Comprehension and Perceived Quality of Warning Pictorials

N. Clayton Silver
University of Nevada, Las Vegas
Las Vegas, NV 89154-5030

Michael S. Wogalter
North Carolina State University
Raleigh, NC 27695-7801

Blair M. Brewster
Electromark Company
Wolcott, NY 14590

Barbara L. Glover La Tondra A. Murray
North Carolina State University
Raleigh, NC 27695-7801

Cheryl A. Tillotson Tallah L. Temple
University of Nevada, Las Vegas
Las Vegas, NV 89154-5030

ABSTRACT

The present study assessed the comprehensibility and quality of warning pictorials in the presence and absence of explicit context. Context was provided by a photograph and a verbal description of an environmental scene in which the pictorial might appear. A total of 248 individuals performed a comprehension test on a randomly-assigned pictorial from each of three referent categories (Keep Out, Electrical Shock, and Do Not Dig). Following this task, 185 participants were shown five pictorials (four others plus the one they had seen) associated with each of the three referent categories and then rated and ranked them on their quality to convey the referent message effectively. Results indicated that the context manipulation enhanced comprehension for pictorials two out of three referent categories. Confidence intervals indicated that comprehension levels of all the Electrical Shock symbols would fall within the ranges specified by ISO's 67% and ANSI's 85% comprehension criteria. Three of the Do Not Dig pictorials and none of the Keep Out pictorials fell within the acceptable ISO and ANSI comprehension criteria. Statistically significant average point biserial correlations were obtained between the comprehension and quality scores for each referent category. Implications for warning pictorial test and design are discussed.

INTRODUCTION

Warning designers have increasingly made greater use of pictorials in hazard communications. The potential benefits of pictorials include being able to see them at greater distances (and in smaller sizes) than words (Jacobs, Johnston, and Cole, 1975), as a way of reaching persons unable to read the language (e.g., foreign speakers, children), and as a way of increasing the salience of the message (Young, 1991).

Although there are several criteria for the design of good pictorials, understandability (i.e., comprehension) is usually considered the most important criterion (Dewar, 1994). Without empirical evidence on comprehension rates, it is not possible to tell whether existing pictorials are adequate on this dimension. The present study examined the comprehension rates of several pictorials intended to convey the concepts of Keep Out, Electrical Shock, and Do Not Dig.

A second purpose of the study was to determine whether the presence of context facilitates icon comprehension compared to its absence. Pictorials are normally viewed in locales that assist understanding of their meaning. However, when pictorials are tested for comprehension, they are usually evaluated without any explicit context. In other words, the test lacks external validity. Without context, the viewer has to generate a mentally-derived locale for the pictorial. For example, a pictorial of a boot could be interpreted in different ways depending on where a sign with this symbol is posted. At the entrance of a shopping mall, it could mean that shoes are required (no bare feet). At a construction site, it could mean that steel-toed boots are required. In other situations, it could designate a shoe store or shoe repair facility.

The issue of combining pictorials with an appropriate context during testing is important because according to international and U.S. standards pictorials need to meet certain levels of comprehension for acceptability. The ISO (1979) minimum comprehension criterion for acceptable pictorials is 67% and the ANSI Z535.3 (1991) standard on Criteria for Safety Symbols designates an 85% comprehension rate as the minimum cutoff for acceptable pictorials. Tests conducted by Collins, Lerner, and Pierman (1982) and Laux, Mayer, and Thompson (1989) indicate that many pictorials in widespread use did not reach the 85% criterion for conveying their designated concepts.

If a pictorial does not meet acceptable levels of comprehension, then it needs to be redesigned and retested until an acceptable level of comprehension is achieved (Wolff and Wogalter, 1993). This iterative process, however, can be costly in terms of effort, time, and monetary resources (Magurno, Kohake, Wogalter, and Wolff, 1994).

One way to reduce costs involved in testing and enhancing external validity is to provide an appropriate context. This enhancement could reduce costs because fewer iterations would be required to reach a stated criterion level of acceptability. Vukelich and Whitaker (1993) found that verbal context facilitated comprehension better than no context. However, Brelsford, Wogalter, and Scoggins (1994) found that pictorial training involving the name of the concept plus additional hazard-related information (e.g., on consequences) had little effect on subsequent comprehension scores compared to the name of the concept alone. The present study examined the effect of context (presence versus absence) using a photograph of the environment where the pictorial would be posted together with a verbal scenario.

