·38·

APPLYING USABILITY ENGINEERING PRINCIPLES TO THE DESIGN AND TESTING OF WARNING TEXT

Michael S. Wogalter
North Carolina State University

Vincent C. Conzola
IBM Corporation

William J. Vigilante, Jr.
Robson Forensic, Inc.

ABSTRACT

This chapter describes methods for creating and testing text messages to be used in product warnings. Design and evaluation of warning text is described using the principles of usability engineering and user-centered design. First, pre-prototype activities, including audience definition and hazard analysis to define the warning's requirements, are described. Second, guidelines for creating and testing prototypes are given. Finally, methodologies for measuring effectiveness before and after product release are presented. This material can serve as a basis for the production of warning text.

INTRODUCTION

As this book demonstrates, there is a wealth of literature concerning various factors that influence warning effectiveness. However, information on the content of the warning text is relatively limited. There are research and guidelines on formatting characteristics, such as font size. However, the recommendations regarding the printed message's content is mainly covered by a few broad statements. The three main types of statements are information about the hazard, instructions on how to avoid the hazard, and a description of potential consequences of noncompliance. Several other guidelines include keeping the warning brief and giving specifics rather than generalities. All of the above guidelines have been supported by research (see, for example, Wogalter, DeJoy & Laughery, 1999).

The guidelines are useful for forming prototype warnings (early candidate warnings), but because products differ in various ways, including their characteristics and use, and in the hazards they potentially pose, the content derived from merely using guidelines may not be optimal for the warning. There has been very little literature describing the process of creating and testing the text messages for product warnings. Frantz, Rhoades, and Lehto (1999) described a general, four-step process (project planning, identifying and understanding product hazards, developing warning prototypes, and evaluating warning prototypes)

Handbook of Warnings, Edited by Michael S. Wogalter, © 2006, Lawrence Erlbaum Associates (Mahwah, NJ).

for developing warnings. Frantz et al. gave an excellent overview of the warning process, but the specifics involved in developing warning text were only covered broadly.

There are several reasons for the limited literature in this area. One is that standards and guidelines address mainly shorter, highly edited word messages. These are warnings that are posted as signs on the wall of an industrial facility, on a fence around an electrical transformer station, or as a warning label on the side of a forklift. Not a whole lot of development time is needed to come up with warning text for an electrical hazard or for a mechanical stamping press, such as "High Voltage" or "Your hands can be crushed." The need for testing these relatively brief textual messages is much less crucial than in other warning applications with more complex messages for products with multiple hazards. The latter application is less frequently tackled in current standards and guidelines, probably because of its complexity. A second reason for the limited literature on the topic is that many of the people who know the basic process do not work with warnings specifically. The persons most knowledgeable about the complete process of design development and evaluation are in areas related to human factors and ergonomics (HF/E). HF/E professionals are extensively trained on the various processes involved in conducting iterative design and evaluation tests. However, their work is usually tied to the growing areas of human use of technology and, more specifically, human-computer interaction. Few HF/E individuals make their living testing warnings because their jobs do not call for it. A third reason for the limited literature is that the process itself is complex and is further complicated by limited time and resources. Every project is somewhat different. The product, its particular hazards, and the environment in which the product may be used will likely vary from project to project. Although extensive pre-project planning may be employed, numerous decisions have to be made on the fly as the project moves along. The deliverable for a contracted project may involve a short report, not a highly detailed document. The lack of documented details combined with limited time and resources has likely slowed the available literature on the topic.

Although there are several reasons for the relatively limited and highly general literature on the development and evaluation of warning text, it is also somewhat surprising. The reason is that there is already a rather large body of literature concerning symbol development and evaluations. There are guidelines and standards (e.g., American National Standards Institute [ANSI, 2002] Z535.3) that describe how to produce prototype symbols and evaluate them with respect to comprehension criteria. For example, for a symbol to be used without accompanying text, it must be shown to produce 85% correct comprehension by a sample of 50 participants, with the errors producing no more than 5% critical confusions. There are no criteria like this in the ANSI standards or any other guidelines with respect to producing or evaluating the textual components of warnings.

It is important to define procedures for developing warning text because many products have multiple hazards that cannot be addressed in simple words or symbols. Consider the hazards associated with a chemical solvent, a pharmaceutical product, or a power tool. These products have multiple hazards, many of which may not be well known by the user populations. This chapter is about developing effective warning text for these kinds of products.

This chapter attempts to show how the principles of usability engineering and user-centered design (UCD) can be applied to the development and testing of warning text messages. These are techniques that have been used successfully in the field of human-computer interaction to design and evaluate computer interfaces. Usability engineering is a practical and systematic process for ensuring that the needs, expectations, and limitations of users are considered during product development. There are three main principles of usability engineering: (a) prototyping, (b) empirical user testing, and (c) iterative design (e.g., Nielsen, 1992). UCD is somewhat broader than usability engineering in that UCD ensures that users' requirements and needs are incorporated and addressed within the product development cycle (International Organization for Standardization, [ISO], 1999; IBM, 2003).

Guidelines are given for many parts of the process, such as who will need the warning, which hazards should be warned against, and how to create preliminary versions of warning text. This is followed by a discussion of prototyping techniques. Finally, a methodology is proposed for formally evaluating the text's effectiveness and for validating its utility once the product is in the marketplace. Parts of this methodology are similar to that described in Annex B to ANSI Z535.3 (2002) for evaluating safety symbols (see also Wolff & Wogalter, 1998), but instead it is applied to text. The methodologies presented in this chapter are only a subset of many that might provide acceptable warning text. They are offered to assist practitioners who may be asked to develop and evaluate warning text. The guidelines and methods proposed here should not be considered definitive, but rather as giving some direction where otherwise these tasks might be done without any explicit guidance. The procedures can be adapted as necessary to satisfy the requirements of a specific situation (Wogalter, Conzola, & Vigilante, 1999; Wogalter & Vigilante, 2001).

UCD PROCESS

Within the usability literature, there are a number of documents describing the activities and steps involved in the UCD process and how the process directly influences the usability and usefulness of consumer and commercial products (e.g., Garrett, 2003; Garrity, 2001; ISO, 1999; Kwahk, Smith-Jackson, & Williges, 2001; Norman & Draper, 1986; Noyes & Baber, 1999; Pradeep, 1998; Vredenburg, Isensee, & Righi, 2002). Although the documents may differ in the number or order of defined steps and the activities performed at each step, there are several basic aspects involved in implementing a UCD process. As a framework for the creation of a UCD-based process for developing warning text, this chapter makes use of a generalized UCD process used by IBM Corporation in the development of consumer and commercial products (IBM, 2003; Sawin, Yamazaki, & Kumaki, 2002; Vredenburg et al., 2002). The six basic steps

as applied to the development and evaluation of warning text are as follows:

1. Identify the target audience: Determine who will need the warning information.
2. Hazard/task analysis: Determine how the product will be used or foreseeably misused, identify potential sources of injury associated with the product's use or misuse, and determine at what points during its use the warning information will be needed.
3. Competitive evaluation: Evaluate any existing warning text used for similar products to identify the strengths and weaknesses of each.
4. Prototype development: Create warning text prototypes based on existing warning research and guidelines/standards as well as the results of the hazard analysis and competitive evaluations conducted previously.
5. Empirical testing: Conduct a formal test of the warning text.
6. Field validation: Verify that the warning text is producing the desired results after it is released into the market by eliciting user feedback and tracking injury reports associated with product use and then modifying the warning text as needed.

