Evaluating Hazard Symbols for the Globally Harmonized System (GHS) for Hazard Communication

Robert G. Hesse, Department of Cognitive Science, Rensselaer Polytechnic Institute
Nicholas H. Steele, Department of Cognitive Science, Rensselaer Polytechnic Institute
Michael J. Kalsher, Department of Cognitive Science, Rensselaer Polytechnic Institute
Claudia Mont'Alvao, Art & Design Department, PUC-Rio, Rio de Janeiro

ABSTRACT

Sweeping globalization has resulted in unparalleled economic growth, including increased international trade of hazardous chemicals. Fundamental differences between nations trading these materials, including language, literacy rates, cultural values, and technical and governmental infrastructures has created an urgent need for a common system of risk communication to reduce the occurrence of deaths and serious injuries that result from unintended chemical exposures. To accomplish this goal, the United Nations (UN) created the Globally Harmonized System for the Classification and Labeling of Chemicals (GHS) in 1992. Unfortunately, there was no requirement for testing of the GHS labeling components, including pictograms intended to depict specific hazards, before their deployment. In Experiment 1, twenty GHS hazard pictograms were subjected to comprehension testing in two non-student samples from the U.S. and Brazil, respectively. In Experiment 2, alternatives for five of the GHS pictograms that were least well understood were created and then re-tested for comprehension. Several of the new pictograms outperformed their "original" GHS counterparts in terms of comprehension and participant preference. Overall, the results of testing showed that only a small portion of the original GHS hazard pictograms reached acceptable levels of comprehension. Therefore, additional systematic work is needed to develop GHS pictogram alternatives that effectively convey safety hazards to a global audience.

INTRODUCTION

Sweeping globalization over the past several decades can be viewed as a double-edge sword. On the one hand, it has given rise to unparalleled economic growth and opportunity as products and services are increasingly freely traded among many different countries throughout the world. On the other hand, the many positive contributions of globalization have been offset by problems arising from fundamental differences between nations involved in international trade including language, literacy rates, laws and cultural values, and technical and governmental infrastructures. This has been particularly true of international trade involving potentially hazardous chemicals—products that solve important practical problems when used as intended, but pose significant risks to people's health and safety when they are used in ways not intended by their manufacturers. One particularly important difference between countries engaged in international trade is the availability (or lack thereof) of regulatory agencies focused on health and safety issues. While the U.S. and other developed nations have sophisticated networks of regulatory agencies designed to prevent its citizens from accidental exposure to hazardous chemicals, third world companies frequently do not (Laughery, 2006). Thus, there is an urgent need for a common system of risk communication that effectively conveys chemical hazards to a wide range of people from many different countries in order to reduce the occurrence of deaths and serious injuries worldwide that result from chemical exposures.

The purpose of the current study was to evaluate comprehension of set of GHS pictograms and a set of newly developed alternative candidate pictograms that may be more effective at conveying certain hazards. The fact that graphical pictograms are a critical component of any risk communication system, and certainly one intended for application to a diverse global audience, underscores the importance of this research. This paper reports the results of two experiments. In Experiment 1, twenty GHS hazard pictograms were comprehension tested with two non-student populations from the U.S. and Brazil using the comprehension estimation procedure (Brugger, 1994; Zwaga, 1989). In Experiment 2, an overlapping set of pictograms containing some of the pictograms used in Experiment 1 and several newly developed candidates were tested using the open-ended comprehension testing procedure (e.g., ANSI Z535.3, 2007). The methodology used in Experiment 2 also incorporated recommendations from the ILO's most recent comprehension testing recommendations (see Annex 6, 2009).

EXPERIMENT 1

Method

Participants. There were a total of 312 non-student participants; 155 were males and 157 were females. Of these, 225 (111 males and 114 females) comprised the U.S. sample and 87 (46 men and 41 women) comprised the Brazilian sample. The average age of participants was 40.6 (*S.D.=14.7*) and 41.3 (*S.D.=17.7*) in the U.S. and Brazilian samples, respectively. The average familiarity of participants with the GHS system was 0.37 (*S.D.=0.53*) and 0.16 (*S.D.=0.45*) in the U.S. and Brazilian samples, respectively. The scale anchors were 0 = Not at all Familiar, 1 = Somewhat Familiar, and 2 = Very Familiar.

Procedure. After providing informed consent, participants were asked to evaluate each of twenty GHS pictograms using the comprehension estimation procedure. Participants were provided with the context in which each pictogram would likely be seen (e.g., see Laughery, 2006) and then asked to estimate the percentage of people in their respective countries would comprehend each pictogram's intended meaning.