METHOD

Subjects

A total of 248 persons participated. The sample included 121 undergraduate students (69 from the University of Nevada, Las Vegas and 52 from North Carolina State University) who received extra credit for participating (mean age of 24.65, *SD* = 8.92). Additionally, 127 nonstudents at public shopping venues (77 from a flea market in Raleigh, NC and 50 from a shopping center in Las Vegas, NV) received a baseball cap or other inexpensive gift in exchange for their participation. In the shopper sample, there were 72 males and 54 females with a mean age of 43.45, *SD* = 14.42).

The ethnic/racial composition of the shoppers was: 65% Caucasian, 17% African-American, 6% Hispanic, 5% Native American, 4% multi-racial, and 4% other. One individual left ethnicity/race blank. Their highest level of education completed was: 1% only elementary school, 13% some high school, 26% high school, 26% some college or trade school, 21% college, and 13% at least some graduate school. One individual left education blank. All 77 individuals from a flea market in Raleigh, NC completed the comprehension portion of the study, however, only 14 of them rated and ranked the pictorials on quality.

Stimulus Materials

This study examined five pictorials from three referent categories: (a) Keep Out, (b) Electrical Shock, and (c) Do Not Dig. These pictorials are depicted in Figure 1. A listing of the pictorials are shown in Table 1.

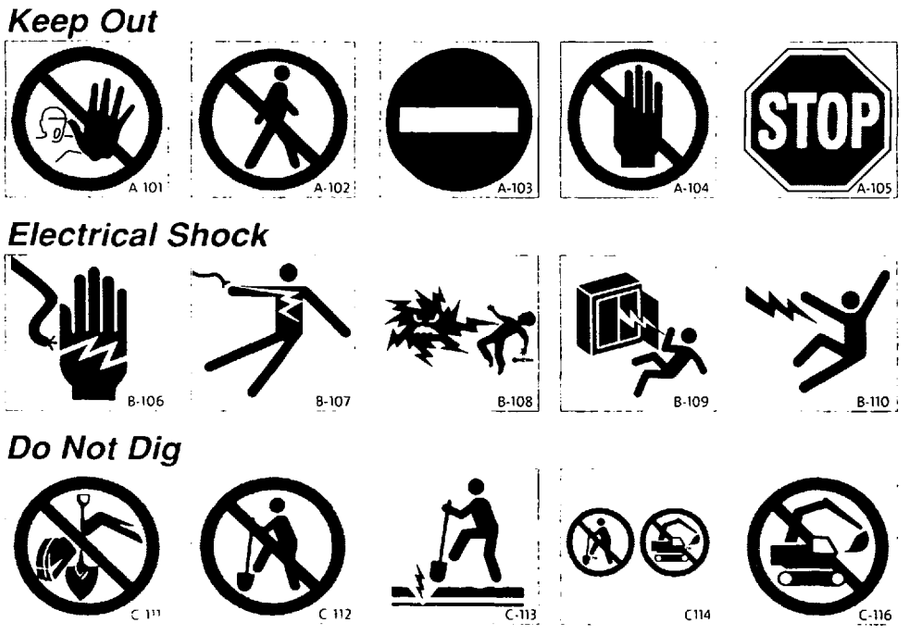
In the context-present condition, participants were shown a photograph of a location in which the pictorial might appear for each of the three referent categories. In addition, a verbal scenario was provided for each. For the Keep Out symbol, participants were shown a photograph of an electrical generator and told that a sign with the pictorial would be posted on a fence around the area. For the Electrical Shock pictorial, participants were shown a photograph of high voltage power lines and tower. They were told to assume that their wallet had been stolen recently, but their driver's license and wallet could now be seen lying inside the fenced-in area. Furthermore, they were told that their wallet could be reached if they squeezed through the locked fence which had a sign on with the pictorial. In the Do Not Dig condition, participants were shown a photograph of a residential construction site and then told that a sign with the pictorial would be placed there. All of the photographs were in color and measured 20.3 x 30.5 cm (8 x 12 inches).

Procedure

All participants received one randomly-selected pictorial from the three referent categories. In each case, the participant was to write down the meaning of the pictorial as specifically as possible. In the context-present condition, they were given an associated photograph and verbal scenario. The procedure for the context-absent condition was the same as the context-present condition except that the participants were provided only with the pictorials (no photograph or verbal scenario was given). Participants were randomly assigned to conditions.

