·0	0·0	0·0	1·6
DON'T	0·0	0·0	0·0	1·5	1·6
CAREFUL	1·8	0·0	0·0	0·0	0·0
PROHIBIT	26·8	15·0	6·0	5·9	10·9
RISKY	1·8	1·7	0·0	1·5	0·0
URGENT	14·3	10·0	2·0	1·5	3·1
ALARM	5·4	1·7	0·0	0·0	0·0
NEVER	1·8	3·3	0·0	0·0	0·0
STOP	0·0	1·7	0·0	0·0	0·0
ALERT	7·1	1·7	0·0	1·5	0·0
HOT	0·0	0·0	0·0	0·0	0·0
VITAL	19·6	20·0	6·0	4·4	10·9
FORBIDDEN	12·5	8·3	0·0	2·9	7·8
CRUCIAL	37·5	16·7	10·0	8·8	9·4
UNSAFE	1·8	1·7	0·0	0·0	0·0
INJURIOUS	19·6	20·0	12·0	5·9	12·5
CAUTION	**7·1**	**0·0**	**0·0**	**1·5**	**0·0**
BEWARE	0·0	0·0	0·0	0·0	3·1
WARNING	**0·0**	**0·0**	**0·0**	**0·0**	**0·0**
SEVERE	23·2	8·3	0·0	4·4	6·2
HAZARD	5·4	1·7	0·0	2·9	1·6
HARMFUL	0·0	0·0	0·0	0·0	0·0
SERIOUS	3·6	1·7	0·0	1·5	0·0
CRITICAL	23·2	13·3	2·0	4·4	4·7
LETHAL	23·2	16·7	10·0	7·4	9·4
DANGER	**0·0**	**0·0**	**0·0**	**0·0**	**0·0**
HAZARDOUS	8·9	5·0	2·0	2·9	1·6
DANGEROUS	1·8	1·7	0·0	0·0	0·0
FATAL	19·6	5·0	2·0	5·9	3·1
TOXIC	10·7	3·3	2·0	2·9	3·1
POISON	0·0	0·0	0·0	0·0	0·0
EXPLOSIVE	3·6	3·3	0·0	1·5	0·0
DEADLY	0·0	1·7	0·0	0·0	0·0

Sheehy and Chapman 1985). This result can be attributed to at least three reasons. First, the younger students may not be able to discriminate between small differences in the levels of hazard, but at the same time, they may recognize that the words indicate a dangerous condition, producing the higher overall carefulness ratings. Second, the younger students' generally inflated scores might represent a response derived from the consequences that could occur if they were not careful when seeing these terms. That

is, children might be punished by protective authority figures (e.g. parents, teachers) if they do not adhere to the behavioural directives of the signal words (e.g. spanking, admonishments). Therefore, not only might these students be affected by the hazard levels connoted by the terms, but also there is a strong likelihood of chastisement and rebuke for not taking appropriate heed to them, which might have instigated the higher carefulness ratings. Third, the difference between the younger and older students might reflect the older group's greater familiarity with the signal words; because of their greater experience and exposure to the words (and to the situations in which they appear), habituation could have produced the lower carefulness ratings.

The results also showed a strong relationship between the understandability ratings of the college students and the missing ratings of the grade-school students. This suggests that both are measuring the same underlying dimension, which is different than carefulness (or strength). The correlations with the other metrics of understandability were smaller, with the exception of the word frequency counts in some instances. As might be expected, missing ratings also decreased with increasing grade level.

In this experiment, the understanding and strength of signal words were evaluated using younger participants. Reliability was found between the results of these participants and two undergraduate student populations (in this study and earlier research). This stability suggests that reasonably accurate predictions of this sort can come from initial tests using college student populations. To further establish the reliability of these findings, the next study examines the connotation and understandability of these words as evaluated by other at-risk populations, the elderly and non-native English speakers.

3. Experiment 2

This experiment is similar to experiment 1 except that elderly US citizens and recent US immigrants who are non-native English speakers evaluated the terms.