In the sections that follow, these steps are discussed in more detail. In each section, the UCD process is described with respect to products *in general*, followed by its specific applicability to warning text. Most of the generalized products are computer related. Concrete examples are provided.

IDENTIFYING THE TARGET AUDIENCE

General Product

When developing a product, the first step in the usability engineering process is to study and understand the intended users and their tasks (Nielsen, 1992). The UCD process begins by determining who will be using the product or has a need for information about the product (IBM, 2003, Sawin et al., 2002; Vredenburg et al., 2002). User characteristics, such as education level, reading and language skills, and past experience, are important for understanding anticipated difficulties and limiting unnecessary product complexity. Once the target audience has been identified, a representative sample of users can be recruited to provide input to the design team developing the product (IBM, 2003).

Warning Text

Similarly, before designing warning text, it is important to know the characteristics of the target audience. If the target audience is the general public, the textual warnings should be designed to be understandable by persons with lower level reading abilities. For the general public, a fourth-to sixth-grade reading level is usually appropriate. If a warning's audience will be restricted to members of a specialized field (e.g., trained health professionals for a medical device), higher levels of reading ability are acceptable, and the use of professional or industry-standard terminology is recommended. For warnings whose audience might include a large percentage of non-native language users, translated text (or symbols) might be necessary. Other user characteristics, such as age, which might affect the user's ability to read small print, must also be considered. As with product development, once the target audience is defined, user profiles can be created to aid in the selection of appropriate representative participants for input in later activities.

HAZARD/TASK ANALYSIS

General Product

When designing products, a clear understanding of tasks likely to be involved with the product is important early in the design process. A task analysis should be conducted to identify user goals and information needs at different stages in the task. A task analysis also considers the environment in which the product will be used. When conducting a task analysis, Nielsen (1992) recommends observing how users interact with systems in their natural environments.

Warning Text

When designing warnings, it is important to understand how the product will be used and at what points during use the warning information is likely to be needed or seen (Frantz & Rhoades, 1993). A task (or hazard) analysis is used to identify potential sources of injury associated with product use and foreseeable misuse (Madden, chap. 45, this volume). According to basic engineering standards and Tort law, manufacturers are responsible for designing a reasonably safe product. There is a basic hierarchy when controlling hazards. The best way is to design out or eliminate the hazard. If this cannot be done, then the next best method is to guard against the hazard to separate users from the hazard. Only when the hazard cannot be effectively and practically designed out or guarded, warnings should be used as the last method of hazard control. In other words, the design team should examine the product to be sure that neither design nor guarding can be done, and if there are potential effective engineering controls, to recommend that they be used instead of a warning.

As mentioned earlier, consideration of foreseeable misuse is necessary. Norman (1988) described how products, which afford certain actions, could be used in ways never intended or foreseen by their designers. For example, people die each year of carbon monoxide poisoning from using barbeque grills in their homes during power outages and inclement weather. Use of this product to heat homes or cook indoors may not be an intended use, but it is foreseeable based on databases of fatality reports (e.g., the National Electronic Injury Surveillance System of the U.S. Consumer Product Safety Commission).

If at all possible, when conducting a hazard analysis, users should be observed interacting with the product in a realistic environment and interviewed about how they have used the product at other times. For example, warning designers for large construction equipment might visit construction sites to observe how their products are used (or misused) by actual users in natural settings. Surveillance cameras or other recorded data can also be used to assess how people use a product in the field. If field observation is not possible, it might be possible to conduct a simulated-use situation where representative users can be observed performing realistic tasks in a lab environment that has been modified to create a pseudo-natural setting. Of course, not all tasks for all products can be realistically simulated in the lab, but frequently high-fidelity simulation is not necessary to get helpful information.

There are simulations in the warning literature. For example, Wogalter, Kalsher, and Racicot (1993) created a simulated chemistry laboratory environment to examine the effects of different warning variations on compliance to take necessary precautions when handling chemicals. Moreover, advances in computer technology now allow the creation of virtual hazard settings in which user behavior can be observed. Glover and Wogalter (1997) used relatively low-fidelity virtual-reality software to simulate a coal mine so that that they could study egress behavior. Moreover, some companies have built high-fidelity simulators to train users. These technologies and the associated knowledge might be adapted for use in warning applications.

In addition to user observation, a hazard analysis should include input from product and domain experts who are highly familiar with the product, its use, and its potential hazards. In designing a set of warning labels for personal watercraft, Rhoades, Frantz, Young, and Wisniewski (2001a, 2001b) included not only experienced personal watercraft operators but also representatives from a number of personal watercraft manufacturers and the U.S. Coast Guard Office of Boating Safety as members of the team that conducted the hazard analysis. Other sources of information for a hazard analysis include accident reports and injury statistics available from sources such as the U.S. Consumer Product Safety Commission and analyzing the warning information found on similar products.

After potential hazards are identified, they should be prioritized. Some products will have only a few hazards, whereas others may have many. In general, the most important ones should be given priority in the warning. Generally, the warnings for the most important hazard should be attached to the product, be listed first, and be larger and more conspicuous than warnings for the less important hazards. Sometimes, there is not enough space on an on-product label to list all of the product's hazards. In these cases, either the label needs to be enlarged to provide additional surface space or an alternative product label design, such as a tag, should be used to add extra space on which to print the warnings (Wogalter & Young, 1994). Less important hazards may need to be placed in supplemental materials (e.g., accompanying product manual or package insert).

How does one determine the prioritization? The process generally requires input from databases regarding past injuries and judgments by domain experts and potential users. The databases can give information on severity and likelihood of injury, but of course, if it is a new product, the databases will not be very helpful unless there are comparable products on the market. Domain experts should be asked to estimate the potential likelihood and severity of injury for each hazard identified in the hazard analysis. This can be done using one of several rating scales or psychophysical measurements, such as magnitude estimation (to be described later). In making their judgments, the experts should consider the complete product life-cycle, from transporting the product, to consumers to removing the product from its packaging, to disposal at the end of its life. Next, a decision is needed on whether the hazards are open and obvious or so well known that a warning is not needed. Generally, open and obvious hazards (e.g., a knife is sharp and can cut) do not require warnings, although sometimes warnings for some hazards may be included to remind users of a known hazard. To make this determination, domain experts should once again be involved. Experts are simply asked whether each hazard is likely to be known (open and obvious) to users. Vigilante and Wogalter (1997) found it useful also to ask about each of the hazard's relative importance in considering prioritization.