After completing the pictorial comprehension test, participants were given all five pictorials for each referent and

Figure 1. The Three Categories of Pictorials



told the intended meaning. In addition, all participants were given the photographs and scenarios that had been provided in the context-present condition. All were asked to rate each pictorial within each set on how well it conveyed the appropriate message. A 5-point Likert-type scale with the following numerical and verbal anchors was used: (1) extremely poor quality; (2) poor quality; (3) moderate quality; (4) good quality; and (5) extremely good quality. Lastly, participants were asked to rank the pictorials within each category on how effectively they conveyed the intended message, with the easiest to understand pictorial placed in position 5, and the most difficult to understand pictorial placed in position 1. For all three parts of the study, the stimuli were randomized between and within subjects.

RESULTS

Comprehension

The responses were scored as either incorrect (0) or correct (1) by one of the senior authors. For Keep Out, any answer synonymous with the referent (e.g., do not enter and stay away) was counted as correct. For Electrical Shock, any response indicating that an individual would be injured by electricity (e.g., shocked or burned) was counted as correct. For Do Not Dig, a correct response had to be synonymous with the referent (e.g., do not excavate or do not use a shovel). An additional individual, independent of the laboratory team, re-scored the comprehension responses in order to establish inter-rater reliability. Of the 248 responses for each category, the two scorers agreed on 99.2% for the Keep Out condition, 100.0% for the Electrical Shock condition, and 99.6% for the Do Not Dig condition.

Separate 2 (context) x 5 (pictorial) between-subjects analyses of variance (ANOVAs) were performed on each of the three categories of pictorials using comprehension as the dependent variable. All subsequent tests were performed using the Student Newman-Keuls multiple range procedure. Means, standard deviations, and confidence intervals for each pictorial are shown in Table 1. Confidence intervals were used to compare the comprehension rates from the present study to the ISO 67% and ANSI 85% comprehension criteria. Table 1 also shows the means and standard deviations for the pictorials, both overall, and as a function of context effects. Only effects that are statistically significant are described.

Keep Out. Table 1 shows that the "Figure Walking with Slash" (A102) produced the highest comprehension. The ANOVA showed a significant context x pictorial interaction, $F(4, 238) = 4.03$, $p < .005$. Tests of simple effects revealed that the pictorial "STOP" (A105) was more understandable in context than without context, $p < .0001$. With context, "STOP" was more understandable than either the "Shout" (A101) or the "Hand with Slash" (A104), $ps < .03$. Without context, the "Figure Walking with Slash" (A102) was more understandable than "STOP" (A105), $p < .04$.

Electrical Danger. The ANOVA showed a significant main effect of context, $F(1, 238) = 7.36$, $p < .01$. For this set of pictorials, the presence of context ($M = .914$) produced higher comprehension than its absence ($M = .797$).

Do Not Dig. There was a significant main effect of pictorial, $F(4, 238) = 21.26$, $p < .0001$. Highest comprehension rates were found for the "Man & Shovel with Slash" (C112) and the "Shovel & Backhoe with Slash" (C111). The "Backhoe with Slash" (C116) and the "Man & Shovel & Shock in Ground" (C113) were less well understood than the other three pictorials, $ps < .0001$.

Comparison to the Standards

In general, pictorial acceptability has been judged with respect to the proportion or percentage correct relative to a criterion specified in standards. Because samples that might be taken from a population vary, it may be better to judge acceptability using confidence intervals which give upper and lower limits of where the population mean lies. The confidence intervals indicate that 95% of the time this interval will fall around the population mean (in this case, comprehension). Thus, if the ISO or ANSI comprehension standards fall within the confidence interval, then there is no statistically significant difference between the standards and the population mean. If the ISO or ANSI criteria are outside the interval, however, then the population mean would be significantly higher or lower than the standards.

As shown in Table 1, none of the Keep Out pictorials reached the 67% ISO or 85% ANSI comprehension standards with respect to the means and confidence intervals. All of the Electrical Shock pictorials met the 67% ISO and 85% ANSI comprehension standards according to the confidence intervals. For the Do Not Dig pictorials, the "Shovel & Backhoe with Slash" (C111), the "Man & Shovel with Slash" (C112) and 2-symbol pictorial (C114) produced comprehension rates that exceeded the ISO and ANSI criteria.

Quality Ratings

Because it is possible that exposure to a particular pictorial in the preceding comprehension task could influence the subsequent quality evaluations. In order to evaluate this potential carryover effect, the variable pictorials within a particular category was used as a factor. Separate 2 (context) x 5 (pictorials within a particular category) x 5 (pictorial) mixed-model ANOVAs, with the latter factor a repeated measures variable, were performed for each of the three pictorial categories. No significant carryover effects were found in the analyses. The quality rating means and standard deviations are presented in Table 1.