3.1. *Method*

3.1.1. *Participants*: A sample of 233 individuals voluntarily participated in this study. Ninety-eight were residents of various retirement homes in Orlando, Florida (Mean age = 74·44, SD = 9·20). The remaining 135 participants were non-native English speakers enrolled in an 'English as a Second Language' class either at the University of Central Florida in Orlando or at the Refugee Resettlement Program in Springfield, Massachusetts (Mean age = 23·75, SD = 5·45). These non-native English speakers varied both in culture (represented by over 30 nationalities and 17 languages) and fluency with the English language.

3.1.2. *Stimuli and procedure*: The stimuli and procedure were identical to that of the grade-school students of experiment 1.

3.2. *Results*

3.2.1. *Analysis of the carefulness ratings*: The carefulness ratings of the same eight terms evaluated in experiment 1 were analysed. These terms are underlined in table 3. Also, the three most common signal words are further designated in bold print in this table. A 2 (participant groups: elderly versus non-native English speakers) $\times$ 8 (signal words: DEADLY, DANGER, WARNING, CAUTION, ATTENTION, CAREFUL, NOTICE, NOTE) ANOVA showed a significant main effect of word, $F(7, 1162) = 61{\cdot}66$, $p < 0{\cdot}0001$ with DEADLY, DANGER, WARNING, CAUTION,

Table 3. Carefulness, means, standard deviations, and percentages of missing ratings for elderly and non-native English speakers.

	Elderly participants			Non-native English speakers		
Word*	Mean	SD	% missing	Mean	SD	% missing
NOTE	4·30	2·45	2·0	4·76	2·37	12·6
NEEDED	4·80	2·30	3·1	5·04	2·29	7·4
REMINDER	4·80	2·07	2·0	4·65	1·93	39·3
NOTICE	5·00	2·20	2·0	4·99	2·07	17·0
NECESSARY	5·06	2·20	2·0	5·44	2·15	13·3
REQUIRED	5·17	2·07	2·0	5·01	2·13	29·6
CAREFUL	5·24	2·19	1·0	5·97	2·02	15·6
PREVENT	5·33	2·11	0·0	5·35	2·04	28·1
ATTENTION	5·43	2·22	1·0	6·05	1·88	6·7
IMPORTANT	5·59	2·16	0·0	6·22	2·03	3·7
NO	5·82	2·31	5·1	5·84	2·55	4·4
ALERT	5·91	2·02	1·0	5·57	2·09	37·8
CAUTION	**5·92**	**2·03**	**1·0**	**5·96**	**2·20**	**17·0**
DON'T	5·94	2·03	0·0	5·71	2·49	5·9
RISKY	5·96	1·90	1·0	5·63	2·25	31·1
INJURIOUS	5·99	2·01	1·0	5·91	1·94	37·0
BEWARE	6·08	2·03	1·0	5·54	1·92	40·0
ALARM	6·09	2·11	0·0	5·86	2·14	12·6
PROHIBIT	6·14	2·04	0·0	5·68	2·26	28·1
NEVER	6·28	2·06	1·0	5·58	2·89	6·7
VITAL	6·31	2·00	1·0	5·14	2·47	35·6
HALT	6·35	2·13	1·0	5·31	2·23	49·6
SERIOUS	6·43	1·70	1·0	5·91	2·02	8·1
CRUCIAL	6·44	1·87	0·0	5·18	2·32	4·0
WARNING	**6·49**	**1·64**	**0·0**	**6·44**	**1·70**	**17·0**
UNSAFE	6·55	1·83	1·0	5·60	2·09	23·0
FORBIDDEN	6·55	2·05	0·0	5·89	2·06	46·7
HOT	6·61	1·96	0·0	5·30	2·51	3·0
HARMFUL	6·68	1·72	2·0	5·69	2·15	37·0
HAZARD	6·72	1·82	0·0	5·51	2·08	44·4
URGENT	6·82	1·73	3·1	6·17	2·00	34·8
SEVERE	6·84	1·55	1·0	5·20	2·03	41·5
CRITICAL	6·89	1·62	2·0	4·96	2·43	22·2
STOP	6·96	1·70	0·0	6·45	2·17	3·0
DANGER	**7·00**	**1·50**	**0·0**	**6·99**	**1·50**	**10·4**
DANGEROUS	7·04	1·38	0·0	7·20	1·30	8·9
TOXIC	7·08	1·79	2·0	6·57	2·09	26·7
HAZARDOUS	7·17	1·51	0·0	5·03	2·26	52·6
FATAL	7·57	1·40	0·0	6·59	2·14	27·4
POISON	7·57	1·32	0·0	7·17	1·56	14·1
LETHAL	7·57	1·38	4·1	6·03	2·40	45·9
EXPLOSIVE	7·62	0·82	1·0	6·75	2·02	38·5
DEADLY	7·69	0·96	0·0	6·51	2·11	20·7
Mean	6·27			5·78		

