Applying Usability Engineering Principles To the Design and Testing of Warning Messages

Michael S. Wogalter, Vincent C. Conzola, and William J. Vigilante, Jr
Psychology Department
North Carolina State University
Raleigh, NC 27695-7801

Very little has been published on methods for creating and testing text messages for product warnings. This paper describes how the three main principles of usability engineering (prototyping, empirical user testing, and iterative design) can be applied to the design and evaluation of warning text. Pre-prototype activities that help define the warning requirements are described, followed by a discussion of guidelines for creating and testing design prototypes. Finally, a methodology for formally testing the effectiveness of warning text is presented. The procedure described should be useful to human factors practitioners who are responsible for creating or evaluating textual warnings.

INTRODUCTION

While there is a wealth of literature examining various factors that influence warning effectiveness, very little has been published on specific methods for creating and testing text messages for product warnings. This paper attempts to fill that void by providing some general guidelines for warning text development based on the principles of usability engineering.

Usability engineering is a practical and systematic process for ensuring that the needs, expectations, and limitations of users are considered during product development. The three main principles of usability engineering are prototyping, empirical user testing, and iterative design (Nielsen, 1992). These principles have been used successfully in the field of human-computer interaction to facilitate the design of computer interfaces. This paper describes how these same principles can be applied to the design of product warning text. Guidelines are given for creating preliminary versions of a warning followed by a discussion of prototyping techniques. Finally, a methodology is proposed for formally evaluating the effectiveness of the warning text. This methodology is similar to that described in Annex B to ANSI Z535.3 (1998) for evaluating safety symbols (see also Wolff & Wogalter, 1998). The methodology presented is one of many that might provide acceptable warning text. The guidelines and methods proposed here should be adapted as necessary to satisfy the requirements of a specific warning situation.

PRE-PROTOTYPE ACTIVITIES

When designing a computer interface, the first step in the usability engineering process is to study and understand the intended users and their tasks (Nielsen, 1992). User characteristics such as education level, reading and language skills, and past computer experience are key to understanding anticipated difficulties and limiting interface complexity.

Before designing warning sign or label text it is important to know the characteristics of the expected audience for the warning. Textual warnings must be understandable by people with lower level reading abilities in the target population. A warning designer can try to determine what reading level is needed, but for the general public, a grade school reading level is usually appropriate. If a warning's audience will be restricted to members of a specialized field, the use of profession or industry standard terminology is recommended. For warnings whose audience might include a large percentage of non-native speakers language translation might be necessary.

When designing computer interfaces, a clear understanding of user tasks is important early in the design process before prototypes are created. A task analysis identifies user goals and information needs at different stages in the task. A task analysis also considers the environment in which the interface will be used. An interface designed for display on a 17-inch, SVGA, color monitor is likely to be quite different than an interface designed for the 3-inch, low resolution, monochrome display of a personal digital assistant. When conducting a task analysis, Nielsen (1992) recommends visiting customer locations and observing how users interact with systems in their natural environments.

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 seen (Frantz & Rhoades, 1993). A task (or hazard) analysis is used to identify potential sources of injury associated with product use or foreseeable misuse (Wogalter, Barlow, & Murphy, 1995). According to basic engineering standards and Tort law a manufacturer is required to design out, guard against, or warn about all foreseeable hazards from the proper use or misuse of a product. If at all possible, when conducting a hazard analysis, users should be observed interacting with the product in a realistic environment. For example, warning designers for power tools might visit building construction sites to observe how their products are used (or misused) by actual users in natural settings. If field observation is not possible representative users can be observed performing realistic tasks in a lab environment that has been modified to create a pseudo-natural setting (see Wogalter, Kalsher, & Racicot, 1993). Also, advances in computer technology now allow the creation of virtual hazard settings in which user behavior can be observed (Glover & Wogalter, 1997). In addition to user observation, a hazard analysis should include input from product and domain experts who are most familiar with the product, its use, and its potential hazards.

For each hazard identified during the hazard analysis, estimate the potential likelihood and severity of injury. In doing this, injury data and product domain experts may need to be consulted. Consider the complete product life cycle - from removing the product from its packaging to disposal at the end of its life. Decide if 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 people need to be reminded of a hazard. If more than one product hazard exists prioritize the hazards according to importance, injury likelihood, and injury severity (Vigilante & Wogalter, 1997). When designing warnings for on-product labels space limitations sometimes do not allow all of the warning information to be presented on the label. Enlist product domain experts to help determine which warnings have the highest priority and need to be presented on the label. Other, less important warnings can be communicated in supplemental materials such as an operator manual or package insert. If forced to choose between two warnings of seemingly equal importance, priority should be given to the one associated with a lesser known or less obvious hazard. Keep in mind that a hazard that may seem obvious to an expert might be quite obscure to some members of the consuming public. If necessary, the available label area can be increased with alternative designs (Wogalter & Young, 1994).

