



Effects of emphasis terminology in warning instructions on compliance intent and understandability

Soyun Kim, * Michael S. Wogalter¹

Department of Psychology, North Carolina State University, 640 Poe Hall, CB 7650, 2310 Stinson Drive, Raleigh, NC 27695-7650, USA

ARTICLE INFO

Article history:

Received 20 October 2014
Received in revised form 27 May 2015
Accepted 23 July 2015
Available online 11 August 2015

Keywords:

Warnings
Vocabulary
Perceived hazard
Terminology
Compliance intent

ABSTRACT

Introduction: The aim of this study was to examine whether differing terminology in warning directives can influence compliance intentions and understandability. Despite its important role for warning effectiveness, warning instructions has not received much attention in warning research. Emphasis terms that can be used in warning directives were investigated. **Method:** Three experiments were conducted. In Experiment 1, participants rated a set of 12 warning directive statements consisted of one basic warning directive, which served as the control and the other 11 one- or two-word emphasis phrases that added to a basic directive. In Experiment 2, participants rated 37 emphasisers on compliance intent. In Experiment 3, participants rated the same emphasisers on understandability. **Results:** The first 2 experiments showed substantial differences in compliance intentions depending on the emphasisers used. For example, some terms and phrases (e.g., “urgent”) produced high compliance intent whereas others showed lower compliance intent (e.g., “recommended”). In Experiment 3, some terms were rated as understandable (e.g., “important”), whereas others were rated as somewhat understandable (e.g., “compulsory”). **Conclusion:** The addition of emphasis terms to the warning directives influenced people’s compliance intent and understandability. In addition, significant correlations were found among compliance intent, understandability, and measures of variability. **Practical application:** . The findings from this research could aid warning designers in selecting understandable wording that gives rise to different levels of compliance intentions.

© 2015 National Safety Council and Elsevier Ltd. All rights reserved.

1. Introduction

Warnings are intended to communicate safety information to users of products and environments (Laughery & Wogalter, 1997, 2006). Most research (e.g., Laughery, Wogalter, & Young, 1994; Wogalter et al., 1987) as well as guidelines and standards (e.g., American National Standards Institute [ANSI], 2011) suggest that warnings should, in many cases, include certain components. For example, the ANSI (2011) Z535 warning standard for product labels and signs recommends that warnings contain four main textual components: (a) a panel including signal word (e.g., DANGER, WARNING, and CAUTION); (b) a statement giving hazard information; (c) a consequences statement telling what may happen if the hazard is not avoided; and (d) instructions statement on how to avoid the hazard. This standard also suggests that warnings contain graphics and color. Fig. 1 contains an example of a warning that illustrates these components. This warning also includes red color and an

alert symbol (triangle enclosing an exclamation point) graphic in the signal word panel.

Research of warnings has verified some, but not all, of the components specified in the ANSI (2011) Z535 standard as having some benefit. The components that have received the most in-depth examination have been symbols (graphics), colors, signal words, and consequences. These components are briefly described below.

With regard to symbols, most studies have examined factors affecting comprehension, and to a lesser extent their attention-getting properties (e.g., see Dewar, 1999; Wogalter, Silver, Leonard, & Zaikina, 2006). To a large extent, symbols that directly represent or depict the hazard-related concept that are more familiar (trained or frequently used) are more likely understood. Symbols are optional in the ANSI warnings (except the alert symbol shown in Fig. 1 is required for hazards that could cause personal injury), but they are recommended. If symbols are used alone without text, they should undergo comprehension testing with acceptable symbols receiving at least 85% correct using a sample of 50 persons with no more than 5% critical confusion errors (i.e., opposite and very wrong answers).

The ANSI Z535.1 standard also specifies certain colors in relation to particular signal words, red with DANGER, orange with WARNING, and yellow with CAUTION. The levels are defined based on probability and severity of the hazard. Parts of these assignments are in accordance

* Corresponding author at: 84 Heukseok-ro, Dongjak-gu, Seoul 156-756, Republic of Korea. Tel.: +82 10 2852 6313.

E-mail addresses: soyun0714@gmail.com (S. Kim), WogalterM@gmail.com (M.S. Wogalter).

¹ Tel.: +1 954 744 4213.



Fig. 1. An example of ANSI Z535.4 (2011) type product warning.

with research findings but other aspects show discrepancies in research measuring people's beliefs or connotation of hazard for these colors. The color Red is almost always the highest rated in terms of hazard level in accord with the standards (e.g., Chapanis, 1994). However, most research does usually find a substantial difference between orange and yellow although both are lower than red (see, e.g., Chapanis, 1994; Smith-Jackson & Wogalter, 2000; Wogalter, Kalsher, Frederick, Magurno, & Brewster, 1998).

Another component of ANSI Z535 is signal words, which have received extensive evaluation in research, particularly concerning their perceived hazard connotation. As mentioned above, ANSI Z535 recommends three signal words associated with hazards: DANGER, WARNING, and CAUTION. Like their associated colors, they are defined based on probability and severity of the hazard. Research has shown that people's understanding of these terms is not entirely in accord with the ANSI standards' definitions of what the terms are supposed to mean. While most research shows that DANGER is perceived to have the highest level of hazard or urgency compared to the other two terms, the other two terms are often not perceived as being different and there is some research showing a reversal of order (e.g., Chapanis, 1994; Dunlap, Granda, & Kustas, 1986; Hellier & Edworthy, 2006; Leonard, Karnes, & Schneider, 1988; Wogalter, Laughery, & Mayhorn, 2012).

Several researchers (Edworthy & Adams, 1996; Hellier & Edworthy, 2006; Wogalter & Silver, 1995) have suggested that the choice of signal words should be congruent with the particular hazard situation. In other words, the terms selected to should be associated or calibrated with the likelihood and severity of injury potential involved.

Besides signal words, very little research has been done concerning terminology in warnings. What has been done has mainly been concerned explicitness of the consequences information. Consequences statements tell what could happen, the potential outcome, if the warning directives or instructions are not followed. While advocated in the ANSI Z535.4 product warning standard as well as research (e.g., Wogalter et al., 1987), consequences information are not always included in warnings. The allowable reason for leaving out the consequences statement is when the outcome is well known and obvious or is easily inferred from rest of the warning. However, there is seldom adequate justification (e.g., data) to support leaving out the consequences statement. While some research has investigated the presence versus absence of consequence information (Wogalter et al., 1987; Young & Wogalter, 1998), most of the other investigation on varying the wording of warnings has been mainly on the extent of explicitness of the consequences information. Research shows that giving general, nonspecific consequences such as "may cause health problems" (as is found in the warning on alcoholic beverage containers in the United States) is less informative and less motivating for compliance than statements that give more specific consequences (e.g., "may cause acute liver failure;" Laughery & Stanush, 1989; Laughery & Paige-Smith, 2006).

Thus, although research has been done on some of the components of warnings (symbols, color, signal words, and explicitness of consequences), other components have not. One important area that has not yet been evaluated is the effect of varying the level of emphasis in the wording used in the instructions (directives) component of warnings. This is examined in the present research.

Although there has not been research on varying the emphasis of the warning directive statements, there has been research on the cost of compliance in performing the instructions (Dingus, Hathaway, & Hunn, 1991; Wogalter, Allison, & McKenna, 1989; Wogalter et al., 1987). People are less likely to comply with warning instructions when there is high cost of complying (when compliance requires effort, time, and money) compared to when there is low cost of complying or easy to do. However, other potential aspects of warning instructions, which could influence warning effectiveness, have not received much attention in research. Warning instructions are important because they are directly tied to people's safety as they are used to inform people on what they should do to avoid getting hurt. If people do not realize the importance of performing the instructed behavior, they may be less willing to comply with them, potentially doing or not doing something potentially resulting in severe injury or property damage (Laughery & Wogalter, 1997). The strength or urgency of the terminology in warning directives might influence compliance intent.

Besides signal words, measuring perceptions for varying terminology used to describe different levels of dimensions on labels has been examined in a few studies. Kreifeldt (1993) has measured the influence of qualifier wording that could be added to instructions to give a better idea of how to modify human behavior. Also, other researchers have measured terms based on their relative meaning along different dimensions on chemical labels and nutrition labels of foods (Kalsher, Wogalter, & Gilbert, 1992; Lehto & House, 1997; Wogalter, Kalsher, & Litynski, 1996).

One way to potentially influence intent to comply is to add motivating language to basic warning directives. Consider a warning instruction statement related to a severe respiration hazard. A basic warning directive might simply state, "Wear XYZ respirator." This "bare" directive could be changed in ways that might influence motivation to comply. One way to potentially accomplish this is to add terminology that gives greater emphasis to the directive such as adding an adjective word "important," such as in the statement "It is *important* to wear XYZ respirator." Also, motivation might further be affected by adding an adverb such as "extremely important." Adding this adverb might further bolster the adjective to provide even greater emphasis to the instruction, as in the statement, "It is *extremely important* to wear XYZ respirator."