* Analyses of the bold and underlined words are described in the text. Words are ordered according to the mean carefulness ratings of the elderly participants.

ATTENTION, CAREFUL, NOTICE and NOTE rated from greatest to least on intended carefulness. Subsequent Newman–Keuls range tests showed significant differences for all pairwise comparisons ($ps < 0{\cdot}05$) except between DEADLY and DANGER and between ATTENTION and CAREFUL. DANGER was rated higher on perceived carefulness than either WARNING or CAUTION. In addition, WARNING was rated significantly higher on carefulness than CAUTION. The ANOVA showed no significant main effect of participant group, but signal word and participant group interacted, $F(7, 1162) = 4{\cdot}95$, $p < 0{\cdot}0001$. Using the correction of Satterthwaite (1946) for testing simple effects in mixed designs, comparisons showed the non-native English speakers rated the term CAREFUL significantly higher than the elderly did, $F(1, 652) = 5{\cdot}49$, $p < 0{\cdot}02$. However, the elderly rated DEADLY significantly higher than the non-native English speakers did, $F(1,652) = 10{\cdot}68$, $p < 0{\cdot}002$. Indeed, the non-native English speakers rated DANGER slightly but not significantly higher than DEADLY. The other terms did not significantly differ between participant groups ($ps > 0{\cdot}05$).

Although there was a significant interaction between signal word and participant group with regard to mean carefulness ratings, the rank order of the terms was fairly consistent, τ_{av} (Kendall's Tau) $= 0{\cdot}86$, $p < 0{\cdot}0015$. In addition, the rank order of the terms by the elderly and the non-native English speakers was consistent with those of the grade-school and college students in experiment 1, W (Kendall's coefficient of concordance) $= 0{\cdot}93$, $p < 0{\cdot}001$. Moreover, the overall carefulness ratings for the elderly were highly correlated with those of the grade-school children and the undergraduates in experiment 1 and in Wogalter and Silver (1990). These correlations ranged from 0·82 to 0·96. However, the correlations for the non-native English speakers were lower in relation to these groups, (r ranged from 0·57 to 0·63). Finally, there was a fairly strong correlation between the mean ratings of the elderly and non-native English speakers, $r = 0{\cdot}60$, $p < 0{\cdot}0001$.

3.2.2. *Analysis of understandability*: As described in experiment 1, the number of missing values can be used as a measure of the participants' understanding of the terms. Examination of table 3 shows that the non-native English speakers left more word ratings blank than the elderly did. Words like HOT and STOP were left blank by less than 3% of the non-native English speakers, whereas words like HALT and HAZARDOUS were left blank by nearly 50% of these participants.

