# Testing the Effects of Label Deterioration on the Legibility and Comprehensibility of Warnings

Nathan T. Dorris and Jerry Davis, PhD, CPE
Auburn University
Auburn, Alabama

On-product warning signs and labels face a variety of potential sources of deterioration (e.g., abrasion, ultraviolet light, and exposure to chemicals). Reviews of the available literature have produced few references regarding the effects of this deterioration on the ability of warnings to communicate. The objective of this study was to provide an initial investigation into this void in the literature. Actual warnings used on forest-harvesting equipment were photographed and rated with respect to the intensity of degradation. These real-world labels were then tested for legibility and comprehensibility using a sample of undergraduate students. The results indicate that at moderate levels of label deterioration persons can often correctly comprehend the intended safety messages. With respect to a degradation intensity rating of 3 (on a 5-point scale), the number of participants that correctly identified the conceptual meaning of the warnings (82) considerably exceeded the number that correctly identified the entire set of text messages (53).

## INTRODUCTION

Over the past three decades a voluminous body of research has developed with respect to the design and effectiveness of warning signs and labels. There are more than a few comprehensive books and literature reviews available for one seeking an overview of this research (Miller and Lehto, 1986; Wogalter et al, 1999; Rogers et al, 2000). A clear preponderance of the current literature focuses on either the design characteristics of a warning or the effectiveness of warnings as a safety measure. Several areas of research with great practical value have gone largely unexplored. Specifically, the effects of degradation to warning signs or on-product labels have largely been ignored.

In a review of the available literature for the Consumer Product Safety Commission, Kotwal and Lerner concluded that "[n]o empirical data have been found which address the effects of durability on label noticeability, legibility, or effectiveness. However, since graphic features such as color or contrast have been shown to be related to label noticeability and legibility, it can be inferred that changes in physical characteristics due to label deterioration will likely reduce the overall effectiveness of the warning label" (1995). In this review, the authors defined durability as the warning's capacity for maintaining the legibility of symbols and/or text messages as well as reasonable degree of color stability.

The ability of safety messages to reach their intended audience may be greatly reduced by substantial deterioration of the signs or labels. An illegible warning will have little value to experienced users, beyond a potential reminder, and even less for new or inexperienced users. The efficacy of any warning can be measured by its ability to communicate the intended precautionary information to its target audience. That is, the ultimate measure of warning adequacy is the comprehensibility of its message. Even if a well-crafted warning is noticed and attended to by a user, degradation can prevent the information from being relayed.

Researchers have suggested that degraded warnings have a decreased ability to communicate, however, there is no research investigating levels of degradation intensity and the respective effects on warning legibility and comprehension. The objective of this study was to provide an initial investigation into this void in the available literature.

From a practitioner's perspective the durability of on-product warnings has been addressed through recommendations in several voluntary consensus standards (e.g., SAE J115, ASAE S441.3, and ANSI Z535.4). Over the years, these standards have provided a broad range of advice. In early versions, a commonly suggested test for "weatherability" was to place potential labels on an angled (45 degrees) southern-facing surface in south Florida or central Arizona for a period of at least two years. If the sign remained legible from the intended viewing distance it was considered to be adequately durable. The most recent advice is that the label should have a reasonable expected life given the expected life of the product and the foreseeable environment of use (ANSI Z535.4-2002). These three standards also provide similar recommendations regarding the appropriate location of warnings in light of potential degradation. In fact, the ANSI standard suggests placement of warnings (when feasible) in locations that may "...provide protection from foreseeable damage, fading, or visual obstruction caused by abrasion, ultraviolet light, or substances such as lubricants, chemicals, and dirt" (ANSI Z535.4-2002).

Warning deterioration may result from environmental sources, such as color fading from exposure to ultraviolet light or physical damage from the expected abrasion and impacts associated with the routine operation of a given product or piece of equipment. Aside from the purpose-built equipment used in mining operations, forest-harvesting equipment may be used in one of the harshest environments with respect to on-product label durability. The nature of the logging process results in persistent abrasion and exposure to potential sources of degradation for warning labels. Further, because of the

hazards inherent in logging operations, much this equipment provides numerous on-product warnings. The warnings employed in the current study are actual warnings found on forest-harvesting equipment used in the Southeast. The condition of each label is as it was observed in the field during inspection of the equipment. Thus, the level of deterioration of each label is authentic and representative of what one would likely find in the field.