Previous warning research has employed several different methodologies to measure warnings effectiveness. Although behavioral compliance is the most important criterion of warning effectiveness, there are relatively few behavioral compliance empirical studies because they are difficult to conduct because of ethical, situational, cost constraints, or limited resources (Kalsher & Williams, 2006; Wogalter et al., 1998). Consequently, many warning researchers have employed measurements of intermediate or pre-behavior stages (Kalsher & Williams, 2006; for C-HIP model, refer to Wogalter, 2006) such as memory tests, knowledge ratings, or perceived hazard (Wogalter et al., 1998). For example, perceived hazard has been found to be strongly related to behavior-related intentions such as intended carefulness and willingness to comply (Wogalter et al., 1998). Some warning research has shown that people's intentions have a strong relation of behavior (Wogalter et al., 1998) and compliance (Kalsher & Williams, 2006). Intention to comply is the main response measure in the present research.

Thus, the present research examines whether addition of certain terms or phrases to warning directives would affect compliance intent. To add some clarity, we have defined certain types of terms for these types of words, albeit somewhat arbitrarily. The general category of terms intended to influence compliance intent are called *emphasizers*.

Some emphasizeers might increase compliance intent, while others might not have any influence or even decrease it. Single-term emphasizeers (adjectives) such as term “important” are called *intensifiers*. Other terms (adverbs) added to intensifiers such as “extremely” are called *qualifiers*. Thus, some manipulations had only intensifiers and others had both a qualifier and an intensifier.

An important consideration for warnings is comprehension (see, e.g., Wogalter, 2006). Besides promoting compliance, another purpose of warnings is to give individuals an adequate appreciation or understanding of the hazard so that people will be able to make informed judgments and convey information on the appropriate compliance behavior (Wogalter, 2006). In the present research, emphasizeers are measured on understandability to aid in selecting potential emphasizeers for use in warning directives. More details and rationale for measuring understandability is presented in Experiment 3.

In the current study, emphasizeers are evaluated individually, which are not embedded within complete and entire warnings. This methodology was chosen based on the assumption that warnings have already attracted people’s attention successfully (due to salience and other factors). This approach comports with the C-HIP model (Communication-Human Information Processing, refer to Wogalter, 2006), which describes warning processing via several stages from warning source to behavior. Early in the process, warnings need to be delivered from a source via some channel. If the warning is adequately conspicuous, it will cause attention to be attracted to it. Extracting meaning from the warning involves separate processes after attention has been attracted. In the current research, we are assessing post-attention processing, namely, comprehension and beliefs about the warning message components.

A final purpose of this research is to provide data on a large list of emphasizeers that could be added to basic warning directives so as to potentially influence compliance intent. Selection would use terms that are highly understandable and that vary in the level of connoted urgency or necessity in complying with the directive. This list could be useful for assisting warning designers in selecting appropriate terms to match the level of hazard being warned about.

2. Experiment 1

In this initial experiment, emphasizeers that could be added to the warning instructions or directives were evaluated. The set of terms were generated based on warnings in varied applications and environments including mining and medical laboratories. Compliance intent was measured.

2.1. Methods

2.1.1. Participants

A total of 21 individuals (12 females, 9 males) participated. Overall average age was 23.1 years ($SD = 5.2$). Most (71%) were undergraduate students from a large southeastern U.S. university ($M = 21.1$ years; $SD = 2.2$), and others (29%) were nonstudent adult volunteers from Raleigh, North Carolina ($M = 28.2$ years; $SD = 7.2$). One-hundred percent were native English speakers.

2.1.2. Materials

Emphasizers consisted of one-word (intensifier only) or two-word (qualifier and intensifier) phrases, such as “important” and “extremely important,” respectively. A control condition was included that had only a basic (unembellished) warning directive that had no emphasizeers. A total of 12 conditions was generated. Each emphasizeer was embedded in warning directive statements in different ways.

To illustrate this, consider the two basic instruction/directive statements without emphasizeers: “Keep all body parts, hair, and loose clothing away from rollers,” and “Wear XYZ Protective Eyewear.” They lack emphasizeers (i.e., control condition). Now consider these same

two statements with emphasizeers as in “We *strongly recommend* that you keep all body parts, hair, and loose clothing away from rollers,” and “*Mandatory*: Wear XYZ Protective Eyewear.” The italicized words in the last sentence were emphasizeers; the first has a qualifier and intensifier and the second has only an intensifier. In some cases, some additional connective words were added along with the emphasizeer.

Participants were shown two warnings consisting of four components that are commonly found in warnings (Wogalter et al., 1987): signal word, hazard information, consequences, and directives. They were as follows: (a) DANGER; Entanglement Hazard; Could Result in Crush, Amputation, Scalping Injury; Keep all body parts, hair, and loose clothing away; and (b) WARNING; Laser; Could Cause Severe Eye Injury; Wear XYZ Protective Eyewear. Note the last components in both warnings are directives (instructions) that give information to people on what to do to avoid hazard. These two example warnings were used to provide the base instruction statement which then was varied according to the emphasizeer term conditions. The base statement only condition was the control.

2.1.3. Procedure

Each participant was given a questionnaire packet that included a consent form, demographics questions (e.g., age, gender), and a set of warnings directives with blanks for participants to mark their ratings. Before rating the warning directives, participants were given the following background and instructional information: warnings are intended to convey information about hazards and motivate people to follow safety instructions and are found on many products, industries, and environment, as for example in industrial conveyor systems and in medical laboratory environments. They were then shown two example 4-component warnings (as described above), and it was pointed out that the last component in each example warning gives instructions or directives on what to do to stay safe. Each example warning was provided along with 12 warning directive statements, which were comprised of the directives with emphasizeers (11 conditions) and the control with no emphasizeer. The list of emphasizeer conditions is shown in Table 1. Participants were asked to rate each of the directive statements listed according to how likely they believe that the warning would be complied with given the wording. A 0- to 8-point rating scale was used with the anchors 0 (“not at all likely to obey”), 2 (“somewhat likely to obey”), 4 (“likely to obey”), 6 (“very likely to obey”), and 8 (“extremely likely to obey”). Each directive statement (experimental and control) had an adjoining blank where participants recorded their rating. Before making ratings, participants were told to review the entire list of words or phrases to familiarize themselves with the range of words that they would be rating.

Two orders of a set of directive statements were administered. One order was a randomized order. The other was the reverse. After completing the ratings, participants were debriefed and dismissed.

Table 1

Mean compliance intent ratings and standard deviations (SDs) for the 12 emphasizeer conditions ordered from high to low.

Emphasizer terms	Mean	SD
Federal law	6.83*	1.5
Mandatory	6.52*	1.1
Extremely important	6.14*	1.4
Required	6.07*	1.3
Absolutely necessary	6.05*	1.6
Strongly recommended	5.02	1.5
Important	4.98	1.4
Strongly suggested	4.91	1.7
Necessary	4.86	1.6
No emphasizeer (control)	4.57	1.8
Recommended	3.98	1.7
Suggested	3.71	1.4

* The terms were significantly different from the control condition ($p < .05$)

2.2. Results

Two main analyses are described. The first involved the entire set of 12 word/phrase conditions. Table 1 shows the means and the standard deviations of the terms arranged in order from high to low compliance intent. The highest rated was “federal law” ($M = 6.83$) and “mandatory” ($M = 6.52$); the lowest rated were “recommended” ($M = 3.98$) and “suggested” ($M = 3.71$). The control condition (no emphasize) was the third lowest-rated ($M = 4.57$). According to the anchors, these extremes ranged from “very likely” to “somewhat likely” to obey.

One-way repeated-measured analysis of variance (ANOVA) indicated a significant effect of emphasize condition, $F(11, 220) = 17.49$, $MSE = 1.21$, $p < .001$. Tukey’s honestly significant difference (HSD) test set at $p = .05$ was found equal to 1.11. This value can be used to compare any two means to determine if the difference between them is significant at $p < .05$. For example, the addition of “federal law,” “mandatory,” “extremely important,” “required,” and “absolutely necessary” to the basic directive statement produced significantly higher compliance intent ratings than the control condition with the basic directive only. These conditions are indicated by asterisks in the Table 1. The highest-rated three were rated higher than “strongly recommended,” “important,” and the remaining conditions. “Recommended” and “suggested” were rated lower than the control (basic directive) but not significantly.