As the number of missing values for each word increased among the non-native English speakers, the college students (in Experiment 1) rated these words lower in mean understandability ($r = -0{\cdot}71$, $p < 0{\cdot}01$). Moreover, words most frequently left blank by the non-native English speakers were also terms used less frequently in the English language ($r = -0{\cdot}62$ and $-0{\cdot}46$ for the Thorndike–Lorge (1944) and the Francis–Kucera (1982) counts, respectively, $p < 0{\cdot}01$). Word length (number of letters) was positively correlated to the number of missing ratings, $r = 0{\cdot}28$, $p < 0{\cdot}05$). Variability (as measured by the standard deviations in the carefulness ratings) did not yield significant relationships with the other understandability metrics, except for word length ($r = -0{\cdot}44$, $p < 0{\cdot}01$).

3.3. *Discussion*

In general, the words connoting greater carefulness to the elderly were the same as those for the non-native English speakers. The pattern was also similar to that of the grade-school children and college students in experiment 1 and Wogalter and Silver (1990). One exception to the earlier results is that ATTENTION has higher carefulness

ratings than CAREFUL among the elderly and non-native English speakers. This pattern was reversed for all other populations. The other exception was that the non-native English speakers connoted greater care to DANGER than DEADLY, which was reversed for all other populations.

The higher carefulness ratings for DANGER as compared to WARNING and CAUTION corroborates the findings of experiment 1 and several earlier studies (Bresnahan and Bryk 1975, Dunlap *et al.* 1986, Wogalter and Silver 1990). Although the significantly higher carefulness ratings for WARNING as compared to CAUTION has little support in the research literature, this finding concurs with the ASU college student ratings in experiment 1. Thus, the present results provide some support for a hazard level difference between WARNING and CAUTION as denoted in the standards and guidelines (ANSI 1991, FMC Corporation 1985). These results notwithstanding, explanations for the equivocal difference between WARNING and CAUTION across research studies are not easily forthcoming. One possibility is that the elderly and non-native English speakers perceive themselves to be more vulnerable than most college or grade-schools students. As a consequence, they make more finer gradations among various levels of hazard. For example, most elderly persons have considerable exposure to warnings across their lifetimes (e.g. with respect to pharmaceuticals and medical devices), and as a result may have formed knowledge structure delineating a difference between WARNING and CAUTION. However, this explanation does not fully account for the non-native English speakers' differentiation of these two terms. Possibly, in their limited exposure to the English language, the non-native English speakers received training on the intended meanings of the terms (perhaps through formal instruction, or paying close attention to the gradations of English word meanings or to verbiage on products manufactured by English-speaking countries). There are other possible explanations, but perhaps the most rational way to deal with the discrepant findings for WARNING and CAUTION at the present time is to withhold judgement until a consistent pattern is replicated in future investigations. This notwithstanding, another perspective to this indeterminate finding is worth noting: Even if these two signal words are found to be statistically different in subsequent research, some practical judgements should also be used, and these include consideration of the adequacy of the difference if found to be reliable and its real-world importance. If the difference is not strong, then the implied levels will probably not be differentiated by substantial numbers of people, and thus its utility to convey two separate hazard levels should be questioned. Other terms can be selected to better delineate different degrees of hazard than these two terms.

The results also showed a strong relationship between the understandability ratings of college students and the percentage of missing ratings of non-native English speakers. The validity of the missing-values scores as a measure of understandability is supported by the significant relations with the words' frequency of occurrence in English and letter length. These results suggest that a single underlying dimension is being captured by these measures, which the authors suggest is understandability.

4. General discussion

The two experiments showed that the three special population of grade-school children, the elderly, and non-native English speakers gave signal word ratings that were, in general, similar to each other and to the most usually tested group, college students. Although the pattern was generally the same, the elderly and grade-school children (fourth- and fifth-graders) provided higher carefulness ratings than college students did.

Table 4. Mean carefulness ratings of signal words known by 95% or more of the fourth and fifth graders and by 80% or more of the non-native English speakers. Also shown are college student and elderly participant ratings.