## METHODOLOGY

During previous research by the authors for the United States Forest Service, 1,581 on-product warnings associated with forest-harvesting equipment used in the southeastern United States were photographed and cataloged into a Microsoft Access database (Davis and Dorris, 2003). Additionally, each of these labels was rated with respect to degradation due to (a) permanent damage, such as abrasion, scratches, or fading, and (b) temporary deterioration resulting from dirt or foreign substances, such as lubricants or chemicals. Three judges used a simple 5-point ordinal scale with verbal anchors (see Table 1 below) to independently rate each label with respect to both permanent and temporary degradation. The final rating was accepted if two of the judges were in agreement and the third did not differ by more than one unit on the scale.

Table 1. Ordinal Scale for Permanent Degradation Intensity

| | |
|---|---|
| 1 | No damage, label or sign appears to be in *as-new* condition |
| 2 | Slightly worn/damaged, able to see the *virtually all* of the message(s) and/or symbol(s) (e.g., small scratches or peeling) |
| 3 | Moderately worn/damaged, able to see *most* (at least 1/2) of the message(s) and/or symbol(s) |
| 4 | Extremely worn/damaged, able to see only a portion of the message(s) and/or symbol(s) |
| 5 | Completely worn/damaged, *unable* to see *any* of the message(s) and/or symbol(s) |

This study focused exclusively on the issue of permanent degradation. Primarily because temporary degradation can often be easily corrected (i.e., washing the equipment) and the prevalence of highly degraded warnings due to temporary mechanisms was relatively low. In fact, only 30 (1.9%) warnings were rated as a level 4 or 5 on the temporary degradation scale (Davis and Dorris, 2003).

***Equipment.*** Survey booklets were developed to conduct legibility and comprehension testing of the selected warnings. Each page (8.5" x 11") contained one color photograph of a warning label presented at approximately 80% of actual size. Ten warnings were selected from the photographic database. Each sign had no significant amount of visual obstruction from dirt or debris, i.e., the label had received a rating of one on the temporary degradation scale. Two signs were chosen for each of the five permanent degradation ratings. Warning label design characteristics, such as text size, color, and use of symbols, varied among the 10 labels. However, the majority of the warnings employed design elements suggested by the ANSI Z535.4 standard (i.e., safety alert symbol, borders, as well as both text messages and safety symbols).

***Participants.*** Fifty undergraduate students from Auburn University were recruited as a sample of a naïve population. That is, they were known to have very little experience or familiarity with forest-harvesting equipment. The mean age was 21.8 years and 27 (54%) of the participants were male. Only one participant reported being colorblind. Eight respondents reported being near-sighted, however, all participants were instructed to wear any prescribed corrective lenses during the study.

***Procedure.*** The surveys asked questions designed to assess the legibility and subsequent comprehensibility of the warnings. Participants were first asked to identify the signal word used in the displayed warning from a list of the three most common signal words or the alternative, that no signal word was used. In order to assess the legibility of a warning, participants were instructed to re-write all text messages that appeared in the warning. A response was considered to be correct only if the entire message was accurately reproduced.

When a symbol was used in the warning, participants were asked to explain the meaning of the symbol. This open-ended comprehension testing procedure is consistent with that described by Lesch and McDevitt as well as outlined in ANSI Z535.3-2002 (2002). The obvious exception is that in some cases the intended meaning of the symbol was presented in the accompanying text messages of the warning being examined. The ideal method of testing symbol comprehension is independent of such text messages. However, the current procedure allows testing the comprehensibility of these symbols in a real world scenario. That is, in a manner similar to how one would view the warning in the field.

Finally, respondents were asked what actions they would take in response to the warning (i.e., the entire warning not just the individual symbols or text messages). Responses to this question demonstrated whether the participants truly understood the warning. Criteria for judging responses as correct, including acceptable variations, were identified prior to conducting the testing. The responses must convey the identified criteria or a conceptually similar response in order to be judged as correct.

## RESULTS

The number of correct and incorrect responses associated with each question and the associated levels of degradation intensity is provided in Table 2. Responses to the open-ended questions have been categorized as either correct or incorrect based on the predetermined criteria. Two warnings were text-only, i.e., they contained no symbols. Thus, a no response category was added for the symbol comprehension category. The 50 no responses associated with both degradation intensity ratings 2 and 4 are appropriate.