A second analysis examined whether the components in some of the emphasize in one-word or two-word phrases (intensifier alone vs. intensifier plus qualifier) combine interact or show a linear pattern. From the initial set of 12 emphasize, 4 of them could be assembled to conduct a two factor repeated-measures analysis of variance involving presence versus absence of qualifier (strongly vs. none) with the two intensifiers “recommended” and “suggested.” Table 2 shows these conditions. A 2 (intensifier: “recommended” and “suggested”) \times 2 (qualifier: none and “strongly”) repeated-measures ANOVA showed a significant main effect of qualifier, $F(1, 20) = 30.57$, $MSE = 26.30$, $p < .001$. Higher compliance intent ratings were produced when the qualifier “strongly” was present ($M = 4.97$) than when it was absent ($M = 3.85$). There was no main effect of intensifier nor an interaction effect, $ps > .05$.

2.3. Discussion

The results showed that some emphasize significantly increased compliance intent (e.g., “mandatory,” “extremely important”) compared to directives with no emphasize (control condition). A few of the emphasize showed lower compliance intent (e.g., “suggested” and “recommended”) although these were not significantly lower than the control condition. The highest-rated emphasize was “federal law.” The results are consistent with earlier warnings research showing that adding specific, reputable, and expert sources (e.g., U.S. Government and Federal Government) to warnings raises the credibility and enhances compliance intent compared to no source (Wogalter, Kalsher, & Rashid, 1997, 1999).

Some of the included emphasize were not ones that one would expect to see in a serious warning, such as “suggested.” Nevertheless, it is not uncommon to find some of the lower rated terms such as “recommended” in warnings for consumer products. Clearly, these low-rated terms would be inappropriate if compliance is necessary to avoid personal injury consequences.

Table 2
Mean compliance intent ratings and SDs (in parentheses) of four conditions.

	No qualifier	Mean (SD)	Qualifier	Mean (SD)
Intensifiers	Recommended	3.98 (1.8)	Strongly recommended	5.02 (1.5)
	Suggested	3.71 (1.4)	Strongly suggested	4.91 (1.7)
Mean		3.85 (1.6)		4.97 (1.5)

The ratings for two of the low-rated emphasize “recommended” and “suggested” were raised when the qualifier “strongly” is added to them. This suggests that the individual intensifiers could be changed and in this case raised by the addition of a modifying adverb. This fact indicates that there is the potential to “fine tune” the level of compliance intent needed. This is further examined in Experiment 2 with a larger set of emphasize terms. As an initial study, only 29 people participated in Experiment 1. In Experiment 2, a larger number of people participated.

3. Experiment 2

Experiment 2 used a larger number of one- (intensifier only) and two- (intensifier and qualifier) emphasize terms to determine their effects on compliance intent. There were also more participants than in Experiment 1.

3.1. Methods

3.1.1. Participants

A total of 138 individuals (79 females, 59 males) participated. Overall average age was 28.4 years ($SD = 11.1$); 28% ($n = 38$) were non-native English speakers. Two population samples were collected: 40.6% were undergraduate students from a large southeastern U.S. university ($M = 21.5$ years; $SD = 3.8$), and 59.4% were nonstudent adult volunteers from central North Carolina ($M = 33$ years; $SD = 12.0$).

3.1.2. Materials and procedure

The materials and procedure was identical to the Experiment 1 except there was a larger set of emphasize, more people evaluated them, and the instructions and example warning statement provided a more general context. The enlarged set of emphasize was constructed using the set in Experiment 1 together with a thesaurus and dictionary. A list of 37 emphasize conditions including one- (intensifier only) and two- (intensifier and qualifier) emphasize terms were constructed. Some of the two-word emphasize were purposely included to examine the effect found in Experiment 1 of the effect of presence versus absence of a qualifier term (as, e.g., “necessary” versus “extremely necessary”). The total set of emphasize terms are listed in Table 3.

Participants initially completed a consent form, followed by several questions asking for demographic information. Before the main experimental task, participants were told that product warning labels make use of different types of words to convince or persuade consumers to follow the safety instructions, and that, for example, the term “important” might be included in an instructions or directives statement in warning sign or label. They were told to rate the added emphasis of the term “important” compared to a basic directive which was simply “Obey the warning.” To illustrate this, using the term “important,” they were provided two example instructions or directives statements: “It is important that you obey the warning” and “Important – Please obey the warning.” The use of the generic warning statement was intended to provide a general as opposed to any specific context (particular products or environments) under which the emphasize terms would be evaluated. A set of 37 emphasize were listed and each emphasize had an adjoining blank where participants recorded ratings. Before making ratings, participants were told to review the entire list of emphasize.

Two orders of emphasize were administered. One order was a randomized order given to approximately one half of the participants. The other was the reverse. The same rating scale as Experiment 1 was used. After completing the ratings, participants were debriefed and dismissed.

Table 3
Mean ratings and standard deviations of compliance intent in descending order of 37 conditions.

Emphasizer	Student	Nonstudent	Overall
	Mean (SD)	Mean (SD)	Mean (SD)
Extremely crucial	5.85 (2.1)	6.51 (1.7)	6.24 (1.9)
Urgent	5.98 (1.7)	6.41 (1.8)	6.23 (1.7)
Extremely vital	6.09 (1.9)	6.28 (1.7)	6.20 (1.8)
Critical	5.78 (1.9)	6.29 (1.9)	6.08 (1.9)
Federal Law	6.13 (2.4)	6.00 (2.2)	6.05 (2.3)
Extremely important	5.91 (1.7)	6.08 (1.8)	6.01 (1.7)
Mandatory	5.72 (1.9)	6.15 (2.0)	5.98 (2.0)
Absolutely necessary	5.94 (1.8)	5.99 (1.9)	5.97 (1.8)
Absolutely crucial	6.07 (1.7)	5.87 (2.1)	5.95 (1.9)
Very crucial	5.76 (1.9)	5.95 (2.0)	5.87 (2.0)
Absolutely vital	5.81 (2.1)	5.82 (2.1)	5.82 (2.1)
Very vital	5.44 (1.9)	6.06 (1.7)	5.81 (1.8)
Vital	5.72 (1.8)	5.86 (2.0)	5.80 (1.9)
State Law	5.67 (2.4)	5.65 (2.2)	5.65 (2.3)
Extremely essential	5.39 (1.7)	5.81 (2.1)	5.64 (2.0)
Absolutely important	5.54 (1.9)	5.68 (1.9)	5.62 (1.9)
Required	5.56 (2.3)	5.66 (1.9)	5.62 (2.1)
Very important	5.46 (1.8)	5.62 (1.9)	5.56 (1.9)
Must	5.22 (2.1)	5.67 (2.1)	5.49 (2.1)
Crucial	5.39 (1.8)	5.54 (2.0)	5.48 (1.9)
Extremely necessary	5.39 (1.8)	5.48 (2.0)	5.44 (1.9)
Absolutely essential	5.39 (1.8)	5.41 (2.3)	5.40 (2.1)
Very necessary	5.30 (1.8)	5.46 (2.0)	5.39 (1.9)
Very essential	5.17 (1.6)	5.46 (2.0)	5.34 (1.9)
Strongly recommended	5.15 (1.6)	5.27 (2.0)	5.22 (1.9)
Essential	4.91 (1.7)	5.18 (2.0)	5.07 (1.9)
Strongly suggested	4.91 (1.6)	5.18 (2.0)	5.07 (1.9)
Imperative	4.85 (2.2)	5.19 (2.2)	5.05 (2.2)
Important	4.85 (1.6)	5.06 (1.8)	4.98 (1.7)
Necessary	4.50 (1.9)	5.14 (1.9)	4.88 (1.9)
Recommended	3.43 (1.9)	4.47 (2.2)	4.05 (2.1)
Compulsory	3.98 (2.0)	3.44 (2.7)	3.66 (2.4)
Suggested	2.83 (1.5)	3.61 (2.0)	3.29 (1.8)
Discretionary	3.26 (2.2)	3.09 (2.2)	3.16 (2.2)
Please	2.93 (1.6)	3.28 (2.3)	3.14 (2.1)
Voluntary	2.20 (2.0)	2.44 (2.0)	2.35 (2.0)
Optional	1.93 (1.7)	2.43 (2.2)	2.23 (2.0)

3.2. Results

Three main analyses were conducted. The first was an analysis of the individual 37 emphasize conditions. Table 3 shows the means and standard deviations of the terms arranged in order from high to low compliance intent. The range of the two extreme means was nearly 4 points on the rating scale, and according to the anchor descriptions accompanying the rating scales, the means ranged from “very likely to obey” to “somewhat likely to obey.” The highest rated items were “extremely crucial” ($M = 6.24$) and “urgent” ($M = 6.23$), while the lowest rated was “voluntary” ($M = 2.35$) and “optional” ($M = 2.23$). A one-way ANOVA on the 37 emphasize conditions indicated there was a significant effect, $F(36, 4752) = 72.77$, $MSE = 2.17$, $p < .0001$. Tukey’s honestly significant difference (HSD) test set at $p = .05$ was found equal to .59. Any comparison of two means in this table is significant if the difference is greater than the HSD.