Results

Table 1 provides a summary of the comprehension estimates for each of the twenty GHS pictograms evaluated. For each pictogram, the table reports the means and standard errors for the U.S. and Brazilian samples, respectively. Pictograms that received the highest comprehension estimates were ones generally intended to depict relatively concrete hazards (e.g., flammability), although the ratings varied significantly between the U.S. and Brazilian samples. Pictograms intended to depict relatively abstract concepts (e.g., reproductive and carcinogenic hazards) tended to receive the lowest comprehension estimates. According to criteria outlined in the American National Standard Institute's Criteria for Safety Symbols (ANSI Z535.3, 2007), symbols and pictograms are considered acceptable if 85% of the study participants are able to understand its meaning with no more than 5% critical confusions. When compared against this standard, only three of the GHS hazard pictograms can be judged acceptable, and only for the U.S. sample: the flammability symbol (*M*=91.6, *S.E.*=0.91), the acute toxicity symbol (*M*=86.6, *S.E.*=1.35), and the marine pollutant symbol (*M*=85.0, *S.E.*=1.55). Even if the 85% criterion is relaxed, as has been recommended for studies employing the comprehension estimate procedure, only the environmental hazard pictogram (for the U.S. sample) can be judged as acceptable (*M*=65.5, *S.E.*=2.04). According to B2.4 of the ANSI Z535.3 document, the comprehension estimation procedure has a 20% margin of error (ANSI, 2007); therefore, scores of 65% or better are most likely to meet the 85% criterion of open-ended testing.

A mixed-model analysis of variance (ANOVA) was performed on the data. The repeated-measures variable was Type of Pictogram and the between-subjects variable was Nationality. Mauchly's test of sphericity was significant, $X^2(189)=1142.17$, $p<.01$, and so the Greenhouse Geisser correction was applied. There was a significant Pictogram x Nationality interaction, $F(12.93, 4008.31) = 19.43$, $p<.01$. Post-hoc tests were carried out to examine differences in comprehension estimates between the U.S. and Brazilian samples. Significant differences were found for pictograms depicting the following hazards: the flammability hazard, the acute toxicity hazard (skull-and-crossbones), the corrosive hazard, the carcinogen hazard, the environmental hazard, the acute hazard, the explosives hazard, the acute toxicity, the organic peroxide hazard, the flammable solid hazard, the pyrophoric liquid hazard, the marine pollutant hazard, and the chronic hazard (*ps*<.05). Differences in the samples' comprehension estimates for the remaining seven pictograms were not significant (*p*s>.05).

Discussion

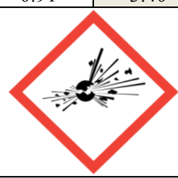
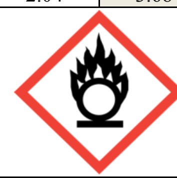
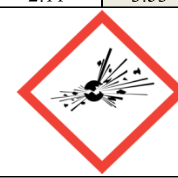
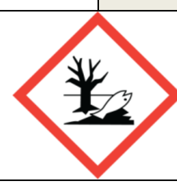
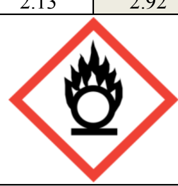
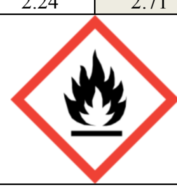
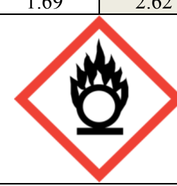
The major finding of Experiment 1 was that a majority of GHS pictograms were not well understood by the study participants. There were clear differences in comprehension estimates between the samples, but these should be interpreted carefully, and in light of a consideration of specific similarities and differences. A comparison of the U.S. and Brazilian samples indicated a number of similarities such as mean age of the participants and roughly equal proportions of males and females. However, the two samples differed in ways that may help to explain, at least in part, the observed differences in their comprehension estimates. Specifically, participants in the U.S. sample rated their familiarity with the GHS system significantly higher than their Brazilian counterparts (*p<.01*). And as a group, participants in the U.S. sample had attained a significantly higher level of education than their Brazilian counterparts (*p<.01*).

These results indicate an urgent need to address deficiencies in the current GHS symbols. In the present study, only four pictograms met the ANSI Z535.3 criteria, or relaxed criteria, for correct comprehension and only one of these met the 85% comprehension criteria in both the U.S. and Brazilian samples. One school of thought has advocated for more and better training to achieve the goal of a shared understanding of the GHS symbols (e.g., see Lesch, 2008A; Lesch, 2008B). However, it is not clear where the resources would come from to pay for this training. Another possibility is to develop and iteratively test new pictograms that more clearly convey their intended meanings, which was one purpose of Experiment 2. A second purpose was to confirm the main findings of Experiment 1 by using open-ended comprehension testing instead of the comprehension estimation procedure.