Keep Out. The ANOVA showed a significant main effect of pictorials, $F(4, 696) = 71.39$, $p < .0001$. Quality was highest for "The Shout" (A101), followed by the "Stop" sign (A105). Quality was lowest for the "Do Not Enter" DOT sign (A103). All pairwise comparisons were significant, $ps < .05$.

Electrical Shock. There was a significant main effect of pictorials, $F(4, 700) = 26.52$, $p < .0001$. The "Hand Shock" (B106) and the "Man Shock (B107) were perceived to be lower quality than the other pictorials, $ps < .01$. Moreover, "Mr. Ouch" (B108) was perceived as higher quality than the "Man & Bolt" (B110), $p < .02$. "Mr. Ouch" and "Shock in a Box" (B109) were perceived to be the highest quality and did not significantly differ.

Do Not Dig. There was a significant effect of pictorials, $F(4, 696) = 54.34$, $p < .0001$. The "Man & Shovel & Shock in Ground" (C-113) and the 2-Symbol pictorial (C-114) were

Table 1. Means and Standard Deviations of Comprehension and Message Quality Measures for Each Pictorial

Pictorials	Comprehension Proportion Correct Mean	SD	95% Confidence Interval	Context Mean	SD	No Context Mean	SD	Message Quality† Rating Mean	SD	Ranking Mean	SD
Keep Out											
A101 - The "Shout" (ISO)	.29	.46	.15 ≤ u ≤ .41	.24	.44	.33	.48	3.4	1.3	4.0	1.3
A102 - Figure Walking with Slash (generic)	.43	.50	.29 ≤ u ≤ .56	.42	.50	.44	.51	2.6	1.3	3.0	1.3
A103 - Do Not Enter (generic)	.30	.46	.17 ≤ u ≤ .41	.37	.49	.22	.42	1.6	1.1	1.5	1.1
A104 - Hand with Slash (generic)	.28	.45	.15 ≤ u ≤ .40	.26	.45	.30	.47	2.9	1.2	3.1	0.9
A105 - STOP sign (generic)	.36	.49	.21 ≤ u ≤ .50	.63	.49	.05	.22	3.1	1.4	3.3	1.3
Electrical Shock											
B106 - Hand Shock (Westinghouse/FMC)	.78	.41	.65 ≤ u ≤ .89	.82	.39	.69	.48	3.1	1.2	2.3	1.3
B107 - Man Shock (Westinghouse/FMC)	.94	.24	.87 ≤ u ≤ 1.0	1.00	.00	.87	.34	3.5	1.1	2.8	1.1
B108 - Mr. Ouch (NEMA)	.79	.41	.68 ≤ u ≤ .90	.90	.30	.72	.46	4.0	1.2	3.7	1.6
B109 - Shock in a Box (Electromark)	.88	.33	.79 ≤ u ≤ .96	.96	.19	.79	.41	3.8	0.9	3.2	1.3
B110 - Man & Bolt (Electromark)	.91	.29	.82 ≤ u ≤ .99	.90	.31	.92	.28	3.7	1.0	2.9	1.4
Do Not Dig (All are from Electromark except C111)											
C111 - Shovel & Backhoe with Slash (APWA)	.82	.41	.71 ≤ u ≤ .92	.86	.36	.77	.43	2.8	1.3	2.3	1.5
C112 - Man & Shovel with Slash	.87	.34	.77 ≤ u ≤ .95	.93	.26	.79	.41	3.8	1.1	3.2	1.3
C113 - Man & Shovel & Shock in Ground	.37	.49	.24 ≤ u ≤ .50	.33	.48	.39	.50	4.0	1.0	3.1	1.5
C114 - 2-Symbol pictorial (C112 & C116)	.77	.43	.64 < u ≤ .88	.80	.41	.71	.47	4.0	1.1	4.0	1.0
C116 - Backhoe with Slash	.25	.44	.12 ≤ u ≤ .37	.28	.44	.26	.45	3.1	1.3	2.4	1.1

† Quality ratings and rankings ranged from 1 to 5 with high numbers indicating better evaluations.

rated highest quality. The "Backhoe with Slash" (C-116) and "Shovel and Backhoe with Slash" (C-111) were rated significantly lower than the other three pictorials, $ps < .0001$.