	Study 1		Study 2	
Word*	4th & 5th Graders	ASU college Students	Elderly	Non-native English speakers
NOTICE	5·39	4·01	5·00	3·64
CAREFUL	5·86	4·76	5·23	5·88
ALARM	6·16	5·01	6·09	4·87
IMPORTANT	5·95	5·06	5·59	5·64
CAUTION	6·64	5·22	5·91	4·75
DON'T	6·12	5·24	5·93	4·54
NO	5·63	5·60	5·81	4·68
SERIOUS	6·90	5·73	6·43	5·52
NEVER	6·09	5·93	6·27	5·34
WARNING	6·52	6·13	6·49	5·58
HOT	6·00	6·21	6·61	4·40
STOP	6·11	6·43	6·95	6·55
DANGER	7·12	6·49	7·00	7·63
DANGEROUS	7·18	6·64	7·04	7·66
POISON	7·49	7·00	7·57	7·93

* Selection of terms were based on missing-value indicators of understandability. Words are ordered according to the mean carefulness ratings of the college students.

One other purpose of this research was to construct a list of words that might be used for special populations. Several criteria could be used to eliminate less appropriate terms. As mentioned earlier, the list of 43 words used in the present experiments were selected from a larger list of 84 words investigated by Wogalter and Silver (1990) based on word length of less than 10 letters and understandability ratings of college students in a previous study. In order to reduce the list further, two other criteria were used: (a) the words had to be known by at least 95% of the youngest grade-school students (fourth- and fifth-grade students combined), and (b) the words had to be known by 80% or more of the non-native English respondents. The basis of these criteria are the missing ratings. Using the first criterion, 20 of the words were eliminated, leaving 23 words known by 95% or more of the fourth- and fifth-grade students. Using the second criterion of 80% comprehension by the non-native English speakers, eight additional words were eliminated. The resulting list of 15 terms is shown in table 4 together with the mean carefulness ratings of the ASU college students, fourth- and fifth-graders combined, the elderly, and the non-native English speakers.

It should be noted that the list of terms in table 4 is not the only list that could be constructed. Given other criteria and different measures of understandability, different terms should be selected. Nevertheless, it should be noted that the terms are on the current list are among the most frequently used words in the English language, and they appear to be suitable for individuals with low-level English reading ability (Fletcher and Abood 1988, Johnson *et al.* 1983).

One other reason for constructing the list is to show that with a fairly simple set of criteria, a group of terms can be derived that are interpretable by a wide range of individuals in the population. Lists such as this one would be useful to individuals designing warnings. Also the list of terms and the descriptive statistics presented in this

report are useful in another respect. The tables can be used as a partial basis for selecting alternative terms that convey various hazard levels including substitutes for other terms. A warning designer should select terms that are the most understandable to the target population(s) with significant (statistical as well as important) differences along the hazard dimension.

Signal words should be chosen to be congruent with the particular hazard situation. Certain hazard words might be inadequate because they are used in a variety of non-hazard contexts. Extensive use of the term WARNING in common messages 'WARNING: Discount coupon will expire at year's end' or 'WARNING: Batteries not included' may dilute its arousal strength. That is, although the guidelines suggest that WARNING be reserved for those situations that could result in severe personal injury or death (FMC Corporation 1985), this particular term is being widely used in a variety of applications that are incongruent with the guidelines. Hence, given its fairly loose language usage, WARNING might not effectively serve the purpose of signalling dangerous conditions (despite the recommendations in warning design standards and guidelines). Words like REMINDER or NOTE might be more appropriate in two of the examples given above (expiration date and battery), whereas terms such as HAZARD or UNSAFE might substitute for WARNING in applications with potentially dangerous consequences.

How do you properly assign signal words to labels and signs associated with agents (products, equipment, and environments) of varied hazard levels? One possible way is to make a best-fit determination by comparing the guidelines' and standards' definitions associated with the three main signal words with specific reference to the type and magnitude of injury that the product or environment may cause. Another potential technique is to have a group of unbiased experts (knowledgeable about all the associated potential hazards and the consequences) rate the agents on degrees of hazard (Silver and Wogalter 1991b) or consequence (Harris and Wiklund 1989), and then match the ratings of the words to the agents. These ratings are likely to be affected by context, so it is important that assignments must be considered in relation to the particular terms and agents involved (Young *et al.* 1990).