EXPERIMENT 2

In Experiment 2, an overlapping set of pictograms containing some of the original GHS pictograms used in Experiment 1 and several newly developed candidates were tested using the open-ended comprehension testing procedure (e.g., ANSI Z535.3, 2007). We elected to use open-ended

Table 1. GHS symbol comprehension estimation results. Results are ordered from highest to lowest comprehension scores (from left to right and from top to bottom).

Symbol											
Intended Meaning		Flammables		Acute Toxicity		Marine Pollutant		Environmental Hazard		Corrosive	
Nationality		U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil
Comprehension Scores	Mean	91.6	65.5	86.6	53.0	85.0	43.7	65.4	46.7	56.7	34.3
	St. Error	0.91	3.40	1.35	4.19	1.55	3.75	2.04	3.68	2.11	3.53
Symbol											
Intended Meaning		Explosives		Acute toxicity		Flammable Liquids		Organic Peroxide		Explosive Divisions	
Nationality		U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil
Comprehension Scores	Mean	55.5	58.6	51.7	34.7	44.9	42.6	41.4	27.2	35.7	21.5
	St. Error	2.05	3.36	2.27	3.35	2.31	3.66	2.33	3.13	2.06	2.74
Symbol											
Intended Meaning		Flammable Solids		Acute Hazard		Chronic Hazard		Compressed Gas		Carcinogens	
Nationality		U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil
Comprehension Scores	Mean	35.1	27.4	29.8	19.1	27.9	17.4	25.0	21.5	24.3	31.0
	St. Error	2.13	2.92	2.24	2.71	2.04	2.77	1.69	2.62	1.75	3.02
Symbol											
Intended Meaning		Oxidizing Chemical		Flammable Gases		Pyrophoric Liquid		Oxidizing Liquid/Solid		Reproductive Hazard	
Nationality		U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil	U.S	Brazil
Comprehension Scores	Mean	23.6	19.8	22.5	28.6	19.3	27.1	18.9	22.4	9.66	14.5
	St. Error	1.54	2.56	1.76	3.46	1.67	3.02	1.58	2.85	1.21	2.76

testing as opposed to the comprehension estimation procedure used in Experiment 1, for purposes of reliability, at least for the overlapping pictograms.

Method

Participants. There were a total of 109 participants (68 males and 41 females). Of these, 85 were students and 24 were non-students. The average age of participants was 25.0 (*S.D.=12.3*). The majority of participants were either attending college or were college graduates. The average familiarity of the participants with the GHS system was 0.49 out of 2 (S.D.=0.70). The scale anchors were 0 = Not at all Familiar, 1 = Somewhat Familiar and 2 = Very Familiar. Only five participants were not natives of the United States.

Experimental Stimuli. Participants viewed a set of pictograms that contained nine of the pictograms used in Experiment 1, plus five alternatives that were created by the research team to depict each of the following hazards: an explosive hazard, an environmental toxicity hazard, a mutagen hazard, a reproductive toxicity, and a respiratory toxicity. All of the stimuli evaluated are presented in Table 2.

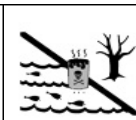
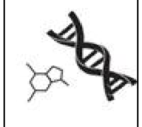
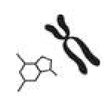
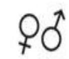
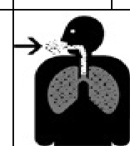
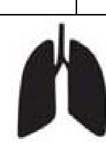
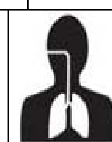
Procedure. After providing informed consent and basic demographic information, participants completed the open-ended comprehension section of the survey. Participants viewed each of the pictograms in one of three different random orderings. For each pictogram, participants were provided a written description of the context in which it would be found (Laughery, 2006) and then were asked to describe its meaning and indicate the action they should take.

For the alternative symbol testing section of the survey, participants viewed several existing GHS pictograms and a number of alternative candidates developed for this study. They were then asked to choose the pictogram they believed best depicted the intended hazard. The section containing the open-ended comprehension questions always appeared before the symbol design preference section.

Results

Table 2 provides a summary of the open-ended comprehension results for each of the nine existing GHS pictograms and five newly developed alternatives. Two judges used both a liberal and a strict criterion to evaluate comprehension. A response was scored as correct according to the liberal criterion if it was partially correct, but was insufficient to ensure adequate safe behavior. A response was judged as correct according to the strict criterion only if it was identical or fully consistent with the intended meaning. The liberal and strict criterions were defined per GHS Annex 6 (ILO Annex 6, 2009). The judges discussed any discrepancies in ratings until agreement was achieved. The rate of critical confusions for each pictogram was also calculated and is included in Table 2. The results for the alternative design preference test are presented in Table 3. The frequency of selection of each choice was determined for each of the design alternatives and subsequently reported as preference percentages.

Table 3. Hazard symbol design preference results.