Rankings and Intercorrelations

There was significant agreement between the quality ratings and rankings (Kendall's Tau ranged from .60 to 1.0, $ps < .05$) for each referent category. Examining the agreement among the comprehension, rating, and ranking scores, Friedman's test was significant only for the Electrical Shock category, $p < .007$. This indicates that there was high agreement in the ordering of pictorials across the three measures. An average point biserial correlation (Silver and Dunlap, 1987; Silver and Hollingsworth, 1989) between comprehension and ratings across the three hazard categories was statistically significant, $r = .21, p < .0001$. Within each category, the average point biserial correlations were: for the Keep Out and Electrical Shock conditions, $rs = .24, ps < .002$, and for the Do Not Dig condition, $r = .17, p < .03$.

DISCUSSION

Context facilitated comprehension for some pictorials and not others. In general, all of the Electrical Shock pictorials were better understood in the presence of context. One of the Keep Out pictorials benefited from context, but none of the Do Not Dig pictorials was significantly aided by context.

The comprehension rate was very high for the Electrical Shock pictorials, and the addition of context facilitated it further. Without context, some of the incorrect answers included: one will be struck by lightning or that a man has been overcome by a dangerous "force." Although there is some rudimentary understanding of danger in these wrong answers, the addition of context produced a more detailed understanding of the actual hazard.

In the Keep Out condition, many of the responses in the no context condition for "STOP" were indeed "STOP." Stop, however, is not necessarily synonymous with Keep Out. This is especially true when driving. In this case, STOP means to bring the vehicle to a stand still, look both ways, and proceed. Without context, the referent Keep Out was rarely mentioned. Only with context did the comprehension rate for STOP —meaning Keep Out—increase significantly.

It is clear that context sometimes helps, and sometimes does not. Why might context not help? One possibility is that the certain types of context are not adequate to cue the proper locale. Further research is necessary to delineate what kinds of context are most appropriate.

The ISO 67% and ANSI 85% comprehension criteria are benchmarks to compare empirically-derived data. Considering only the percentage correct scores, all five of the Electrical Shock pictorials met the ISO criterion, however, only three met the ANSI criterion. Within the Do Not Dig category, three of the pictorials fulfilled the ISO criterion, whereas only one satisfied the ANSI criterion. None of the Keep Out pictorials met the criteria specified by the standards. A similar, though not identical picture is revealed by the confidence intervals. The 95% confidence intervals

indicate that all of the Electrical Shock pictorials and three of the Do Not Dig pictorials met both criteria. Thus two Do Not Dig pictorials met the ANSI criterion using confidence-interval analysis—an outcome that was not apparent using the absolute percentage-correct criterion.

Both standards set arbitrary levels of comprehension for acceptability. It may be difficult to create a pictorial for some highly complex concepts that people will readily understand. Keep Out is one of those cases. This notwithstanding, it might be better to use a pictorial that fails to meet the arbitrary acceptability levels when coupled with a verbal warning than to not use any pictorial whatsoever (as long as it does not produce confusion). Moreover, pictorials should be improved to maximize comprehension. Consider that evaluation of a safety-related pictorial is found to be understood by 86% of those tested. This is higher than either the ISO or ANSI criterion, and so can be considered acceptable. However, if that pictorial can be improved, then safety concerns should dictate further enhancement.

In all three hazard categories, there were message quality differences across pictorials. For the Keep Out pictorials, all had mean ratings from poor to moderate quality. Although "The Shout" was perceived as the highest quality, it was next to lowest in comprehension ratings. This pictorial also happens to be a double negative. That is, the figure is signaling to keep out, but with the slash middle, this icon technically denotes that it is safe to enter. This response, however, was not given by any participant. Nevertheless, this symbol might be easier to understand without the slash.

All of the Electrical Shock pictorials received high quality ratings. "Mr. Ouch" was perceived as conveying the message most effectively among the five. One reason for this finding is that "Mr. Ouch" was initially developed (NEMA, 1982) to provide hazard information to children. Possibly, because of its regular use over the years, it has become known as illustrating electrical shock hazards. Other research by the present authors suggests that the "Shock in the Box"— a pictorial not significantly different from Mr. Ouch—better describes where the electrical hazard comes from than Mr. Ouch. This characteristic may be useful in some applications.