In addition to domain experts, a sample of representative or prospective users should be involved. Similar questions as described earlier should be asked. The reason both domain experts and potential product users should be included is that the two groups sometimes differ in their assessments. Although domain experts know much about the product, they might also misjudge what users know and so might not accurately assess users' knowledge of product hazards. Users (or prospective users) offer a different perspective, and their judgments should carry weight in prioritizing the hazards. They are the ones most dependent on the warning's effectiveness. Keep in mind that sometimes a hazard that seems obvious to an expert might be quite obscure to some members of the consuming public.

The job of the designer is to combine the judgments from the responses of both groups to establish a prioritization. How best to accomplish this has not yet been established definitively by research (Vigilante & Wogalter, 1997), but it is probably justified to weight certain responses more heavily than others in assigning overall importance (prioritization) scores. A hazard that is severe and reasonably probable (according to domain experts) and not well known by a large proportion of users should be given a higher overall prioritization score than a hazard that is less severe, very unlikely, and already well known by users. After the scores have been determined, the warnings can be rank ordered according their levels of prioritization.

COMPETITIVE EVALUATION

General Product

The next step in the UCD process assumes that other versions of a product (competitors) are available against which to compare the strengths and weakness of the target product. This may be

accomplished by (a) asking domain experts to do a formal analysis of this type, or (b) having a sample of representative users complete tasks using one or more of the products to assess their performance and to rate their satisfaction with each product, or (c) both.

Warning Text

With regard to the design of warning text, a competitive evaluation might include the evaluation of the warning text from similar products. If available, users should be asked to make judgments about the warning text's salience (how well the warning attracts attention), clarity, believability, connoted hazard, and likelihood of complying. Because others in the industry may not have good warnings, the data involving competitors' warnings may be lacking in their ability to show warning development potential. More on this topic is given later in this section.

Another important assessment involves comprehension. Do participants understand what is being said on the label and does it provide the information needed or intended? Comprehension assessment can be performed by asking participants what the text means to them. Once their unprompted answers have been given, further probing can be done to solicit additional knowledge about the meaning of the warnings that they did not provide initially. These responses can be compared with the hazard analysis performed earlier by the domain experts to determine if users might have incomplete understanding or erroneous beliefs about product hazards. This information can be useful later when developing prototype text.

It is important to note that competitors may not be using appropriate warnings. Even if similar text is used on several competitive products that does not necessarily mean it is appropriate. That is, the standard warning text used by an industry may be defective. This derives from the fairly common "copy cat" method of writing warnings. The assumption is that other manufacturers have exercised due diligence in developing their warnings. Consider the difficulty one might have in a product liability case if the only explanation of how the warning for some product was developed was that the warning was simply copied from another product. A much better response would be to say a proper, formal assessment like that described in this chapter was used. Improper warnings on existing products are sometimes a result of not only lack of care, but also "reworking" by other entities in the company, such as marketing, who may have concerns other than safety, such as sales (see also chap. 52, Cox, this volume).

There might not be a viable competitor available to conduct a direct competitive analysis. In these cases, it is still essential to conduct some assessment at this stage. The assessment would be similar to those discussed in this section, but modified to fit the characteristics of the particular product and other available products in its class. The warning designer should be familiar with warning text for other products in the target product's general class. Ratings like those mentioned earlier could be conducted on related products to provide some idea of what may be needed in the target product's warning text.

Regardless of the presence or absence of competitive products, it is important for the warning designer to be aware of warning standards (e.g., ANSI (2002) Z535.4 for product safety labels and signs, and ANSI (2000) Z129.1 for hazardous chemicals), guidelines, and any other relevant standards for specific products, and also to have at least some knowledge of current research on warnings. The warning designer can use these materials to become familiar with the types of formatting and content that meet or exceed state-of-the-art criteria for effective warnings. Standards are minimum criteria, and it is not the case that simply meeting them is adequate by itself. Indeed, one purpose of this chapter is to describe some of the missing aspects of warning text design that are currently absent in standards.

With the information gathered from the competitive evaluation process, the warning designer can begin to understand how different types of wording and messages will fare in subsequent UCD steps.

PROTOTYPE DESIGN AND EVALUATION

Message Design

General Product. The next step in the UCD and usability engineering process is to design and evaluate prototypes. Prototyping gives the designer an opportunity to try different ideas and weigh alternatives early in the design cycle. Prototype designs should make use of knowledge gained from studying user characteristics, analyzing ratings and task performance, and performing a competitive evaluation as described in the earlier stages. Prototypes should also be based on recognized design standards/guidelines and research. General usability guidelines include giving and using feedback and reducing working memory load (Shneiderman, 1998).

Warning Text. Several chapters in this Handbook provide general guidelines for the wording of warning messages based on empirical research. Two other sources include Laughery and Wogalter (1997) and Wogalter, Conzola, and Smith-Jackson (2002). In general, a warning message should include three components: (1) a description of the hazard or hazards, (2) specific instructions on how to avoid the hazard(s), and (3) an explicit description of the consequences of failing to comply with the warning (Wogalter et al., 1987). Research has shown that explicitly worded warnings connote greater hazard (Laughery & Stanush, 1989) and are more likely to be recalled (Trommelen & Akerboom, 1999) than less explicit warnings. Methods for increasing warning explicitness include adding detailed descriptions of potential injury outcomes (Laughery & Stanush, 1989) or product injury statistics (Conzola & Wogalter, 1998) to the warning text (Table 38.1).

In a practical sense, however, it is not always possible to include complete explicit information about each hazard on the label. As described earlier, the size of the product package often limits the amount of space available for warnings. Although alternative package designs that include such surface-area-expanding features as wings, foldouts, or supplemental tags

TABLE 38.1. General Guidelines for Warning Wording and Formatting

Wording
- Be brief, using as little text as necessary to clearly convey the message.
- Use short sentences rather than long complicated ones.
- Be explicit—tell in specific terms what the hazard is, what to do or not do, and the consequences of not avoiding the hazard.
- Use short, familiar words.
- Avoid technical terms, jargon, and abbreviations, except when assured the target audience has knowledge of them.
- Use standard signal words (DANGER, WARNING, CAUTION) to convey hazard level. Other terms may be used if understandable and correspond to the hazard, e.g., DEADLY).
- Use bulleted lists to communicate points or steps.
- Use the active voice rether than passive voice.
- Use concrete rether than abstract words.
- Avoid using words or statements with multiple interpretations.