Environmental Hazard					
Percent of Respondents	**76.2**	22.9	0.92		
Mutagen Hazard					
Percent of Respondents	**46.8**	34.9	11.9	6.42	
Reproductive Hazard					
Percent of Respondents	**69.7**	23.9	5.51	0.92	0.00
Respiratory Hazard					
Percent of Respondents	**87.2**	4.59	4.58	3.67	

In general, comprehension results for the existing GHS pictograms paralleled those of Experiment 1. Specifically, the pictograms intended to depict a corrosive hazard, flammability, and acute toxicity received more than 85% correct responses according to the liberal criterion (ANSI Z535.3, 2007). These pictograms were also among the five pictograms that received the highest comprehension estimates based on the estimation procedure of Experiment 1. As in Experiment 1, pictograms depicting relatively abstract hazards, including hazards associated with compressed gas, oxidizing agents, and the hazards encompassed by the "starman" symbol were among the least well-understood pictograms tested in Experiment 2. Although the group of participants surveyed in Experiment 2 contained a relatively high percentage of students, the patterns of results for the two experiments were strikingly similar.

According to the strict criterion, the only pictogram to reach the 85% correct response criterion was the one depicting a corrosive hazard. None of the pictograms tested using the open-ended comprehension protocol exceeded the maximum of 5% critical confusions (ANSI Z535.3, 2007). However, two pictograms—the newly developed candidate intended to depict a mutagen hazard and the original GHS

Table 2. Open-ended comprehension study of GHS symbols (Original) and new alternative warning symbol designs (Alternative).

Symbol							
	Original	Original	Original	Original	Original	Original	Original
Intended Meaning	Corrosive	Flammable	Explosive	Environmental	Toxic	Danger	Reproductive Hazard
% Correct (Liberal Criterion)	94.5	92.7	72.5	79.8	89.0	84.4	20.2
% Correct (Strict Criterion)	90.8	83.5	59.6	68.8	82.6	79.8	15.6
% Critical Confusion	0.92	0.92	0.00	0.92	0.92	0.00	0.00
Symbol							
	Original	Original	Alternative	Alternative	Alternative	Alternative	Alternative
Intended Meaning	Oxidizer	Compressed Gas	Explosive	Environmental	Respiratory Hazard	Reproductive Hazard	Mutagen
% Correct (Liberal Criterion)	20.2	20.2	72.5	68.8	82.6	53.2	39.4
% Correct by (Strict Criterion)	5.50	15.6	61.5	56.9	66.1	36.7	27.5
% Critical Confusion	0.00	3.67	0.92	1.83	0.92	0.92	3.67

pictogram intended to depict a pressurized gas hazard—each produced a critical confusion rate of 3.67%, which approaches the maximum recommended by ANSI Z535.3.

It is noteworthy that all three of the alternative designs for the "starman" pictogram outperformed the original. The GHS actually uses "starman" to depict several different hazards, including reproductive hazard, mutagenic hazard, and respiratory hazard. Still, none of the "starman" alternative designs achieved the 85% correct comprehension criterion, and so iterative re-design is required. Candidates created as alternatives to existing GHS pictograms intended to depict two other hazards—an environmental hazard and an explosive hazard—did not perform better than their "original" counterparts. Overall, these results show that even though some of the alternative pictograms debuted in this study outperformed their GHS counterparts, the new alternatives must be improved to achieve minimum acceptable comprehension criteria or replaced with ones that do.

The results of the symbol design preference section of this experiment generally support the results of open-ended comprehension testing by showing that the "starman" pictogram is never preferred over alternative designs. Interestingly, the pictogram designed as an alternative to the existing environmental hazard pictogram was preferred by a large margin, even though it was less well understood according to the comprehension results.

REFERENCES

American National Standard Institute. (2007). *ANSI Z535.3-2007: Criteria for safety symbols*. Washington, DC: National Electrical Manufacturers Association.

Brugger, C. (1994). Public information symbols: A comparison of ISO testing procedures. Proceedings of Public Graphies. 26.1-26.11.

ILO. "Annex 6." (2009). Globally Harmonized System. *Proceedings of Comprehension Testing Methodology*, Geneva. 3rd ed. New York: United Nations, 431-42.

Laughery, K.R. (2006). Safety communications: Warnings. *Applied Ergonomics*, *37*, 467-478.

Lesch, M.F. (2008A). A comparison of two training methods for improving warning symbol comprehension. *AppFd Ergonomics*, *39*, 135-143.

Lesch, M.F. (2008B). Warning symbols as reminders of hazards: Impact of training. *Accident Analysis and Prevention*, *40*, 1005-1012.

Zwaga, H.J. (1989). Comprehensibility estimates of public information symbols: Their validity and use. *Proceedings of the Human Factors and Ergonomics Society*, *33*, 979-982.