All of the Do Not Dig pictorials ranged from moderate to good quality. These results were surprising because whereas the "Man & Shovel & Shock in Ground" had the highest perceived quality, its mean comprehension rate was beneath both standards. Conversely, the "Shovel & Backhoe with Slash" had the lowest perceived message quality, but when examining the confidence intervals, had comprehension rates that fulfilled both standards. Although the 2-symbol pictorial ("Backhoe with Slash" and "Man & Shovel with Slash") were rated higher than either of the icons individually, the comprehension rate was lower than one of its component icons ("Man & Shovel with Slash"). This drop in comprehension might be due to the inclusion of the less understandable "Backhoe with Slash." Therefore, it appears that an "averaging" effect occurred in comprehension, but not for the quality evaluations. Additional work would seem warranted on issues related to the use of multiple pictorials.

There was a substantial rank order correlation between the ranking and rating data. This indicates that these procedures may be used interchangeably. However, it should be noted that in this study, the ranking always followed the rating, thus potentially producing carryover effects.

Although the average point biserial correlation was statistically significant between comprehension rates and quality of the message, it only accounted for about 4% of the total variance. Therefore, although rating and ranking are easier and less costly than evaluating comprehension rates, it did not predict comprehension rates well. Zwaga (1989) suggests that taking the median of the estimates of the percentage of individuals that would understand a pictorial predicts comprehension scores. Hence, this or some other predictive method (e.g., see Brugger, 1994) might reduce the costs involved in testing pictorial understandability.

ACKNOWLEDGEMENTS

The authors would like to thank Ann W. Carter, Jason G. Daurity, Michael A. Ettefagh, James A. Keen, Jane Jackson, Julie A. Swindell, and William J. Vigilante for their assistance in data collection. We thank Amy B. Magurno for organizing many of the procedures and operations.

REFERENCES

ANSI Z535.1-5 (1991). *American National Standard for Safety Signs and Colors*. Washington, DC: National Electrical Manufacturers Assn.

Brelsford, J.W., Wogalter, M.S., and Scoggins, J.A. (1994). Enhancing comprehension of safety-related pictorials. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 836-840). Santa Monica, CA: Human Factors and Ergonomics Society.

Brugger, C. (1994). Public information symbols: A comparison of ISO testing procedures. Proceedings of Public Graphics (pp. 26.1-26.10) The Netherlands: University of Utrecht.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Dewar, R. (1994). Design and evaluation of graphic symbols. *Proceedings of Public Graphics* (pp. 24.1-24.18). The Netherlands: Univ. of Utrecht.

Collins, B.L., Lerner, N.D., and Pierman, B.C. (1982). *Symbols for industrial safety*. (Technical Report NBSIR 82-2485). Washington, DC: U.S. Department of Commerce.

International Organisation for Standardisation (1979). *ISO 7001: Public Information Symbols-Index, Survey and Compilation of the Single Sheets*. Geneva: Author.

Jacobs, R.J., Johnston, A.W., and Cole, B.L. (1975). The visibility of alphabetic and symbolic traffic signs. *Australian Road Research, 5*, 68-86.

Laux, L., Mayer, D.L., and Thompson, N.B. (1989). Usefulness of symbols and pictorials to communicate hazard information. *Proceedings of Interface 89* (pp. 79-83). Santa Monica, CA: Human Factors Society.

Magurno, A., Kohake, J., Wogalter, M.S., and Wolff, J.S. (1994). Iterative test and development of pharmaceutical pictorials. *Proceedings of the International Ergonomics Association, 4*, 144-147.

National Electrical Manufacturers Association (1982). *Meet Mister Ouch*. Publication No. 260. Washington, DC: Author.

National Safety Council (1994). *Accident facts*. Itasca, IL: Author.

Silver, N.C., and Dunlap, W.P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology, 72*, 146-148.

Silver, N.C., and Hollingsworth, S.C. (1989). A FORTRAN 77 program for averaging correlation coefficients. *Behavior Research Methods, Instruments, & Computers, 21*, 647-650.

Vukelich, M., and Whitaker, L.A. (1993). The effects of context on the comprehension of graphic symbols. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 511-515). Santa Monica, CA: Human Factors and Ergonomics Society.

Wolff, J.S., and Wogalter, M.S. (1993). Test and development of pharmaceutical pictorials. *Proceedings of the Interface 93: The Eighth Symposium on Human Factors and Industrial Design in Consumer Products* (pp. 187-192). Santa Monica, CA: Human Factors Society.

Young, S. (1991). Increasing the noticeability of warnings: Effects of pictorial, color, signal icon, and border. *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 580-584). Santa Monica, CA: Human Factors Society.

Zwaga, H. J. (1989). Comprehensibility estimates of public information symbols: Their validity and use. *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 979-983). Santa Monica, CA: Human Factors Society.