Formatting
- Use mixed case. Avoid using all upper case letters except for signal words or to emphasize individual words.
- Left-justify text.
- Use larger-sized print for longer view distances and older adults and others with reduced visual acuity.
- Consistently align and position component elements.
- Orient messages to read from left to right, top to bottom.

are possible, there are ways to design the text to convey the most meaning in the smallest amount of space. In general, it is best practice to use as little text as necessary to convey the message. Use simple vocabulary and sentence structure to increase comprehension. Use short, familiar words and avoid technical terms and jargon, except in cases where the target audience for the warning is familiar with the terminology, such as medical device warnings targeted at trained medical personnel. If a signal word is included in the warning text to attract attention, use a standard signal word (DANGER, WARNING, or CAUTION) that conveys the appropriate hazard level. In some cases, where the hazard is particularly heinous, a nonstandard signal word such as DEADLY can be used, as research has shown this term to connote a higher level of hazard than the standard signal words (e.g., Wogalter & Silver, 1995; Wogalter, Frederick, Magurno & Herrera, 1997; Wogalter, Kalsher, Frederick, Magurno & Brewster, 1998). Other characteristics regarding signal words, such as panel/print color and the alert symbol, can be found in ANSI (2002) Z535 (see also Peckham, chap. 33, this volume).

Abbreviations and acronyms should be avoided unless they are understood by the target audience (e.g., CPR, FBI). Warning text should use simple, direct sentences that contain, at most one or two subordinate clauses. Complex, indirect, passive sentences with subordinate clauses and modifying statements tend to make the material less readable, decrease comprehension, and lengthen processing time. If communicating discrete points or steps, use bulleted lists rather than complete sentences. Warning statements should be written using the active rather than the passive voice. When possible, statements should be written in the affirmative versus the negative. For example, the statement, "If you smell gas, leave area immediately" is preferred to "If you smell gas, do not hesitate before leaving area." The exception is the relatively common warning instruction where a prohibition is required (e.g., No smoking). Double and triple negatives in text statements should be avoided. Concrete, as opposed to abstract, words and phrases should be used, and words or statements that have the potential for vague or multiple interpretations (e.g., near, quickly, use adequate ventilation) should be avoided.

Similarly, statements that imply some doubt or uncertainty as to the actuality or seriousness of the hazard should be avoided. For example, a hazard statement on a bottle of aspirin that says there is a "slight risk of stomach bleeding" from using the product leaves "slight" open to interpretation of the reader. Nevertheless, it is important to give a fair assessment of the state of knowledge if it is truly unclear what the extent and probability of the risk is. Research (Freeman and Wogalter, 2002) indicates that people want to be informed about risks even if there are some doubts involved. The text describing the risk should be as unbiased and as informative as possible.

Finally, the warning text should be designed to be understandable by persons with limited reading skills. As mentioned at the outset of this chapter, warning text that is targeted at the general public should be written at or around a fourth-to sixth-grade reading level. However, electronic-graded readability indices need to be treated with some degree of skepticism. Lehto and House (1997) found that such indices had little validity in predicting the comprehensibility of warning statements. Others have also expressed their doubts about the utility of automatic reading indices (e.g., Duffy, 1985). The indices can be used to note potential problems, but the method of asking a representative sample of users what the text means is the "gold standard" for measuring comprehension.

The format of the text can also affect warning effectiveness. For ease of reading, text should be left justified. Centered text requires readers to reorient their eyes to the starting point of each line, and fully justified text adds uneven spaces between words. Text should be printed in mixed case (both upper and lower case). Use of all caps should be restricted to signal words or to emphasize particular words within the text. The font size should be selected considering the space available and the anticipated distance of the reader from the sign or label. Warnings expected to be viewed from a distance and by older adults

(a)

Primary symptoms of TSS are sudden high fever (usually 102° or more), and vomiting, diarrhea, fainting, or near fainting when standing up, dizziness or rash that looks like a sunburn. There may also be other symptoms of TSS such as aching of muscles and joints, redness of the eyes, sore throat and weakness. If you have sudden high fever and one or more of the other symptoms, remove your diaphragm and consult your physician immediately. Women with a known or suspected history of TSS should not use the diaphragm.

(b)

Primary symptoms of TSS are sudden high fever (usually 102° or more), and one or more of the following:

- vomiting
- diarrhea
- fainting, or near fainting when standing up
- dizziness
- rash that looks like a sunburn
- aching of the eyes
- redness of the eyes
- sore throat
- weakness

If you have sudden high fever and one or more of the other symptoms, remove your diaphragm and consult your physician immediately. Women with a known or suspected history of TSS should not use the diaphragm.

FIGURE 38.1. Toxic shock syndrome (TSS) symptoms in the (a) paragraph and (b) list format. From "Evaluation of List vs. Paragraph Text Format on Search Time for Warning Symptoms in a Product Manual," by M. S. Wogalter and E. F. Shaver, 2001, in A. C. Bittner, P. C. Champney, and S. J. Morrissey (Eds.), (*Advances in Occupational*) *Ergonomics and Safety*, Amsterdam: IOS Press.

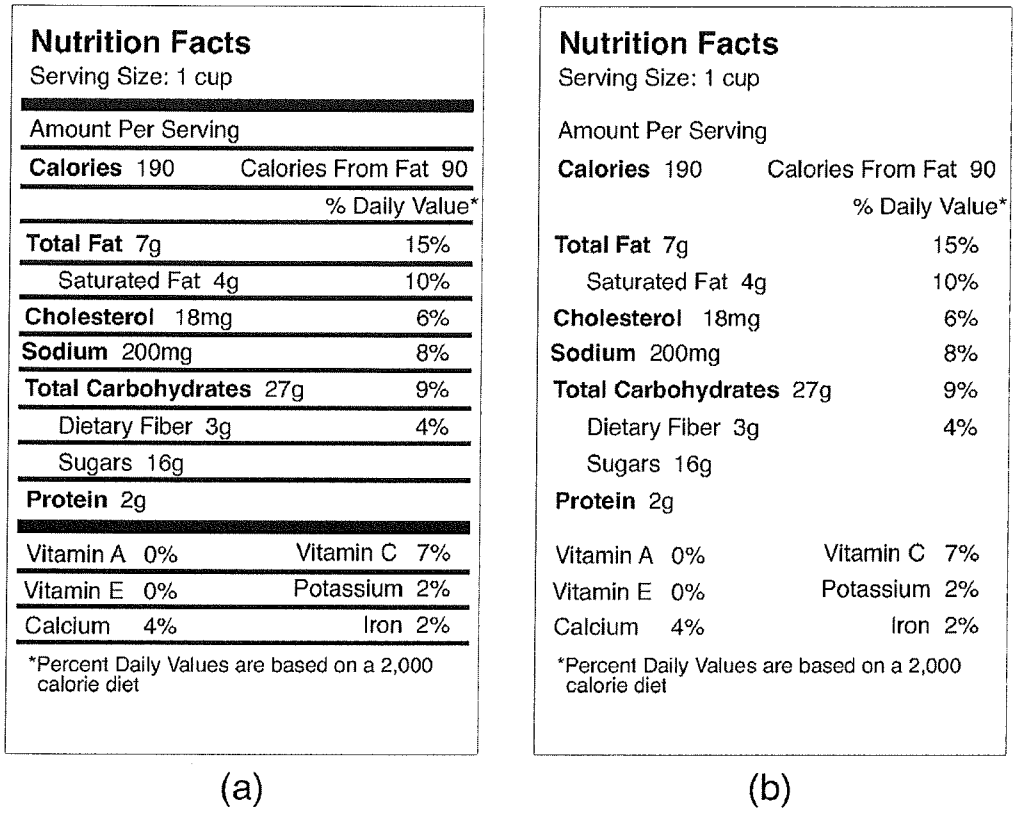
should be printed in larger sizes. Woodson, Tillman, and Tillman (1992) and ANSI (2002) Z535.2 and Z535.4 provide details on the appropriate font size to use for text that will be read from a given distance. Within the warning text, related information should be grouped. Warnings with multiple parts and subparts should be organized into sections with headings and subheadings, and the groups arranged so that they will be read from left to right and top to bottom. Multiple avoidance instructions or consequence descriptions related to a particular hazard should be presented together in the form of bulleted or numbered lists (Wogalter & Shaver, 2001; Wogalter, Shaver, & Chan, 2002). Research has shown that people are faster in examining material in a list format as opposed to a paragraph format. Examples of list and paragraph formatting are shown in Figs. 38.1 and 38.2 (Wogalter & Shaver, 2001; Wogalter, Shaver, & Chan, 2002).