A second analysis examined the effect of adding a qualifier (e.g., strongly) to intensifiers (e.g., “important”). Twenty emphasize conditions from the 37-item list could be assembled to form a two factor analysis of intensifiers with and without an intensifier. A 5 (intensifier: “important,” “crucial,” “essential,” “necessary,” and “vital”) \times 4 (qualifier: no qualifier, “very,” “absolutely,” and “extremely”) repeated-measures ANOVA showed significant main effects for both intensifier, $F(4, 536) = 14.51$, $MSE = 2.51$, $p < .0001$, and qualifier, $F(3, 402) = 25.43$, $MSE = 2.14$, $p < .0001$.

For the main effect of intensifier, Tukey’s HSD at $p = .05$ is .52. The terms “vital” ($M = 5.88$) and “crucial” ($M = 5.87$) had the highest mean ratings. Both were significantly higher than “essential” ($M =$

5.34). “important” ($M = 5.51$) was the next higher but not significantly differing from any other intensifiers. For the main effect of qualifiers, Tukey’s HSD test at $p = .05$ is .45. The term “extremely” ($M = 5.88$) was the highest rated, followed by “absolutely” ($M = 5.72$), “very” ($M = 5.57$), and no qualifier ($M = 5.21$). The only significant difference was between “extremely” and no qualifier.

The interaction was also significant, $F(12, 1608) = 5.22$, $MSE = 1.03$, $p < .0001$. Tukey’s HSD at $p = .05$ was found equal to .39. The cell means for the interaction are displayed in Fig. 2. The analysis of simple effect showed that interaction appeared to be due to intensifier “necessary” having a somewhat different pattern (compared to other intensifiers) when paired with some of the qualifiers. The rating of “necessary” when combined with the qualifier “absolutely” was much higher than other intensifiers but was much lower when paired with the term “extremely.” The rest of the intensifier plus qualifier combinations showed roughly parallel increases across the board from no qualifier, “very,” “absolutely,” to “extremely.”

Given the finding in Experiment 1 that showed that “federal law” was the highest-rated emphasize condition, it was re-examined with the addition of “state law” that was added to the list in this experiment. As can be seen in Table 3, both were given high ratings of compliance intent, with “federal law” being among the highest-rated emphasize conditions. Note that, too, that it was not significantly lower than “extremely crucial,” which received the highest mean rating in the entire set. A planned comparison indicated that “federal law” was rated significantly higher than “State Law,” $F(1, 137) = 21.46$, $p < .001$.

Additional analyses were done with respect to student versus non-student groups and younger participants versus older participants. For the age analysis, a median split at 23.5 years was used to divide the sample by age into older and younger adults. Analyses used a 2 (participant group) \times 37 (emphasizers) mixed model ANOVA. Neither participant group analysis showed a significant effect of group or interaction ($ps > .05$).

3.3. Discussion

The findings support those found in Experiment 1. Indeed a correlation using the means of the corresponding emphasize terms between the two Experiments was $r = .95$ ($p < .05$). While the ordering of the emphasize conditions in common between the two experiments was quite high, there were differences in the absolute magnitude of the means between the two experiments. The reason for the differences might be due to the use of somewhat different instructions and the context of having different words in the list. Other researchers have found that the composition of the list of warnings can have an effect on the ratings (Chen, Gilson, & Mouloua, 1997).

The emphasize terms showed a broad range of compliance intent. Some emphasize conditions (e.g., “urgent” and “extremely crucial”) were given high mean ratings that indicated “very likely to obey,” whereas others (e.g., “voluntary,” “optional”) indicated “somewhat likely to obey” on the rating scale.

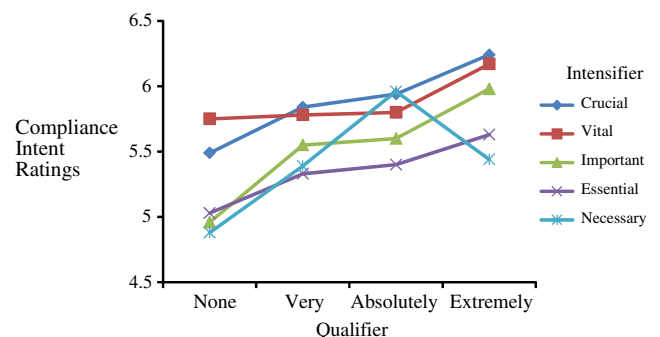


Fig. 2. Means of the intensifier \times qualifier interaction on compliance intent.

“Federal law” was among the highest-rated emphasizeers, and was rated significantly higher than “State Law.” Possibly this is due to frequently federal law almost always takes precedence over state law and may be reflective of the fact that the former is more powerful than the latter. The high ratings are supported by research, suggesting that law-related cues and consequences (such as fines for violation) have been shown to enhance personal-protection compliance behavior for helmet and seat belt warnings (e.g., Lehto & Foley, 1991). These results suggest that it may be worthwhile to include mention of legal requirement warnings if it is applicable. However, other results concerning variability raise a question about the use of law in the directives.

The factorial examination of two-word phrases emphasizeers revealed that the presence of one of the qualifiers, “extremely,” made compliance intent stronger when added to the intensifiers compared to no qualifier with the exception of “necessary.” The other qualifiers in this analysis, “very” and “absolutely,” were generally higher than no qualifier but not significantly. Adding one- or two-word emphasizeers will make a lengthier warning, which will be in opposition to a fundamental guideline that warnings should be as brief as possible. This, however, should not present much of a problem in most cases because we are talking about a few terms that on the positive side appear to affect compliance intent in a positive manner. Clearly, one would not want to waste valuable space by using emphasizeers that do not affect compliance intent or reduce it.

In addition, the interaction pattern showed that when the qualifiers “absolutely” or “extremely” were added to “necessary,” a somewhat different pattern was observed compared to the other intensifiers. A definitive reason for this difference is unknown, but one possible reason may relate to current usage in English where the phrase “absolutely necessary” is commonly used for emphasis, whereas “extremely necessary” is used less commonly.

Table 3 not only shows the means but also the standard deviations. Higher variability (i.e., high standard deviations) is indicative of confusion or interpretation differences among participants (cf. Wogalter & Silver, 1995). If some of the standard deviations were relatively high, it would suggest lower understanding among participants as it indicates different interpretations of the words. In this study, the standard deviations are small and relatively consistent, suggesting that they are all approximately interpreted similarly. There were a few that had somewhat higher variability than others such as “compulsory” and “imperative.” Note, too, that the two emphasizeers with “law” in them, “federal law,” and “state law” were on the upper range of standard deviations. Thus, while there is some indication that the two law-related items are capable of raising compliance intent, there is also some indication of variability in people’s belief about them.

4. Experiment 3

Although the results thus far suggest that adding emphasizeers to warning directives/instructions is beneficial in affecting compliance intent, the first two experiments did not explicitly consider the terms’ understandability. However, the relatively small standard deviations (i.e., low variability) suggest that there is consistency in participants’ interpretations—otherwise there would be high variability in the ratings. These variability numbers are an indirect measure of understandability. One of the goals of Experiment 3 is to determine more directly the terms’ understandability using other measures. Understandable warnings are essential for effective safety communications. The Communications-Human Information Processing (C-HIP) model (see Wogalter, DeJoy, & Laughery, 1999; Wogalter, 2006) has comprehension as an important processing stage prior to behavioral compliance (Laughery & Wogalter, 2006). Comprehension of warnings and their components can be measured in many ways. One way is to use open-ended recall-type tests of knowledge. Other methods include recognition type tests such as multiple choices, true–false, and matching.

Another common method of measuring understandability is to obtain people’s judgments on this dimension on the warning or its components. In the current research, participants were asked to rate the degree to which they or others would understand the warning messages. The same 37 emphasizeers as Experiment 2 were used.

The emphasizeers were also examined using more objective measures that are commonly used as a basis of readability measurements (e.g., Flesch, 1948). More frequently used words in the English language are generally better understood than less frequently used words (Wogalter & Silver, 1995). Additionally, shorter words in terms of numbers of letters and syllables are generally more understandable (Duffy, Kalsher, & Wogalter, 1995; Laughery & Wogalter, 1997). In the current research, understandability ratings and these other measurements (e.g., frequency of occurrence, length of words and syllables) are employed to evaluate emphasizeers in warnings instruction.