DESIGNING AND EVALUATING PROTOTYPES

The next step in the usability engineering process is to design and evaluate prototypes. Prototyping gives the designer an opportunity to experiment with different ideas and weigh alternatives early in the design cycle. Prototype designs should make use of knowledge gained from studying user characteristics and analyzing tasks. Prototypes should also be based on recognized standards and guidelines for design. For computer interfaces general usability guidelines include maintaining consistency, providing feedback, and reducing short-term memory load (Shneiderman, 1998).

For warnings, Laughery and Wogalter (1997) offer guidelines for the wording of warning messages based on empirical research. In general, a warning message should describe the hazard, provide instructions on how to avoid the hazard, and describe the consequences of failing to comply with the instructions. Specific guidelines for wording and formatting warning messages are listed in Table 1.

Early prototypes of computer interfaces are often simple paper mockups that show menu options and screen layouts. They are evaluated heuristically by interface design experts who use checklists to compare the design elements against interface design guidelines and make recommendations for improvement to the design team. At the beginning, several different versions of the interface might be created – each by a different designer working independently - so that a number of unique concepts might be considered. As the interface 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 both the interface designers and the developers. As the interface is refined, users representative of the target population are brought in to evaluate the designs and provide feedback. Typically, prototype evaluations involve qualitative methods and focus more on learning what is wrong with an interface than on how much it is wrong (Nielsen, 1992).

Warning text messages can be evaluated using a similar prototyping technique. Start by writing several versions of each warning that try to communicate the same hazard information using different words. Early on, do not be overly concerned with how the warnings will appear on the product or in the product packaging.

Table 1. ***Guidelines for warning message wording and formatting***

Wording

- Use as little text as necessary to clearly convey the message.
- Use short sentences rather than long complicated sentences.
- Be explicit – tell the reader exactly what to do or not do.
- Use short, familiar words. Avoid technical terms and jargon.
- Use standard signal words (DANGER, WARNING, CAUTION) to convey hazard level.
- Avoid using abbreviations unless they have been tested on the user population.
- Use bulleted lists to communicate points or steps.
- Use the active voice rather than passive voice.
- Use concrete rather than abstract wording.
- Avoid using words or statements that might have multiple interpretations.

Format

- Use mixed case. Avoiding using all caps except for signal words.
- Left-justify text.
- Consistently position component elements.
- Orient messages to read from left to right.

Concentrate on communicating the necessary hazard information as clearly as possible. Once some prototype warnings have been created subject matter experts are recruited to evaluate the prototypes and offer suggestions for improvement. Both experts in the field of warnings and risk perception and domain experts with extensive knowledge of the environment and hazards under consideration should be included. Warnings experts are most qualified to evaluate the warnings against accepted design standards and guidelines. Domain experts will likely have additional input about hazards that were not considered in earliest versions of the warnings. In selecting experts avoid choosing individuals whose professional or financial standing might bias their opinions away from product safety and a desire to communicate product hazards.

Rewrite the warnings as often as necessary based on the expert's feedback. Try to combine the best features of the different versions into 2 or 3 warnings for each hazard that satisfy the experts requirements. Once the experts are satisfied, bring in individuals representative of the target population to evaluate the warning text. Explain the product's purpose and the nature of all product hazards that have been identified. Instruct the evaluators to read each warning and identify portions of the text that they find difficult (or which they believe others may find difficult). In particular, evaluators should be asked to point out words, phrases, or sentences they do not understand or find ambiguous. Ask for ideas on how to better phrase the difficult parts of warnings. Rewrite the warning text as necessary based on evaluator feedback. Repeat this evaluation cycle until both evaluators and designers are confident that one version of each warning clearly communicates the appropriate level of hazard and provides instructions for safe behavior.

EMPIRICAL USER TESTING

Once a fairly stable design point has been reached and the prototype's major problems have been identified and corrected the next step in the usability engineering process is to subject the design to more formal and comprehensive testing. For computer interfaces, representative users perform specific tasks that exercise portions of the interface that support key system functions. If the actual interface is not available because coding is not yet completed, a limited function but 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 task completion times and error rates have been achieved.

Similarly, once the prototype warnings have reached an acceptable level, more formal, quantitative testing of their effectiveness should be conducted before they are put into use. Individuals representative of the target population(s) should be used as test participants, including persons with lower levels of reading ability. As rule of thumb, at least 30 participants who were not involved in the prototyping activities should be used.

Measures

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 of warnings. Four open-ended questions that are used to assess warning comprehension are:

(1) What, in your own words, is the meaning of the warning?
(2) What hazard(s) is 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?