Evaluating Early Prototypes

General Product. Early prototypes for products in development are evaluated heuristically (critiqued) by marketing and design experts who use formal and informal checklists to compare the design elements against existing guidelines and then make recommendations for improvement to the design team. Several different versions of the product might be created—sometimes each by a different designer working independently—so that several concepts derived from unique perspectives might be considered. As the product is refined, the best elements from the various designs are kept and incorporated into later designs. It might take several iterations before a single design is reached that satisfies all interested parties. As the product is refined, users representative of the target population are asked to evaluate the designs and provide feedback. Typically, prototype evaluations focus more on learning what is wrong with a design (qualitative) than on how much it is wrong (quantitative; Nielsen, 1992).

Warning Text. Warning text messages can be evaluated using a similar prototyping technique. Several versions of the warnings should be written that communicate the hazard information using different words. The warning designer should concentrate on communicating the necessary hazard information as clearly as possible. Once some prototype warnings have been created, subject domain experts should be asked to evaluate the prototypes and offer suggestions for improvement. Domain experts will likely have additional input about hazards and potential injury scenarios that were not considered in earliest versions of the warnings. When possible, do not use subject domain experts whose professional or financial standing might be suggestive of bias. After the evaluation and input from subject domain experts, one or more experts in warnings and risk perception should be consulted, particularly if the warning designer does not have extensive background in this field. Warnings experts are most likely to know design standards/guidelines and research and can evaluate the wording heuristically in the preevaluation stages. They may be able to point out deficiencies and opportunities for improvement that have been thus far overlooked.

Rewrite the warnings as often as necessary based on the experts' feedback. Try to combine the best features of the different versions into two to three warnings for each component hazard that satisfy the experts' requirements. Once the experts are satisfied, ask users representative of the target population to evaluate the warning text, and, if possible, the warnings should be given in context (e.g., a preliminary mockup of the product, photographs). Initially, it is instructive to get initial feedback on what they interpret from the warning(s). Then explain to them the product's purposes and the nature of all product hazards that have been identified. Instruct the participants to read each warning and identify portions of the text that they find difficult (or that they believe others may find difficult). In particular, they should be asked to point out words, phrases, or sentences that they do not understand or find ambiguous. Ask them for ideas on how to better phrase the difficult parts of warnings, and then rewrite the warning text as necessary based on their feedback. Repeat this evaluation cycle until both the representative users and the warning designer are confident that the resulting components clearly communicate the appropriate level of hazard, nature of the hazard, and consequences, and provide instructions for safe behavior. In this iterative prototyping stage of repeated redesign and retest, successive versions will not only yield improvement to the prototypes but also increase the investigator's confidence that a satisfactory end stage has been reached.



Nutrition Facts
Serving Size: 1 cup

Amount Per Serving

Calories 190	Calories From Fat 90
	% Daily Value*
Total Fat 7g	15%
Saturated Fat 4g	10%
Cholesterol 18mg	6%
Sodium 200mg	8%
Total Carbohydrates 27g	9%
Dietary Fiber 3g	4%
Sugars 16g	
Protein 2g	

Vitamin A 0%	Vitamin C 7%
Vitamin E 0%	Potassium 2%
Calcium 4%	Iron 2%

*Percent Daily Values are based on a 2,000 calorie diet

(a)

Nutrition Facts
Serving Size: 1 cup

Amount Per Serving

Calories 190	Calories From Fat 90
	% Daily Value*
Total Fat 7g	15%
Saturated Fat 4g	10%
Cholesterol 18mg	6%
Sodium 200mg	8%
Total Carbohydrates 27g	9%
Dietary Fiber 3g	4%
Sugars 16g	
Protein 2g	

Vitamin A 0%	Vitamin C 7%
Vitamin E 0%	Potassium 2%
Calcium 4%	Iron 2%

*Percent Daily Values are based on a 2,000 calorie diet

(b)

Nutrition Facts Serving Size: 1 cup (16oz), Amount per serving: **Calories** 190, Calories From Fat 90, **Total Fat** 7g (25% DV), Saturated Fat 4g (10% DV), **Cholesterol** 18mg (6% DV), **Sodium** 200mg (8% DV), **Total Carbohydrates** 27g (9% DV), Dietary Fiber 3g (4% DV), Sugar 16g, **Protein** 2g, Vitamin A (0% DV), Vitamin C (7% DV), Vitamin E (0% DV), Potassium (2% DV), Calcium (4% DV), Iron (2% DV)
Percent Daily Values (DV) are based on a 2,000 calorie diet

(c)

FIGURE 38.2. Three nutrition label formats (a) list format with horizontal lines, (b) list format without lines, and (c) paragraph format. From "List vs. Paragraph Formats on Time to Compare Nutrition Labels." by M. S. Wogalter, E. F. Shaver, and L. S. Chan, 2002, in P. T. McCabe (Ed.), *Contemporary Ergonomics* 2002 (pp. 458–462), London: Taylor & Francis.