4.1. Methods

4.1.1. Participants

A total of 97 individuals (40 females, 56 males, 1 missing a gender indication) participated. Average age was 30.3 years ($SD = 13.2$). Samples from two population pools were collected: 52% were undergraduates from a large southeastern U.S. university ($M = 22.8$ years; $SD = 7.8$) and 48% were nonstudent adult volunteers from the surrounding community ($M = 38.3$ years; $SD = 13.1$). Twenty-two percent ($n = 21$) were non-native English speakers.

4.1.2. Materials and procedure

All stimuli and procedure were identical to Experiment 2 except for evaluated dimension, perceived understandability. Participants were asked to rate how understandable they believed each of the 37 items to be. In making their ratings, they were asked to consider all persons in living in the United States, including children, immigrants, and all adults including low skill readers. A 9-point Likert scale was provided with text labels assigned to the even-numbered anchors: (0) not at all understandable, (2) somewhat understandable, (4) understandable, (6) very understandable, and (8) extremely understandable.

Several objective indications of understandability of the emphasizeers were examined. One indication is the frequency of occurrence in the English language. For this, Davies (2008) Corpus of Contemporary American English was used. It is composed of one of the largest body of American English words available online, containing 450 million words from diverse genres and domains including spoken, fiction, magazine, newspaper, and academic journals. The measure of standardized frequency of occurrence (per million) from this database was used. Another indication of understandability is variability in the ratings. A similar measure was collected in Experiment 2 with the compliance intent ratings. Thus, the standard deviations for both the understandability ratings from the present experiment and compliance intent from Experiment 2 were examined. The last set of understandability indications used is word length, specifically letter and syllable counts. These measures are shown in Table 4 with the exception of the standard deviations of compliance intent, which are given in Table 3 from Experiment 2.

4.2. Results

4.2.1. Perceived understandability ratings

Understandability ratings were collapsed across participants to form mean scores for each word. The means and standard deviations of 37 emphasizeers are shown Table 4. A one-way repeated measures ANOVA indicated a significant main effect of emphasizeer, $F(36, 3312) = 38.15$, $MSE = 2.53$, $p < .0001$. Tukey’s HSD test at $p = .05$ equals to .76. According to the anchor descriptions accompanying the rating scales, these means ranged from “somewhat understandable” to a level between “very understandable” and “extremely understandable.” The highest-rated emphasizeer on understandability was “very important” ($M =$

Table 4
Mean (and SD) perceived understandability ratings and objective understandability measures for 37 words/phrases in descending order.

Emphasizer	Understandability, mean (SD)	Frequency of occurrence/million	Number of letters	Number of syllables
Very important	6.79 (1.3)	25.27	13	5
Required	6.62 (1.7)	94.5	8	2
Extremely important	6.60 (1.5)	2.63	18	6
Important	6.52 (1.6)	354.57	9	3
Must	6.34 (1.7)	413.06	4	1
Urgent	6.26 (1.8)	10.73	6	2
Absolutely necessary	6.18 (1.7)	1.13	19	8
Federal law	6.08 (2.2)	5.66	10	4
Absolutely important	6.07 (1.7)	0.05	19	7
Strongly recommended	6.04 (1.7)	0.2	19	6
State law	6.02 (2.1)	4.52	8	2
Extremely necessary	6.01 (1.7)	0.01	18	7
Please	5.98 (2.4)	94.47	6	1
Necessary	5.94 (1.8)	99.76	9	4
Very necessary	5.89 (1.8)	0.23	13	6
Recommended	5.76 (1.9)	28.23	11	4
Strongly suggested	5.56 (1.7)	0.21	17	5
Mandatory	5.51 (2.0)	11.4	9	4
Very essential	5.24 (1.9)	0.09	13	5
Extremely essential	5.22 (1.9)	0	18	6
Suggested	5.18 (2.0)	69.7	9	3
Critical	5.07 (2.1)	88.94	8	3
Absolutely essential	5.05 (2.0)	0.82	19	7
Optional	5.03 (2.4)	10.45	8	3
Absolutely crucial	4.94 (2.0)	0.6	17	6
Extremely crucial	4.94 (2.2)	0.02	16	5
Essential	4.94 (2.2)	47.02	9	3
Extremely vital	4.89 (2.1)	0.01	14	5
Very crucial	4.85 (2.2)	0.24	11	4
Voluntary	4.72 (2.4)	13.08	9	4
Very vital	4.68 (2.3)	0.1	9	4
Vital	4.66 (2.1)	26.65	5	2
Crucial	4.44 (2.1)	34.83	7	2
Absolutely vital	4.36 (2.3)	0.17	15	6
Imperative	3.92 (2.2)	7.14	10	4
Compulsory	2.53 (2.5)	2.46	10	4
Discretionary	2.19 (2.0)	2.9	13	

Table 5
Mean ratings (standard deviations) and mean difference of understandability between age groups of 37 conditions.

Emphasizer	Younger mean (SD)	Older mean (SD)	Mean difference
Very important	7.02 (1.1)	6.57 (1.4)	.45
Required	6.96 (1.4)	6.29 (1.9)*	.67
Extremely important	6.67 (1.4)	6.53 (1.7)	.14
Important	6.88 (1.2)	6.16 (1.9)*	.71
Must	6.38 (1.6)	6.31 (1.8)	.07
Urgent	6.31 (1.9)	6.20 (1.6)	.11
Absolutely necessary	6.21 (1.6)	6.14 (1.9)	.65
Federal law	6.71 (2.0)	5.47 (2.3)*	1.24
Absolutely important	6.27 (1.7)	5.90 (1.8)	.38
Strongly recommended	6.23 (1.5)	5.86 (1.9)	.37
State law	6.71 (1.7)	5.35 (2.3)*	1.36
Extremely necessary	6.19 (1.6)	5.84 (1.7)	.35
Please	5.92 (2.5)	6.04 (2.4)	-.12
Necessary	5.96 (1.7)	5.92 (1.8)	.04
Very necessary	6.17 (1.3)	5.61 (2.1)	.55
Recommended	5.94 (1.8)	5.60 (2.0)	.35
Strongly suggested	5.73 (1.6)	5.42 (1.7)	.31
Mandatory	5.81 (1.9)	5.20 (2.1)	.61
Very essential	5.54 (1.7)	4.94 (2.1)	.60
Extremely essential	5.29 (1.8)	5.17 (2.1)	.12
Suggested	5.40 (1.9)	4.96 (2.1)	.44
Critical	5.08 (2.1)	5.06 (2.1)	.02
Absolutely essential	5.21 (2.0)	4.90 (2.0)	.31
Optional	5.75 (2.2)	4.33 (2.4)*	1.42
Absolutely crucial	5.27 (1.9)	4.61 (2.1)	.66
Extremely crucial	5.52 (2.1)	4.37 (2.1)*	1.15
Essential	5.44 (1.8)	4.35 (2.1)*	1.09
Extremely vital	5.50 (1.9)	4.20 (2.3)*	1.30
Very crucial	5.15 (2.1)	4.55 (2.2)	.60
Voluntary	5.54 (2.3)	3.92 (2.3)*	1.62
Very vital	5.21 (2.0)	4.16 (2.4)*	1.05
Vital	5.17 (1.9)	4.16 (2.2)*	1.00
Crucial	4.94 (1.8)	3.96 (2.3)*	.98
Absolutely vital	5.15 (2.1)	3.56 (2.2)*	1.58
Imperative	4.27 (2.2)	3.57 (2.1)	.70
Compulsory	2.96 (2.6)	2.10 (2.3)	.86
Discretionary	2.58 (2.0)	1.80 (1.8)*	.79

* Significantly difference between age group, $p < .05$

6.79), which was followed by “required” ($M = 6.62$), “extremely important” ($M = 6.60$), “important” ($M = 6.52$), and “must” ($M = 6.34$). The lowest-rated emphasize on understandability was “discretionary” ($M = 2.19$), with other low rated items being “compulsory” ($M = 2.53$), “imperative” ($M = 3.92$), “absolutely vital” ($M = 4.36$), and “crucial” ($M = 4.44$).

Additional analysis was conducted on the perceived understandability ratings with the demographic factors, age, and students versus non-students groups. Two ANOVAs were conducted: (a) 2 (age group) \times 37 (emphasizers) mixed model and (b) 2 (students vs. nonstudents) \times 37 (emphasizers) mixed model ANOVA. Age group (median split at 24 years) and students versus nonstudents group were between subjects factors used individually in each ANOVA and the emphasize variable was the within subjects factor. For the ANOVA involving age-group, there was significant main effects of this factor, $F(1, 91) = 6.43$, $MSE = 54.77$, $p < .05$, and emphasize, $F(36, 3276) = 38.84$, $MSE = 2.50$, $p < .001$, and a significant interaction effect between age and emphasize, $F(36, 3276) = 2.16$, $MSE = 2.50$, $p < .05$. In general, younger participants gave higher ratings ($M = 5.65$) than older participants ($M = 5.0$). Simple effects analysis showed that interaction effect appeared due to younger participants giving higher ratings to some emphasize than the older participants, such as the two law-related emphasize (“federal law,” “state law”) and emphasize that had the term “vital” (e.g., “very vital,” “absolutely vital,” and “extremely vital”) as well as several others. Table 5 shows means and standard deviations of both groups as well as the mean difference between the two groups.