In addition, it is sometimes valuable to probe participants for other foreseeable product hazards not addressed by any of the warnings. This is done as a "sanity check" to verify that no hazards were overlooked during the pre-prototype and prototype design phases. When scoring open-ended questions two or more independent judges should evaluate the correctness of each response so that a measure of inter-rater reliability can be obtained.

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/understand the hazard implied in the warning) but people might attribute the wrong level of hazard to the warning (e.g., people may believe a dangerous product is not very hazardous). A sample rating scale that is used to assess warnings along several dimensions (e.g., dangerousness, likelihood of injury, severity of injury, believability, and motivation to comply with the warning) is shown in Figure 1. For persons with low reading abilities oral interviews can be used in place of printed open-ended questions and rating scales.

Procedure

Any number of procedures can be used to test warning effectiveness. The procedure presented here assumes that several different warnings, each for a specific product hazard (or set of hazards) have been created and have gone through prototype evaluations.

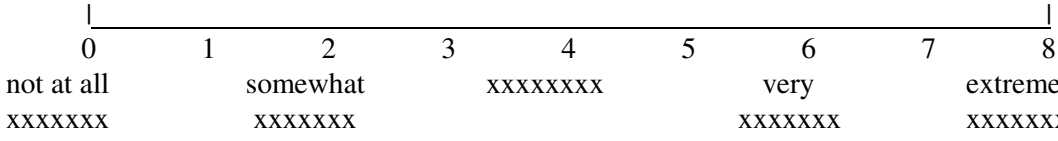
Booklets containing the warnings are created such that the text of one warning appears on each page of the booklet. It is important, when evaluating warnings, that participants have some idea of the context in which the warning is likely to be found. Ideally, sample warning labels affixed to the product or product packaging are available to provide context. If not, a photograph showing the context in which the warning might be placed can be substituted. Participant are given a booklet and instructed to open it, examine the first warning, think about the context in which the warning will be placed, and then complete the questions and ratings discussed above. A similar procedure is followed for each warning in the booklet. If at least 90% of the participants correctly comprehend a warning's meaning, identify an appropriate hazard level, and there are less than 5% critical confusions (answers that convey the opposite or dangerously wrong meaning of the warning), then that warning is acceptable for use. The 90% comprehension level is a conservative value based on ANSI standards which specify 85% comprehension and 5% critical confusion rates for symbols without text (ANSI Z535.3, 1998).

For warnings that fail to meet these criteria an error analysis should be conducted to help identify sections of the warning text that were difficult to comprehend and/or caused critical confusions. To determine how and why comprehension errors were made, participants' responses to the open ended questions are analyzed for meaning. Participant ratings are also analyzed to determine if the warnings are conveying the appropriate hazard level. Based on these analyses the warnings are rewritten and re-tested. If during testing or error analysis, hazards or foreseeable uses or misuses of the product are discovered that had not previously been considered, additional warnings must be created to address these issues. If re-writing a warning requires more space than is available on the label, a portion of the warning information will have to be moved to the supplemental material as mentioned earlier. The testing process continues in this manner, using new participants at each iteration, until all product warnings meet the acceptance criteria.

SUMMARY

In this short article we have provided some general guidelines for evaluating and testing warning text. Usability engineering principles are accepted as an effective means for creating usable human-computer interfaces. We have shown how the same basic principles of prototyping, empirical user testing, and

Figure 1. *Rating scale for assessing warning effectiveness, where xxxxxxx indicates a rating dimension (e.g., important, hazardous, likely to be injured, etc.)*



iterative design that have proven so useful for computer interfaces can also be used to produce warnings that effectively communicate the nature and degree of hazard associated with product use. 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 depending on the needs of a given product or situation.

The ultimate measure of warning effectiveness is behavioral compliance. Ideally, all product warnings would be tested for compliance with real users in real world situations before they are put in to use. However, this type of testing can be very costly, can expose participants to some degree of risk, and in some cases be considered unethical. Therefore, comprehension and behavioral intention measures such as those recommended here often must suffice.

Finally, just as the evaluation of computer interfaces does not end when the product is shipped, neither should the evaluation of warnings end once they are put into use. Computer interfaces 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 done 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

ANSI (1998). *Accredited standard on safety colors, signs, symbols, labels, and tags, Z535.1-5*. Washington, DC: National Electrical Manufacturers Association.

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

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

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

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

Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd ed.).* Reading, Massachusetts: Addison-Wesley.

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

Wogalter, M.S., Barlow,T., & Murphy, S. (1995). Compliance to owner's manual warnings: Influence of familiarity and the task-relevant placement of a supplemental directive, *Ergonomics, 38*, 1081-1091.

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

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