EMPIRICAL USER TESTING

General Product

Once a fairly stable design point has been reached and the major problems have been identified and corrected, the next step in the UCD and usability-engineering process is to subject the design to more formal and comprehensive testing. For consumer products, representative users are asked to perform specific tasks that make use of product features that support key system functions. If the actual product is not available, e.g., because manufacturing is not yet completed, a limited function yet high-fidelity mockup can be used instead. The goal of user testing is to identify deeper usability problems and to determine if pre-defined usability goals, such as reduced task completion times and error rates, have been achieved.

Warning Text

Similarly, once the prototype warnings have been developed by iterative prototyping, more formal, quantitative testing of their effectiveness should be conducted. Individuals representative of the target populations should be used as test participants, including persons with lower levels of

reading ability and sensory-perceptual difficulties, such as older adults.

Any number of methods might be used at the warning text evaluation stage. Ideally, one could place the text on a product, ask a sample of representative users to perform various tasks with the product, and observe whether they did or did not comply with each warning. Afterward, users who did not comply could be asked if they noticed the warning, read it, understood it, and so forth. This type of study, which includes both performance measures and self-reports, can provide valuable information regarding both success and failure in the context of use, which generally has greater face and external validity than other methods. In designing this assessment, the warning designer should ensure that, if participants do not comply with the warnings, they do not get hurt in the process. Unfortunately, this form of assessment is time consuming and relatively difficult to carry out. Nevertheless, the results should provide assurances that the warning text is doing its intended job in promoting safety.

There are alternative procedures that can be valuable and productive. A description of a particular evaluation procedure follows. It assumes that several different component warnings, each for a specific product hazard (or set of hazards), have been created and have gone through prototype evaluations.

In this method, booklets are created such that the text of each component warning appears on a different page, ordered based on the prioritization described earlier. It is important, when evaluating warnings, that participants have some idea of the context surrounding the warning. Ideally, the actual product prototype (or a photograph) is available to provide context. Scenarios of use conditions and an indication of where a warning might be placed should be given first. Each participant is given a booklet and instructed to open it, examine the first warning, think about the context in which the warning will be placed, and complete a set of questions and ratings similar to those described in the next paragraph. A similar procedure is followed for each component warning in the booklet.

Measures that can be used to assess warning effectiveness include open-ended questions and rating scales. A knowledge test using open-ended questions is an effective way to test participants' understanding and comprehension. Four open-ended questions that might be used to assess warning text comprehension are:

1. What, in your own words, is the meaning of the warning?
2. What hazard(s) is (are) implied in the warning?
3. What should you do or not do to avoid the hazard(s) implied in the warning?
4. What consequences can result from failing to comply with the warning?

If at least 90% of the participants correctly comprehend the meaning and identify an appropriate hazard level, and there is less than 3% critical confusions (answers that convey the opposite or dangerously wrong meaning of the warning), the warning can be considered acceptable for use. These values are somewhat more rigorous (i.e., more conservative) than the ANSI (2002) Z535.3 standard for safety symbols, which specify 85% comprehension and 5% critical confusion rates for symbols presented without text. These text-comprehension levels are, however, arbitrary and are mentioned only as a rule of thum. If higher levels of comprehension and lower levels of critical confusions can be attained, it is advisable to do so. When scoring open-ended questions, two or more independent judges should evaluate the correctness of each response so that an assessment of reliability can be obtained. Unbiased, scripted oral interviews can be used in place of printed, open-ended questions and rating scales.

Rating scales should be used in conjunction with the open-ended questions to determine the extent to which the warnings communicate the intended hazards and hazard levels. For example, a warning may be well written (people comprehend the hazard implied in the warning), but users might attribute the wrong level of hazard to the warning (e.g., people may believe a dangerous product is not very hazardous). There are numerous rating scales that can be used. They may range from a 5-point scale (1 to 5) to a 101-point scale (0 to 100). Generally, the numbers go from low to high listed from left to right. Some have a neutral point and some are unidimensional (from absence of a quantity to maximum quantity). It is important to have verbal anchors tied to the numbers along the scale to help participants interpret them and to later aid the interpretation of the rating scores by readers of the eventual report. One example of a unidimensional rating scale that can be used for assessing warning text along several dimensions (e.g., hazardousness, likelihood of injury, severity of injury, believability, and motivation to comply with the warning) is shown in Fig. 38.3. The xxxxxxx can be substituted for the names of the dimensions being measured.

For warnings that fail to meet the comprehension criteria or fail to attain ratings that correspond with appropriate levels of hazard (derived from the earlier hazard analysis by subject domain experts), an error analysis should be conducted. An error analysis involves analyzing participants' responses to identify reasons why the discrepancies might have occurred. Based on these analyses, the warning text is rewritten and, preferably, retested. If, during testing or error analysis, hazards or foreseeable uses or misuses are discovered that had not previously been considered, additional warnings should be created or existing ones modified to address these issues. If rewriting the warnings ends up requiring more space than is available on the existing label, try to determine if the usable label space can be increased or a supplemental label added. Alternatively, a portion of the text containing lower priority warnings might need to be placed in accompanying supplemental material, such as a product manual or package insert. The testing process continues in this manner,

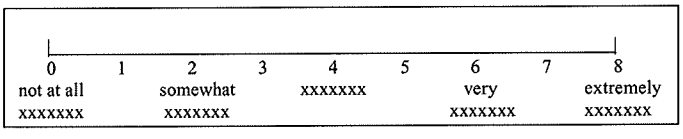


FIGURE 38.3. Rating scale for assessing warning effectiveness, where "xxxxxxx" indicates a rating dimension (e.g., important, hazardous, likely to be injured, etc.)

using new participants for each iteration, until all component warnings meet the acceptance criteria.

It is also advisable, before this stage is completed, to test the entire set of component warnings together as a set. The reason is that the component warnings will not appear in isolation in real conditions of use, and they may benefit from being presented with other warnings around them. Two types of "whole" warning conditions might be considered. One is to present all of the warnings, both the on-product labeling and any supplemental printed material (e.g., product manual) together. The other is similar but without having the product manual available. The latter condition is important because many people misplace product manuals (Wogalter, Vigilante, & Baneth, 1998).

Finally, it is sometimes valuable to ask participants if they know of other product hazards not addressed by any of the warnings. This is done simply as a basic check to verify that no hazards were overlooked during the pre-prototype and prototype design phases.

FIELD VALIDATION

General Product

The last step in the UCD process is to ensure that users are satisfied with the final product design and that it is working as expected. This is usually done via benchmark testing, customer satisfaction surveys, return and defect rate examination, and by tracking support and help center calls.

Once the new product is delivered to the marketplace, customer satisfaction surveys may be conducted to determine if users are satisfied with the new product design. Customer satisfaction surveys can be carried out in many ways. Methods include providing a customer satisfaction postcard with the product packaging or randomly contacting and surveying registered customers. The surveys should be designed to touch on each aspect of the new product that the designers feel is important to the customer's overall satisfaction. The supporting materials (e.g., user's guide, quick-start guide), packaging materials, and overall buying experience may also be addressed.