For the ANOVA involving students and nonstudents, there was significant main effects for students versus nonstudents, $F(1, 91) = 4.36$, $MSE = 1.51$, $p < .05$ and emphasize, $F(36, 3276) = 38.84$, $MSE =$

2.50, $p < .001$. Mean ratings showed that students ($M = 5.59$) gave higher ratings than nonstudents ($M = 5.06$). There was no significant interaction effect between students and emphasize ($p > .05$).

To examine the effects of combination of intensifiers and qualifiers systematically, 20 of the 37 emphasize could be assembled to form a repeated-measures factorial design to examine the effects of adding different qualifiers (or no qualifier) to several intensifiers (i.e., qualifier plus intensifier vs. intensifier only). This analysis is similar to the one examined in Experiment 2 for compliance intent. A 4 (qualifiers: no qualifier, “very,” “absolutely,” and “extremely”) \times 5 (intensifiers: “important,” “crucial,” “essential,” “necessary,” and “vital”) two-way repeated-measures ANOVA showed significant main effects for both intensifier, $F(4, 368) = 50.85$, $p < .01$, $MSE = 4.79$, $p < .01$, and qualifier, $F(3, 276) = 3.95$, $MSE = 1.76$, $p < .01$. Tukey’s HSD at $p = .05$ for the intensifier main effect is .86 and for the qualifier main effect is .49. Comparisons among the intensifier means indicated that “important” had the highest mean rating ($M = 6.51$), but it was not significantly higher than “necessary” ($M = 6.02$), both of which were significantly higher than “essential” ($M = 5.11$), “crucial” ($M = 4.79$), and “vital” ($M = 4.66$), which did not differ among themselves. Comparisons among the qualifier means indicated that “extremely” was rated highest ($M = 5.54$) but was not significantly different from the other qualifiers, “very” ($M = 5.51$), “absolutely” ($M = 5.32$), and no qualifier ($M = 5.30$). Thus, in this case while the ANOVA showed a small overall effect of qualifier, none of the paired comparisons were different.

The interaction effect was also significant, $F(12, 1104) = 2.58$, $MSE = 1.43$, $p < .01$, with Tukey’s HSD ($p = .05$) equal to .54. The cell means for the interaction are shown in Fig. 3. The interaction effect

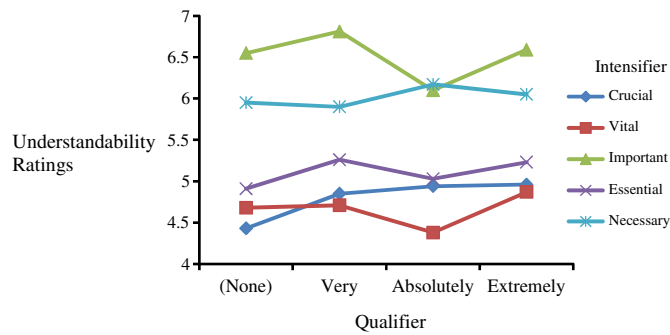


Fig. 3. Means of the intensifier \times qualifier interaction on understandability.

appeared due to certain combinations of intensifiers and qualifiers. When “absolutely” was paired with “vital” and “important,” the mean ratings dropped relative to the levels with the other qualifiers.

4.2.2. Relationships between measures

Pearson product moment correlations were conducted among the measures using the data from both Experiments 2 and 3. These included the compliance intent means and standard deviations (from Experiment 2), and understandability means and understandability standard deviations, frequency of occurrence, number of letters, and number of syllables (from Experiment 3) so as to reveal relationships among them (Dickter, 2006). The correlation matrix is shown in Table 6. The compliance intent means had a significantly positive relationship with the understandability means ($r = .35, p < .05$) but had a negative relationship with the understandability SDs ($r = -.39, p < .05$). This indicates that higher compliance intent is associated with higher understandability of the emphasizees as well as lower variability (SDs) in the latter ratings. Emphasizer understandability was negatively associated with the SDs for both compliance intent ($r = -.46, p < .05$) and understandability ($r = -.63, p < .05$). The SDs for both compliance intent and understandability were positively associated ($r = .49, p < .05$).

The frequency-of-use measure had negative relationships with number of letters ($r = -.45, p < .05$) and number of syllables ($r = -.49, p < .05$), indicating that the fewer letters and syllables, the more frequent the term is used in the English language. Lastly, the number of letters had a significant negative relationship with understandability SD ($r = -.34, p < .05$).

4.3. Discussion

It is a general principle that the content of warnings should have language that the target audiences clearly understand (Laughery & Wogalter, 1997). Thus, if emphasizees are to be used in warning directives, they ought to be readily understood since their job would be to influence people by affecting and promoting compliance. Emphasizees give information about how important it is to comply and this would be missed if people do not realize what the terms were meant to convey.

Participants rated terms such as “important,” “required,” “must”, and “urgent” as “very understandable” (according to the rating scale), whereas they rated other words such as “compulsory,” “discretionary,”

“imperative,” “crucial,” and “vital” as only “somewhat understandable” (according to the rating scale). Thus, in terms of understandability alone, the former group of terms would be better to use as emphasizees than the latter group.

Understandability in this experiment and compliance intent in the last experiment correlated moderately and significantly ($r = .35$). This is not unexpected. Terms that people do not understand are also likely not to promote high compliance intent. Interestingly, the standard deviations for both measures related to each other.

Significant relationships among compliance intent, understandability, and other measurements suggest that more understandable emphasizees may do a better job at motivating people to comply with the warning directives. Additionally, the results indicated that lower variability in the understandability ratings are associated with greater understandability and higher willingness to comply. This confirms previous studies that have found that variability measures to be an indication of how well the terms are understood (Wogalter & Silver, 1995).

The SDs for both compliance intent and understandability were fairly consistent. The highest correlation in Table 6 shows higher mean understandability is associated with lower understandability SDs ($r = -.63$). The fact that compliance intent SDs were significantly negatively correlated with understanding indicates that one could use SDs of the rated variable as an indication of their understandability—without needing to collect specific separate data on the terms understandability (Wogalter & Silver, 1995). At this point, however, measurement of understanding would probably be advisable instead of using a proxy variability measure. That is, ratings of terms that show a wide amount of rating variability along a dimension by participants could be an indication that people are not all consistently understanding the terms. Varied interpretations of warning messages could cause negative consequences as participants may fail to realize the importance and necessity of carrying out the compliance activity. In selecting words to use in warnings, the selection ought to consider measures of understandability using terms with high understandability and low SDs in the measures.

Interestingly, there was no substantial relationship with any of the measures associated with the objective counts (frequency, letters, and syllables) with compliance intent and understandability (and their associated SDs). It would be interesting to see there might be other objective measures that relate to participants' compliance intent and understandability evaluations. If so, they might be used to help predict the utility of phrases and wording in warnings, which would be useful in early phases of warning development in culling warning verbiage that are less likely to be effective.

One unexpected result was found in Table 6's correlation matrix. This was a small but significant negative relationship that was found between number of letters and understandability SDs ($r = -.34$). This would indicate that the greater the number of letters of the emphasizees, the lower the understandability SD. This seems the opposite of expected although few would predict a strong relationship in any case. Given the .05 probability criterion, one in 20 correlations will be significant by chance alone. A replication in future research would help to determine whether this result is reliable.

Age and participant group showed some small but significant effects in the understandability ratings of the emphasizees. These two factors were overlapping to some extent since many of the younger adults

Table 6
Correlation coefficients of measurements of 37 word conditions.

	Compliance intent SD	Understand ability mean	Understand ability SD	Frequency of use	Number of letters	Number of syllables
Compliance intent mean	-.27	.35*	-.39*	-.06	.27	.24
Compliance intent SD		-.46*	.49*	-.10	-.13	-.12
Understandability mean			-.63*	.31	.06	-.02
Understandability SD				-.27	-.34*	-.28
Frequency of use					-.45*	-.49*
Number of letters						.94*

* $p \leq .05$

were students and many of the nonstudent community volunteers were older than the students. Both analyses showed main effects of participant groups with the students and younger participants giving higher ratings to the emphasizees than nonstudents and older participants. There was a small interaction in the analysis of younger versus older participants where certain terms were rated higher in understandability by younger than by older participants (e.g., “federal law,” “state law,” “very vital,” “absolutely vital,” “extremely vital,” “optional,” and “voluntary”). The reason for this finding is not clear at this point without additional data but may be due to generational and educational differences in exposure to the terms. Additional research would be needed to clarify why this particular pattern was found for certain words and not for others.