Another approach would be to examine the written materials used in teaching English to children and non-native speakers to determine whether any of these terms are already being taught and to select signal words based in part on this early-training criterion. However, if examination of the teaching materials reveals that very few (or no) safety-related words are being taught, then one potential solution would be to develop a list of the most important terms related to safety communications and incorporate them into early English language training curricula to ensure that the terms' meanings are learned in a timely fashion (Westaway and Apolloni 1978). Moreover, pictorials and icons might be useful in assisting hazard communication when the verbal information cannot be read (Leonard and Karnes 1993, Mayer and Laux 1989, Wolff and Wogalter 1993, Young and Wogalter 1990), although the benefits of pictorials depend on the quality of depiction and the population tested (Dewar 1994, Frascara and Yau 1986, Olmstead 1994).

A comment regarding the generalizability of the present results should be mentioned. In these experiments, the words were shown to participants in list form without a real-world context (for example, on signs and product labels). Therefore, it is not known whether the current results would generalize to more externally valid situations such as in context with other warning information and in appropriate environments (for example, on products and on signs purporting to show hazardous

situations). This kind of research (Wogalter *et al.* 1994) is needed to verify whether signal words in ecologically valid contexts effectively communicate appropriate hazard levels to special populations.

Lastly, most research on signal words has focused exclusively on hazard level connotation. However, warning design guidelines (Westinghouse 1981) describe another purpose of these terms, and that is to capture people's attention to warnings. This purpose has not received much empirical study except for recent research by Laughery *et al.* (1993) and Young (1991). Young (1991) found that alcoholic-beverage warnings captured people's attention faster when a signal word on labels was present compared to its absence. Future studies using reaction time and eye movement measures (Laughery and Young 1991) will provide more information on the attention-getting aspects of signal words and thereby provide a better basis upon which to determine their role in warnings.

Acknowledgements

The authors would like to thank Amy S. Barlow, Dana S. Gammella, Jill R. Kohake, Carol Montanari and Nicole Solberg for their help in the data collection phases of this research. Portions of this article were presented at the 35th and 37th Annual Meetings of the Human Factors and Ergonomics Society (Silver and Wogalter 1991a, Silver *et al.* 1993).

References

ANSI 1991, American National Standard for Product Safety Signs and Labels: Z535.4 (National Electrical Manufacturers Association, New York).

BITUMINOUS CASUALTY CORP. *v.* BLACK & DECKER MFG. CO. 1974, 518 S.W. 2d 868 (Texas).

BRESHAHAN, T. F. and BRYK, J. 1975, The hazard association values of accident-prevention signs, *Professional Safety*, **January**, 17–25.

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

COHEN, J. 1969, *Statistical Power Analysis for the Behavioural Sciences* (Academic Press: New York).

DEWAR, R. 1994, Design and evaluation of graphic symbols, *Proceedings of Public Graphics: Visual Information for Everyday Use* (Stichting Public Graphics Research, Department of Psychonomics, Utrecht, The Netherlands), pp. 24.1–24.18.

DUNLAP, G. L., GRANDA, R. E. and KUSTAS, M. S. 1986, Observer perceptions of implied hazard: Safety signal words and color words. Research Report No. Tr 00.3428 (IBM, Poughkeepsie).

FLETCHER, D. and ABOOD, D. 1988, An analysis of the readability of product warning labels: implications for curriculum development for persons with moderate and severe mental retardation, *Education and Training in Mental Retardation*, **23**, 224–227.

FMC CORPORATION 1985, *Product Safety Sign and Label System* (FMC Corp., Santa Clara).

FRANCIS, W. N. and KUCERA, M. 1982, *Frequency Analysis of English Usage* (Houghton Mifflin, Boston).