After the product has been on the market for a while, return and defect rates may be examined to determine why products are being returned. The results can then be analyzed to determine if the returns are because of actual product defects or if users cannot understand how to properly use the product. Similarly, by tracking the support and help center calls, one can determine the types of problems users are having with the product and determine whether these issues are because of an inadequate design, poor documentation, or other factors.

Warning Text

The UCD techniques for determining post-sale satisfaction and success with a new product can be used to evaluate the success of printed warnings. For example, by examining returned products, one can determine how people are misusing the product, whether the misuse resulted in personal or property damage, and whether the misuse was foreseen during the development of the product and warning text. If the misuse was foreseen and warned about, yet the product was still misused, it might be necessary to redesign the product to eliminate any affordances that led to the misuse or to redesign the warning text to make it more salient, explicit, or easier to understand. If the misuse was not foreseen, the product and warning text should be redesigned to account for the potential misuse.

Benchmark testing can also be conducted in the form of field validation. Field validation includes examining customers' use of the product in real-life situations and environments to determine if they are using the product correctly and if the warning text is providing the needed information in a timely and understandable fashion. In some cases, it may not be possible to conduct the field evaluation because of risks, privacy rights, and other reasons. The techniques used in the field validation, therefore, may need to be modified according to the available methods, given the particulars involved.

Regardless of the manner in which the field validation is conducted, the goal should be the same: to determine if the warning text is effective and, if not, how it can be modified to make it effective. Finally, the UCD process implies an iterative design cycle, where the product's design is constantly evaluated and updated as new information becomes available. Warning text is no different and should be evaluated repeatedly over time to ensure it remains relevant and effective.

PARTICIPANTS

In the development and evaluations processes described in this chapter, there were three participant groups. One was subject domain experts. These individuals may be recruited from a number of sources (e.g., the company's own engineering department or outside engineering consultants). Excluded should be individuals having a potential conflict of interest. Two to three subject domain experts may be sufficient.

Another group mentioned is experts in warnings and risk perception. Use of these individuals is important if the warning investigator does not have education, training, and past experience in the warnings area. However, if the warning investigator has on HF/E background and knows the field of warnings, the use of outside warning experts is optional.

The third group is comprised of a sample of representative users of the product for which the warning text will be targeted. They are involved in earlier (e.g., iterative prototyping) stages and later (e.g., empirical testing) stages of the project. The number of participants depends on the resources and time available. Research in UCD (e.g., Virzi, 1992) suggests that relatively few participants are needed to determine the main problems with prototypes. However, with warning design, it is extremely important to try to include certain kinds of representative persons from the target population, even when the time and resources available to complete the process are limited. If participant numbers are constrained at the earlier stages of developing prototypes, it is important to select individuals who come from the "worst-case" groups, that is, persons who are likely to be most

disadvantaged by poorly designed text. For example, if the general population is the target audience, it is probably wise to oversample individuals with lower levels of reading skill and sensory-perceptual difficulties (e.g., older adults).

The total number of users at the early stages will depend on how many iterative prototype evaluations are needed to be satisfied that one or more high-quality prototypes have been attained. A larger group of participants is usually employed in the later stages than in the earlier stages. Currently, there is no stated rule, either by standards/guidelines or research, on how many participants should be used in the empirical-testing evaluation stage. As a rule of thumb, a sample of 30 to 50 representative users who were not involved in the prototyping activities should be employed. This range in sample size is based in part on the normal curve for statistical sampling and the number of participants indicated in the ANSI (2002) Z535.3 standard for conducting symbol comprehension evaluations.

SUMMARY

In this chapter, general guidelines for evaluating and testing warning text are presented. They are based, in part, on well-accepted usability engineering and UCD principles for creating usable products. These processes have been shown to increase the ease-of-use and satisfaction with many different types of consumer and commercial products, including laptop computers (Sawin et al., 2002), welding machines (Burmester, Beu, Hackl, & Niedereder, 2002), software for children (Robertson, 1994), computer-based process information systems (Kontongiannis & Embrey, 1997), decision support systems for crop production (Parker & Sinclair, 2001), and Web designs (Cato, 2001; Garrett, 2003; Kwahk et al., 2001).

The same basic UCD principles of audience definition, task/hazard analysis, competitive evaluation, iterative prototyping, empirical user testing, and field validation, can be applied to the process of developing and evaluating textual warnings that effectively communicate information about hazards. Specific methodological recommendations have been made where possible; however, the procedures presented here are by no means rigid nor exhaustive and can be adapted or added to (or deleted), depending on the needs of a given product or situation. Use of each of the steps (and even the need to include certain steps) depends on what is already known by the warning designer and users. Some of the steps can be forestalled or lengthened, depending on the circumstances. This chapter offers more specific guidance where there have been mainly broad generalities in the development and evaluation of textual material used in warnings.

The ultimate measure of warning effectiveness is behavioral compliance. Ideally, all warnings would be tested to assure high levels of compliance with real users in real-world situations before they are put into use. However, this type of testing can be very costly, can expose participants to some degree of risk, and, in some cases, could be unethical if inadequate care is not taken.

Finally, just as the evaluation of consumer products does not end when the product is shipped, neither should the evaluation of warnings end once they are put into use. Consumer products are usually modified with each product version release based on feedback from field studies and an evaluation of user support center calls. Similarly, follow-up assessments of warning effectiveness should be performed after the product has been released. If possible, consumer feedback and reports of injuries related to product use should be examined and the warning text updated as needed.

References

American National Standards Institute. (2000). *Hazardous industrial chemicals-Precautionary labeling, Z129.1*. Washington, DC: Chemical Manufacturers Association.

American National Standards Institute. (2002). *Accredited standard on safety colors, signs, symbols, labels, and tags, Z535.1-5*. Arlington, VA: National Electrical Manufacturers Association.

Burmester, M., Beu, A., Hackl, H., & Niedereder, F. (2002). User centered design for a digital welding machine. *Behaviour and Information Technology, 21*, 321-326.

Cato, J. (2001). *User-centered web design*. Harlow, NY: Addison-Wesley.

Conzola, V. C., & Wogalter, M. S. (1998). Consumer product warnings: Effects of injury statistics on recall and subjective evaluations. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 559-563). Santa Monica, CA: Human Factors and Ergonomics Society.

Duffy, T. M. (1985). Readability formulas: What's the use? In T. M. Duffy & R. Waller (Eds.), *Designing usable texts*. Orlando, FL: Academic Press.

Frantz, J. P., & Rhoades, T. P. (1993). A task analytic approach to the temporal and spatial placement of warnings. *Human Factors, 35*, 179-730.