Table 2. Summary of Responses by Degradation Intensity

| Question and Response | | Permanent Degradation Rating | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Signal Word | Incorrect | 0 | 0 | 0 | 50 | 7 |
| | Correct | 100 | 100 | 100 | 50 | 93 |
| Symbol Comprehension | Incorrect | 14 | 3 | 28 | 50 | 100 |
| | Correct | 86 | 47 | 72 | 0 | 0 |
| | No response | 0 | 50 | 0 | 50 | 0 |
| Text Message Legibility | Incorrect | 2 | 0 | 47 | 98 | 99 |
| | Correct | 98 | 100 | 53 | 2 | 1 |
| Warning Label Comprehension | Incorrect | 2 | 5 | 18 | 97 | 100 |
| | Correct | 98 | 95 | 82 | 3 | 0 |

***Signal Words.*** Interestingly, many respondents (i.e., 47, 48, and 92 at degradation levels 3, 4, and 5 respectively) were able to identify the correct signal word even when they were not able to correctly reproduce the text messages (i.e., legibility) in the same warning label. Signal words are often presented in a larger text size (up to 50% larger) and employ a color-coding scheme consistent with the ANSI standard (e.g., safety red is associated with Danger, safety orange is associated with Warning). Further, three dissimilar words are customarily used as a signal word on warnings in the United States (i.e., Danger, Warning, and Caution). These factors may have assisted the participant when attempting to decipher a signal word that is only partially legible.

In spite of these characteristics, when this portion of the warning experiences high levels of degradation, it can also become obscured. All 50 incorrect answers for the signal word question at intensity rating 4 were associated with one particular warning. The heaviest damage was localized around the signal word panel resulting in 30 participants selecting the wrong signal word and 20 leaving the question blank.

***Symbol Comprehension.*** As with any question of comprehension, the effect of degradation was confounded with the design of that symbol. That is, a symbol may not be well understood by a particular audience even with no deterioration. Because different symbols were employed on the various warning signs, it is difficult to determine whether incorrect answers were from label degradation alone or if the inexperienced audience simply did not understand the intended message. For example, 98% of the text messages were correctly reproduced on the warnings with a level one degradation rating, however, only 86% of the responses to the symbol comprehension question were correct. Although the deterioration of the label was trivial, more of the participants than expected responded incorrectly. It should be noted, however, that many of the incorrect responses to the symbol comprehension testing questions were from incomplete or vague responses. This is a known disadvantage of using open-ended comprehension testing (Lesch and McDevitt, 2002).

***Legibility and Comprehension.*** Similar results were observed with respect to legibility of text messages and comprehension of the entire warning. Both measures had a high percentage of correct responses for labels with minor levels of degradation (i.e., ratings of 1 or 2). Alternatively, both measures had a low percentage of correct responses associated with labels with more substantial levels of degradation (i.e., ratings of 4 or 5). These results are consistent with what one might intuitively presume. That is, almost all respondents were able to correctly read and reproduce (98% and 100%) as well as understand (98% and 95%) the safety messages when the signs were in a nearly new condition. It should be noted that the only two correct responses for legibility with degradation intensity rated as 4 were associated with the warning which had the heaviest damage focused on signal word panel.

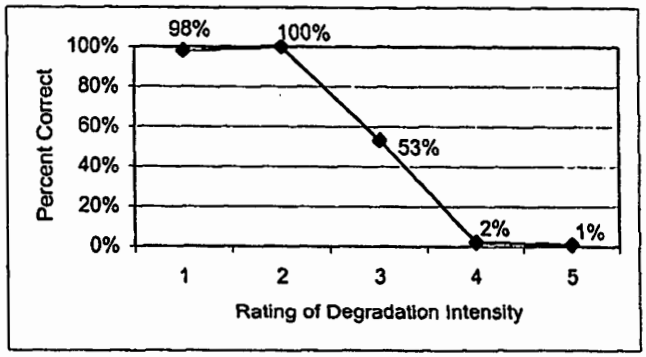


Figure 1. Text Message Legibility by Degradation Intensity

Conversely, when signs were completely deteriorated respondents were rarely able to correctly decipher the messages (i.e., 3% and 0% correct comprehension respectively at level 4 and 5 degradation). Thus, as level of degradation increases, resulting in a higher rating on the degradation scale, the legibility and comprehension levels decrease (see Figure 2 below).

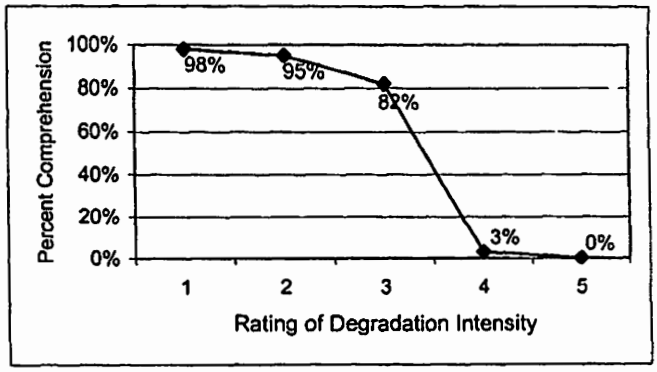


Figure 2. Warning Comprehension by Degradation Intensity

Results for legibility and comprehension of the warnings were also consistent with expectations for the moderate level of degradation (i.e., a rating of 3). Although, only 53% of the respondents were able to correctly reproduce all of the text

messages presented in these warnings, 82% were able to correctly understand the conceptual meaning of the precautionary information put forth. The participants are unfamiliar with this equipment and its associated hazards, yet they were often able to make inferences from the available information (i.e., the legible portions of the text messages and the information conveyed through symbols as well as the context in which these communications were being transmitted) to fill in the missing text and decipher the intended messages.