FRASCARA, J. and YAU, T. 1986, Evaluation of safety symbols, appropriateness ranking tests and comprehension recognition tests. Report submitted to Research Contribution Advisory Committee, Standards Council of Canada (Department of Art and Design, University of Alberta, Edmonton.

HARRIS, J. E. and WIKLUND, M. E. 1989, Consumer acceptance of threatening warnings in the residential environment, *Proceedings of the Human Factors Society 33rd Annual Meeting* (Human Factors Society, Santa Monica), 989–993.

HOWARTH, C. I. and LIGHTBURN, A. 1981, A strategic approach to child pedestrian safety, in H. C. Foot, A. J. Chapman and F. M. Wade (eds) *Road Safety: Research and Practice* (Praeger, Eastbourne), 127–134.

JOHNSON, D. D., MOE, A. J. and BAUMANN, J. F., 1983, *The Ginn Word Book for Teachers* (Ginn, Lexington).

LAUGHERY, K. R. and YOUNG, S. L. 1991, An eye scan analysis of accessing product warning information, *Proceedings of the Human Factors Society 35th Annual Meeting* (Human Factors Society, Santa Monica), 585–589.

LAUGHERY, K. R., YOUNG, S. L., VAUBEL, K. P. and BRELSFORD, J. W. 1993, The noticeability of warnings on alcoholic beverages, *Journal of Public Policy & Marketing*, **12,** 38–56.

LEONARD, S. D., HILL, G. W. and KARNES, E. W. 1989, How does the population interpret warning signals? *Proceedings of the Human Factors Society 33rd Annual Meeting* (Human Factors Society, Santa Monica), 550–554.

LEONARD, S. D. and KARNES, E. W. 1993, Development of warnings of resulting from forensic activity, *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (Human Factors and Ergonomics Society, Santa Monica), 501–505.

LEONARD, S. D., KARNES, E. W. and SCHNEIDER, T. 1988, Scale values for warning symbols and words, in F. Aghazadeh (ed.), *Trends in Ergonomics/Human Factors V* (Elsevier, Amsterdam), 669–674.

LEONARD, S. D., MATTHEWS, D. and KARNES, E. W. 1986, How does the population interpret warning signals? *Proceedings of the Human Factors Society 30th Annual Meeting* (Human Factors Society, Santa Monica), 116–120.

LEVY, K. L. 1976, A multiple range procedure for independent correlations, *Educational and Psychological Measurement*, **36,** 27–31.

LOPEZ *v.* ARO 1979, 584 S.W. 2d. 333 (Texas).

MAYER, D. L., and LAUX, L. F. 1989, Recognizability and effectiveness of warning symbols and pictorials, *Proceedings of the Human Factors Society 33rd Annual Meeting* (Human Factors Society, Santa Monica), 984–988.

OLMSTEAD, W. T. 1994, Cultural differences in comprehension of public information symbols for health care facilities, *Proceedings of Public Graphics: Visual Information for Everyday Use* (Stichting Public Graphics Research, Department of Psychonomics, Utrecht, The Netherlands), pp. 25.1–25.11.

OWSLEY, C., BALL, K., SLOANE, M. E., ROENKER, D. L. and BRUNI, J. R. 1991, Visual/cognitive correlates of vehicle accidents in older drivers, *Psychology and Aging*, **6,** 403–415.

RAMIREZ *v.* PLOUGH INC. 1992, Calif. Ct. App. No. F01950 (5th District, 1992).

RAO, C. R. 1970, *Advanced Statistical Methods in Biometric Research* (Hafner, New York).

SATTERTHWAITE, F. E. 1946, An approximate distribution of estimates of variance components, *Biometrics Bulletin*, **2,** 110–114.

SCHLAG, B. 1993, Elderly drivers in Germany: fitness and driving behaviour, *Accident Analysis and Prevention*, **25,** 47–55.