5. General discussion

Although there has been extensive research on signal words and other components of warnings there has been little research with respect to the directives/instructions statement. Directives are important because they provide information in warning on what to do to avoid the hazard or consequence being warned about. Specifically, this research examined compliance intent and understandability for a set of emphasizeer terms (qualifiers and intensifiers) that could be added to warning directives to potentially influence the extent of importance or the relative necessity of complying with the base directive. Experiment 1 showed that some emphasizees (e.g., “mandatory,” “federal law”) produced higher compliance intent compared no emphasizeer (control condition). Two emphasizees (e.g., “recommended,” “suggest”) produced lower compliance intent than the control condition although it was not significant. Thus, it depends on the emphasizeer words themselves as to whether they might add any urgency in compliance intent. Because some of the effects of the emphasizeer words were low and that adding these words to a warning directive adds to its length (violating to some extent the brevity guideline), the selection of emphasizeer used in a directive should not be casual and haphazard.

Experiment 2 used more terms and more participants than in Experiment 1 and showed that emphasizeer terms can influence compliance intent ratings differentially with some terms rated higher and other rated lower. Clearly, if one wants to raise compliance intent over that of a base directive, some of the highest-rated emphasizees could be used.

This study also examined the effect of emphasizees containing only intensifiers (e.g., “important,” “necessary”) or those having both an intensifier and a qualifier (e.g., “very important,” “absolutely necessary”). The results showed that having both a qualifier and an intensifier generally resulted in higher compliance intent. Although these two factors produced a small interaction, the qualifier main-effect means generally showed a pattern in which “extremely” was the highest rated, followed by “absolutely,” “very,” and no qualifier.

Experiment 3 measured perceived understandability to the 37 emphasizees and a wide range of mean ratings were shown (e.g., “very important” as very understandable and “discretionary” as somewhat understandable). Unlike the compliance intent results, adding qualifiers to intensifiers did not change much with respect to their understandability, although the intensifiers differed with “important” and “necessary” higher in understandability than “crucial” and “vital.”

Using the data from both Experiments 2 and 3, it was found that compliance intent means of the facilitators was positively related to their rated understandability. This makes sense because people may be more likely to comply with directives in which they understand what is being expressed.

Analysis of the variability measurements showed some interesting findings. First, higher compliance intent was associated with lower standard deviations on that measure. Second, greater understandability was associated with lower standard deviations on that measure as well as the standard deviations of the compliance intent rating. The objective

measurements of comprehension (frequency of use and letter and word counts) did not show any strong association with the subjective measures collected.

Apparently, emphasizees added to the warning directives influenced people's willingness to comply, some heightening it (e.g., “necessary” or “required”) and some reducing it (e.g., “suggested” or “recommended”). Emphasizees such as “extremely important” may evoke a subjective understanding such as connoted sense of hazard, which in turn may affect comprehension of the subsequent statements (e.g., “wear XYZ protective eyewear”). On the other hand, weak emphasizees such as “suggested” could fail to evoke strong perceived hazard or necessity in performing behavior as directives, and that may yield lower compliance intent than the bare directive.

The law-related emphasizees (e.g., “federal law,” “state law”) were rated very understandable (Experiment 3) and also raised intentions of compliance in both Experiment 1 and 2. These results were consistent with the previous studies that specific and expert sources produced higher credibility ratings compared to no source alone (Wogalter et al., 1997, 1999) and are persuasive to change beliefs and attitudes of the warning messages (Cox & Wogalter, 2006). It is also possible that people may believe that the warning directives with those items may cause negative consequences (e.g., fine, penalty) for them if do not comply. However, both of these two law emphasizees had relatively high variability in the compliance intent and understandability ratings, which indicates that there is some inconsistency among participants on what they imply. According to these results some caution might be invoked when using these terms in warnings.

Because non-native English speakers might evaluate the terms differently than native English language users, additional analyses were performed. The analyses used the data in Experiment 2 and 3 in which 28% and 22% of the participants, respectively, reported being non-native English language. The mean compliance intent ratings and variability (i.e., standard deviations) of the non-native English speakers were generally higher than native speakers. These findings might be due to non-native English speakers having different exposure or education than native English speakers on the intended meaning of the terms. A more conclusive explanation would need a larger data sample of non-native English speakers.

Another purpose of the current research was to provide data on a large list of emphasizees that could be added to warning directives that vary in their connotation. This has been done by way of the tables of compliance intent and understandability means and SDs. These data could be useful in selecting emphasizees that best match the level of hazard involved (e.g., Edworthy & Adams, 1996). Over time and exposure, repeatedly used words can lose their connoted level of hazard (Kim & Wogalter, 2009). One way to avoid habituation of words that have the highest connotation of urgency or necessity to comply with the directive is to use them sparingly and only when there is a need for additional emphasis. If used only under the most important circumstances, the terms would have a slower rate of habituation. Thus, some of the lower rated emphasizees (or no emphasizeer) might be used for less serious warnings directives and higher rated terms for more serious warnings. With regard to making selections of words or terms, prior research suggests several criteria. Wogalter and Silver (1990) used understandability, variability, frequency of use, and conciseness/brevity as criteria for selecting signal words. We have provided these same kinds of data in the tables in this research. We suggest that emphasizees should be understandable with low variability in ratings, and brief. Terms like “urgent,” “very important,” and “absolutely necessary” are examples of terms that fit those criteria.

Highly understandable emphasizees (with low variability) might be classified into three levels based on compliance intent mean ratings of 6 (“very likely to obey”), 5 (between “very likely” and “likely”), and 4 (“likely to obey”). Table 7 shows a tentative list of emphasizees to illustrate a few terms based on three levels of compliance intent. However, this is only one example list that could be formed; others could be

Table 7

An example list of selected emphasizeers that could be added to warning directives.

Emphasizer	Compliance intent mean (SD)	Understandability mean (SD)
<i>Very high hazard</i>		
Urgent	6.23 (1.7)	6.3 (1.8)
Extremely important	6.01 (1.7)	6.6 (1.5)
Absolutely necessary	5.97 (1.8)	6.2 (1.7)
<i>High hazard</i>		
Absolutely important	5.62 (1.9)	6.1 (1.7)
Very important	5.56 (1.9)	6.8 (1.3)
Extremely necessary	5.44 (1.9)	6.0 (1.7)
Strongly recommended	5.22 (1.9)	6.0 (1.7)
<i>Intermediate hazard</i>		
Important	4.98 (1.7)	6.5 (1.6)
Necessary	4.88 (1.9)	5.9 (1.8)

formed based on other criteria. Selection based on reasonable criteria should be beneficial in assisting warning designers when they choose terms for warnings for their intended perceived hazard.

The current research has some limitations. One limitation is that many of the participants were undergraduate university students. We attempted to capture data from another group comprised of nonstudent adults from various venues and locales in the state of North Carolina. While the results found relatively few differences between these two groups in their ratings of the words, there is still the open question whether the results might differ for other potential populations that would see warnings. Further research would be needed to assess whether the data are generalizable to other groups of people.

Generalizability of the listed emphasizeers should also be considered. Emphasizer terms were evaluated in Experiments 1 and 2 as embedded in specific and general contexts, respectively. However, it is not known whether the current list of emphasizeers generalize to more externally valid situations such as complete warnings or in particular real situations such as more dangerous products. Future research is needed to evaluate whether the listed emphasizeers in actual warnings or valid contexts effectively communicate appropriate hazard levels to populations (Wogalter & Silver, 1995).

One direct follow-up is to actually place the words in entire actual warnings and measure changes in response. Another extension of this is to measure different aspects in a behavioral compliance task and measure performance levels such as speed and accuracy. Other types of follow-ups are studies that manipulate other characteristics of the wording such as the extent to which the seriousness and probability of consequences is conveyed. Other types of wording (e.g., giving the mechanism providing potential injury) could be explored.

6. Practical application

The results of this research showed that addition of emphasizeers to the warning directives or instructions influences individual's understandability and compliance intentions. Use of appropriate emphasis terms may be helpful in communicating beliefs or relative necessity of complying with an associated warning directive. The findings from this research could aid warning researchers or designers in selecting understandable terminologies that evoke different levels of compliance intentions.