The authors were, however, surprised by the similarity of the results associated with the minor levels of deterioration (i.e., between signs with ratings of 1 and 2). This pattern was also apparent when comparing results from signs with more substantial levels degradation (i.e., ratings of 4 and 5). Although the level of degradation increased (e.g., from level 1 to 2) this had little effect on legibility and comprehension.

## DISCUSSION

Largely the results of this study were consistent with the authors' hypothesis. Noticeable exceptions include the similarity of legibility and comprehension rates between the minor degradation levels (i.e., 1 and 2) as well as between the substantial degradation levels (i.e., 4 and 5). Although the amount of deterioration due to abrasion (or from other modes of damage) obscured an appreciably larger portion of the sign, the respondents were able to decipher the messages correctly at comparable rates (e.g., 98% and 95% for level 1 and 2 respectively). Perhaps broader categories, such as with a 3-point scale (i.e., low, medium, and high) for rating degradation intensity would have sufficed.

The most significant finding was the apparent threshold after which the deterioration of a warning noticeably impedes its ability to deliver the intended message. A steep drop-off was observed in comprehension rates between warnings with a rating of 3 and 4 (i.e., from 82% to 3%). With further investigation more concrete recommendations can be developed for equipment manufacturers and owners/operators regarding the most appropriate time (as well as when it becomes a necessity) for replacing deteriorated on-product warnings. Additionally, warnings could be grouped by their respective degradation intensity ratings into two categories: those for which comprehension is likely to be impaired by degradation (i.e., highly degraded warnings with intensity ratings of 4 or 5) and those unlikely to be impaired (i.e., ratings of 1, 2, or 3).

The ability of this group of naïve users to correctly understand warnings in which the entire message was not legible was another important finding. The cues provided by the legible partitions of text, the legible portions of the symbols, and the context of the communication provided a sufficient framework for many of the participants to correctly decipher the intended message by making inferences. Comparison between this naïve sample and a sample of experienced users will provide a better understanding of the role of inference making and prior knowledge when attempting to read and understand degraded warnings. A greater knowledge of this type of equipment, the associated hazards, and logging operations in general (i.e., typical of an experienced population) may provide an enhanced ability to decipher messages under degraded conditions. Goldman and Wolfe note that "[a]n important issue for models of text comprehension is the contribution of prior knowledge of the text topic and structure to the formation of coherent representations. Texts are seldom completely explicit with respect to the connections among text elements so comprehenders often have to make inferences to bridge these 'gaps'" (2001). A future study with a sample of experienced equipment operators is planned.

## CONCLUSION

Nearly a quarter (23.6%) of the warnings surveyed for forest-harvesting equipment were found to be highly degraded, i.e., degradation intensity ratings of 4 or 5 (Davis and Dorris, 2003). The findings of this paper suggest that new, inexperienced users of this equipment would be unable to comprehend those warnings. Additional research to assess the effects of degradation on other measures such as noticeability, likelihood of attending to the warning, and hazard perception should be performed.

## REFERENCES

American National Standards Institute (2002). *ANSI Z535.4 Product Safety Signs and Labels*. Washington DC: National Electrical Manufacturers Association.

American National Standards Institute (2002). *ANSI Z535.3 Criteria for Safety Symbols*. Washington DC: National Electrical Manufacturers Association.

Davis, J and Dorris, NT (2003). Current Status of Warning Systems in Forest Harvesting Equipment. USDA Forest Service Research Agreement No. SRS 02-CA-11330132-087.

Goldman, SR and Wolfe, MBW (2001). International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Elsevier.

Kotwal, BM and Lerner, ND (1995). *Product Labeling Guide: Literature Review*. CPSC-C-93-1132. Washington DC: Consumer Product Safety Commission.

Lesch, MF and McDevitt, JR (2002). Methodological Issues in Testing Comprehension of Safety Symbols. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Miller, JM and Lehto, MR (1986). Instructions and Warnings. Ann Arbor, MI: Fuller Technical Publications.

Rogers, WA, Lamson, M and Rousseau, GK (2000). Warnings Research: An Integrative Perspective. *Human Factors*. 42(1) p. 102-139.

Wogalter, MS, DeJoy, DM and Laughery, KR (1999). Warnings and Risk Communication. London: Taylor and Francis.