SHEEHY, N. P. and CHAPMAN, A. J. 1985, Adults' and children's perceptions of hazard in familiar environments, In T. Garling and J. Valsiner (eds), *Children within Environments: Towards a Psychology of Accident Prevention* (Plenum Press, New York).

SILVER, N. C. and BURKEY, R. T. 1991, A FORTRAN 77 program for testing the differences among independent correlations, *Educational and Psychological Measurement*, **51,** 641–644.

SILVER, N. C., GAMMELLA, D. S., BARLOW, A. S. and WOGALTER, M. S. 1993, Connoted strength of signal words by elderly and non-native English speakers, *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (Human Factors and Ergonomics Society, Santa Monica), 516–519.

SILVER, N. C. and WOGALTER, M. S. 1991a, Strength and understanding of signal words by elementary and middle school students, *Proceedings of the Human Factors Society 35th Annual Meeting* (Human Factors Society, Santa Monica), 590–594.

SILVER, N. C. and WOGALTER, M. S. 1991b, Pest-control products: Hazard perception, product type, and label characteristics, *Proceedings of Interface '91: The 7th Symposium on Human Factors and Industrial Design in Consumer Products* (Human Factors Society, Santa Monica), 106–110.

TAJIMA, N., ASAO, K., HILL, G. W. and LEONARD, S. D. 1991, Comparison of meanings of warnings in Japan and the USA, in W. Karwowski and J. W. Yates (eds) *Advances in Industrial Ergonomics and Safety III* (Taylor & Francis, London), 739–741.

THORNDIKE, E. L. and LORGE, I. 1944, *The Teacher's Word Book of 30,000 Words* (Teachers College Press, New York).

URSIC, M. 1984, The impact of safety warnings on perception and memory, *Human Factors*, **26,** 677–682.

U.S. CONSUMER PRODUCT SAFETY COMMISSION 1980, *Preliminary NEISS Estimates of National Injuries* (US CPSC, Washington, DC).

WESTAWAY, A. and APOLLONI, T. 1978, *Becoming Independent: A Living Skills System* (Edmar Associates, Bellevue).

WESTINGHOUSE 1981, *Westinghouse Product Safety label Handbook* (Westinghouse Printing Division, Trafford).

WOGALTER, M. S., GODFREY, S. S., FONTENELLE, G. A., DESAULNIERS, D. R., ROTHSTEIN, P. R. and LAUGHERY, K. R. 1987, Effectiveness of warnings, *Human Factors*, **29,** 599–612.

WOGALTER, M. S. JARRARD, S. W. and SIMPSON, S. N. 1992, Effects of warning signal words on consumer-product hazard perceptions, *Proceedings of the Human Factors Society 36th Annual Meeting* (Human Factors Society, Santa Monica), 935–939.

WOGALTER, M. S. JARRARD, S. W. and SIMPSON, S. N. 1994, Influence of signal words on perceived hazard levels of consumer products. *Human Factors*, **36,** 547–556.

WOGALTER, M. S. and SILVER, N. C. 1990, Arousal strength of signal words, *Forensic Reports*, **3,** 407–420.

WOLFF, J. S. and WOGALTER, M. S. 1993, Test and development of pharmaceutical pictorials, *Proceedings of Interface 93* (Human Factors and Ergonomics Society, Santa Monica), 187–192.

YOUNG, S. L. 1991, Increasing the noticeability of warnings: effects of pictorial, colour, signal icon, and border, *Proceedings of the Human Factors Society 35th Annual Meeting* (Human Factors Society, Santa Monica), 580–584.

YOUNG, S. L., BRELSFORD, J. W. and WOGALTER, M. S. 1990, Judgments of hazard, risk, and danger: Do they differ? *Proceedings of the Human Factors Society 34th Annual Meeting* (Human Factors Society, Santa Monica), 503–507.

YOUNG, S. L. and WOGALTER, M. S. 1990, Comprehension and memory of instruction manual warnings: conspicuous print and pictorial icons, *Human Factors*, **32,** 637–649.