Acknowledgments

Portions of this article were presented at the National Safety Council annual meeting (Kim, Wogalter, & Cowley, 2007a) in Chicago, Illinois (Experiment 1), and the 51th Annual Meeting of the Human Factors and Ergonomics Society (Kim, Cowley, & Wogalter, 2007b), Baltimore, Maryland (Experiment 2).

References

- American National Standards Institute (ANSI) (2011). *Accredited Standards Committee on Safety Signs and Colors. Z535.1-4*. Arlington, VA: National Electrical Manufacturers Association.
- Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. *Ergonomics*, 37, 265–275.
- Chen, J. Y. C., Gilson, R. D., & Mouloua, M. (1997). Perceived risk dilution with multiple warnings. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 831–835). Santa Monica, CA: Human Factors and Ergonomics Society.
- Cox, E. P., III, & Wogalter, M. S. (2006). Warning source (Chap. 8). In S. Wogalter (Ed.), *Handbook of Warnings* (pp. 111–122). Mahwah, NJ: Lawrence Erlbaum Associates.
- Davies, M. (2008). The Corpus of Contemporary American English: 450 Million Words, 1990–Present. Available online at <http://corpus.byu.edu/coca/>
- Dewar, R. (1999). Design and evaluation of graphic symbols. In H. J. G. Zwaga, T. Boersema, & H. C. M. Hoonhout (Eds.), *Visual information for everyday use: design and research perspectives* (pp. 285–303). London: Taylor & Francis.
- Dickter, D. N. (2006). Basic statistical analyses. In F. T. Leong, & J. T. Austin (Eds.), *The psychology research handbook* (pp. 293–305). London: SAGE.
- Dingus, T. A., Hathaway, J. A., & Hunn, B. P. (1991). A Most Critical Warning Variable: Two Demonstrations of the Powerful Effects of Cost on Warning Compliance. *Proceedings of the Human Factors Society*, vol. 35. (pp. 1034–1038).
- Duffy, R. R., Kalsher, M. J., & Wogalter, M. S. (1995). Interactive warning: an experimental examination of effectiveness. *International Journal of Industrial Ergonomics*, 15, 159–166.
- Dunlap, G. L., Granda, R. E., & Kustas, M. S. (1986). *Observer Perceptions of Implied Hazard: Safety Signal Words and Color Words*. IBM Technical Report TR 00.3428.
- Edworthy, J., & Adams, A. (1996). *Warning design: a research prospective*. London: Taylor and Francis.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 2211.
- Hellier, E., & Edworthy, J. (2006). Signal words (Chap. 30). In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 407–417). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kalsher, M. J., & Williams, K. J. (2006). Behavioral Compliance: Theory, Methodology, and Results (Chap. 23). In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 313–331). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kalsher, M. J., Wogalter, M. S., & Gilbert, C. M. (1992). Connoted quantity of food-label modifier terms. *Proceedings of the Human Factors Society*, 36. (pp. 528–532).
- Kim, S., Cowley, J., & Wogalter, M. S. (2007a). Emphasis terms for warning directives on compliance Intent. *Proceedings of the Human Factors and Ergonomics Society*, 51. (pp. 569–573).
- Kim, S., & Wogalter, M. S. (2009). Habituation, dishabituation, and recovery effects in visual warnings. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, 53. (pp. 1612–1616).
- Kim, S., Wogalter, M. S., & Cowley, J. (2007b). *Effect of emphasis terms for warning directives on compliance intent*. Congress & Expo 2007. Chicago, IL: National Safety Council.
- Kreifeldt, J. G. (1993). Expert opinions, fuzzy probabilities, and warnings: toward a mathematical method for evaluating warning 'read and heed' effectiveness. *Safety Science*, 16, 729–750.
- Laughery, K. R., & Paige-Smith, D. (2006). Explicit information in warnings (Chap. 31). In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 419–428). Mahwah, NJ: Lawrence Erlbaum Associates.
- Laughery, K. R., & Stanush, J. A. (1989). Effects of warning explicitness on product perceptions. *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 431–435). Santa Monica, CA: Human Factors and Ergonomics Society.
- Laughery, K. R., & Wogalter, M. S. (2006). Designing effective warnings. In Williges (Ed.), *Reviews of Human Factors and Ergonomics*, vol. 2. (pp. 241–271). Santa Monica, CA: Human Factors and Ergonomics Society.
- Laughery, K. R., & Wogalter, M. S. (1997). Warnings and risk perception. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1175–1197). New York: Wiley-Interscience.
- Laughery, K. R., Wogalter, M. S., & Young, S. L. (Eds.). (1994). *Human Factors Perspectives on Warnings: Selections from Human Factors and Ergonomics Society Annual Meetings 1980–1993*. Santa Monica: Human Factors and Ergonomics Society.
- Lehto, M. R., & Foley, J. P. (1991). Risk-taking, warning labels, training, and regulation: are they associated with the use of helmets by all-terrain vehicle riders. *Journal of Safety Research*, 22, 191–200.
- Lehto, M. R., & House, T. (1997). *Evaluation of the comprehension of hazard communication phrases by chemical workers*. International Ergonomics Association 13th Triennial Congress, Tampere, Finland, June 29–July 4.
- Leonard, S. D., Karnes, E., & Schneider, T. (1988). Scale values for warning symbols and words. In F. Aghazadeh (Ed.), *Trends in Ergonomics/Human Factors V* (pp. 669–674). Amsterdam: Elsevier.
- Smith-Jackson, T. L., & Wogalter, M. S. (2000). Users' hazard perception of warning components: an examination of colors and symbols. *Proceedings of the International Ergonomics Association & the Human Factors and Ergonomics Society Congress*, 44(6), 55–58.
- Wogalter, M. S. (2006). Scope of warnings (Chap. 1). In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 3–9). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wogalter, M. S., Allison, S. T., & McKenna, N. A. (1989). The effects of cost and social influence on warning compliance. *Human Factors*, 31, 133–140.
- Wogalter, M. S., DeJoy, D. M., & Laughery, K. R. (1999a). Organizing framework: a consolidated communication-human information processing (CHIP) model. In M. S. Wogalter, D. M. DeJoy, & K. R. Laughery (Eds.), *Warnings and Risk Communication* (pp. 15–24). London: Taylor and Francis.
- Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors*, 29(5), 599–612.

- Wogalter, M. S., Kalsher, M. J., Frederick, L. J., Magurno, A. B., & Brewster, B. M. (1998). Hazard level perceptions of warning components and configurations. *International Journal of Cognitive Ergonomics*, 2, 123–143.
- Wogalter, M. S., Kalsher, M. J., & Litynski, D. M. (1996). Influence of food label quantifier terms on connoted amount and purchase intention. *Proceedings of the Human Factors and Ergonomics Society*, 40. (pp. 385–389).
- Wogalter, M. S., Kalsher, M. J., & Rashid, R. (1997). Effect of warning signal word and source on perceived credibility and compliance likelihood. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association, IEA'97*, 3. (pp. 478–480).
- Wogalter, M. S., Kalsher, M. J., & Rashid, R. (1999b). Effect of signal word and source attribution on judgments of warning credibility and compliance likelihood. *International Journal of Industrial Ergonomics*, 24, 185–192.
- Wogalter, M. S., Laughery, K. R., & Mayhorn, C. B. (2012). Warnings and hazard communications (Chap. 32). In G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics* (pp. 889–911) (3rd ed.). New York: Wiley.
- Wogalter, M. S., & Silver, N. C. (1990). Arousal strength of signal words. *Forensic reports*, 3, 407–420.
- Wogalter, M. S., & Silver, N. C. (1995). Warning signal words: connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics*, 38, 2188–2206.
- Wogalter, M. S., Silver, N. C., Leonard, S. D., & Zaikina, H. (2006). Warning symbols (Chap. 12). In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 159–176). Mahwah, NJ: Lawrence Erlbaum Associates.
- Young, S. L., & Wogalter, M. S. (1998). Relative importance of different verbal components in conveying hazard-level information in warnings. *Proceedings of the Human Factors and Ergonomics Society*, 42. (pp. 1063–1067).

Soyun Kim is currently a research fellow at college of pharmacy at Chung-Ang University in Seoul, South Korea. Before joining, she completed a postdoctoral research fellow in the department of oral health policy and epidemiology at Harvard School of Dental Medicine. She received her Ph.D. in Human Factors and Ergonomics program in psychology from the North Carolina State University in 2011. She received a B.A. and M.A. in Psychology from Ewha Womans University.

Mike Wogalter is a professor emeritus in psychology at North Carolina State University. He received a Ph.D. in psychology from Rice University in 1986. He received a B.A. in Psychology from the University of Virginia and an M.A. from the University of South Florida. He was previously on the faculties of the University of Richmond and Rensselaer Polytechnic Institute. He currently resides in the Fort Lauderdale area.