Frantz, J. P., Rhoades, T. P., & Lehto, M. R. (1999). Practical considerations regarding the design and evaluation of product warnings. In M. S. Wogalter, D. M. Dejoy, & K. R. Laughery (Eds.), *Warnings and risk communication* (pp. 291-311). London: Taylor and Francis.

Freeman, K., & Wogalter, M. S. (2002). On informing women of child bearing age about seat belt risk during pregnancy. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 943-946). Santa Monica, CA: Human Factors and Ergonomics Society.

Garrett, J. J. (2003). *Elements of user experience: User-centered design for the Web*. Indianapolis, IN: New Riders.

Garrity, E. J. (2001). Synthesizing user centered and designer centered IS development approaches using general systems theory. *Information Systems Frontiers, 3*, 107-121.

Glover, B. L., & Wogalter, M. S. (1997). Using a computer simulated world to study behavioral compliance with warnings: Effects of salience and gender. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1283-1287). Santa Monica, CA: Human Factors and Ergonomics Society.

IBM. (2003). *Ease of use: User centered design*. Retrieved January 20, 2003 from http://www.ibm.com/ibm/easy/eou_ext.nsf/Publish/570

International Organization for Standardization. (1999). *Human-centered design process for interactive systems, ISO 13407.* Berlin, Germany: Beuth.

Kontongiannis, T., & Embrey, D. (1997). A user-centered design approach for introducing computer-based process information system. *Applied Ergonomics, 28,* 109–119.

Kwahk, J., Smith-Jackson, T. L., & Williges, R. C. (2001). From user-centered design to senior centered design: Designing Internet health information portals. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 580–584). Santa Monica, CA: Human Factors and Ergonomics Society.

Laughery, K. R., & Stanush, J. A. (1989). Effects of warning explicitness on product perceptions. In *Proceedings of the Human Factors and Ergonomics Society 33rd Annual Meeting* (pp. 431–435). Santa Monica, CA: Human Factors and Ergonomics Society.

Laughery, K. R., & Wogalter, M. S. (1997). Risk perception and warnings. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed, pp. 1174–1197). New York: Wiley-Interscience.

Lehto, M. R., & House, T. E. (1997). Evaluation of the comprehension of hazard communication phrases by chemical workers. In *Proceedings of the 13th Triennial Conference of the International Ergonomics Association* (pp. 255–257). Tampere, Finland: International Ergonomics Association.

Nielsen, J. (1992). The usability engineering life cycle. *Computer, 25,* 12–22.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Norman, D. A., & Draper, S. W. (1986). *User-centered system design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Noyes, J. M., & Baber, C. (1999). *User-centered design of systems.* London: Springer.

Parker, C., & Sinclair, M. (2001). User-centered design does make a difference. The case of decision support systems in crop production. *Behaviour and Information Technology, 20,* 449–460.

Pradeep, H. (1998). *User-centered information design for improved software usability.* Boston: Artech House.

Rhoades, T. P., Frantz, J. P., Young, S. L., & Wisniewski, E. C. (2001a). Revisions of labeling for personal watercraft: Part 1—label development. *Injury Control and Safety Promotion, 8,* 71–81.

Rhoades, T. P., Frantz, J. P., Young, S. L., & Wisniewski, E. C. (2001b). Revisions of labeling for personal watercraft: Part 2—label evaluation. *Injury Control and Safety Promotion, 8,* 211–227.

Robertson, J. (1994). Usability and children's software: A user-centered design methodology. *Journal of Computing in Childhood Education, 5,* 257–271.

Sawin, D., Yamazaki, K., & Kumaki, A. (2002). Putting the "D" in UCD: User-centered design in the ThinkPad experience development. *International Journal of Human-Computer Interaction, 1,* 349–368.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer Interaction* (3rd ed.). Reading, MA: Addison-Wesley.

Trommelen, M., & Akerboom, S. P. (1999). Explicit warnings for child-care products (pp. 119–125). In H. J. G. Zwoga, T. Boersema, & H. C. M. Hoonhout (Eds.), *Visual Information for everyday use: Design and research perspectives.* London: Taylor & Francis.

Vigilante W. J., Jr., & Wogalter, M. S. (1997). On the prioritization of safety warnings in product manuals. *International Journal of Industrial Ergonomics, 20,* 277–285.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors, 34,* 457–468.

Vredenburg, K., Isensee, S., & Righi, C. (2002). *User-centered design: An integrated approach.* Upper Saddle River, NJ: Prentice Hall.

Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics, 33,* 231–240.

Wogalter, M. S., Conzola, V. C., & Vigilante W. J., Jr. (1999). Applying usability engineering principles to the design and testing of warning messages. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 921–925). Santa Monica, CA: Human Factors and Ergonomics Society.

Wogalter, M. S., DeJoy, D. M., & Laughery, K. R. (1999). *Warnings and Risk Communication.* London: Taylor & Francis.

Wogalter, M. S., Frederick, L. J., Magurno, A. B., & Herrera, O. L. (1997). Connoted hazard of Spanish and English warning Signal words, colors, Symbols by native Spanish language users. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association, IEA' 97, 3,* 353–355.

Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors, 29,* 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., & Racicot, B. (1993). Behavioral compliance with warnings: Effects of voice, context, and location. *Safety Science, 16,* 637–654.

Wogalter, M. S., & Shaver, E. F. (2001). Evaluation of list vs. paragraph text format on search time for warning symptoms in a product manual. In A. C. Bittner, Jr., P. C. Champney, & S. J. Morrissey (Eds.), *Advances in occupational ergonomics and safety 4* (pp. 434-438). London: Taylor and Francis.

Wogalter, M. S., Shaver, E. F., & Chan, L. S. (2002). List vs. paragraph formats on time to compare nutrition labels. In P. T. McCabe (Ed.), *Advances in ergonomics 2002,* 458–462. London: Taylor & Francis.

Wogalter, M. S., & Silver, N. C. (1995). Warning signal words: Cannoted strength and understand ability by children, elders, and non-native English speakers. *Ergonomics, 38,* 2188–2206.

Wogalter, M. S., & Vigilante, W. J., Jr. (2001). *Proposed annex to American National Standard (ANSI) Z535.4: A method for composing and testing warning messages.* Unpublished manuscript.

Wogalter, M. S., Vigilante, W. J., & Baneth, R. C. (1998). Availability of operator manuals for used consumer products. *Applied Ergonomics, 29,* 193–200.

Wogalter, M. S., & Young, S. L. (1994). Enhancing warning compliance through alternative product label designs. *Applied Ergonomics, 24,* 53–57.

Wolff, J. S., & Wogalter, M. S. (1998). Comprehension of pictorial symbols: Effects of context and test method. *Human Factors, 40,* 173–186.

Woodson, W. E., Tillman, B., & Tillman, P. (1992). *Human factors design handbook* (2nd ed.). New York: McGraw-Hill.