International Journal of Occupational Safety and Ergonomics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tose20

The Relevance of Error Analysis in Graphical Symbols Evaluation

Dominic Paul T. Piamonte[a]
[a] Luleå University of Technology, Sweden
Published online: 08 Jan 2015.

INTERNATIONAL JOURNAL OF
OCCUPATIONAL SAFETY AND ERGONOMICS 1999, VOL. 5, NO. 4, 513–528

The Relevance of Error Analysis in Graphical Symbols Evaluation

Dominic Paul T. Piamonte

Luleå University of Technology, Sweden

In an increasing number of modern tools and devices, small graphical symbols appear simultaneously in sets as parts of the human-machine interfaces. The presence of each symbol can influence the other's recognisability and correct association to its intended referents. Thus, aside from correct associations, it is equally important to perform certain error analysis of the wrong answers, misses, confusions, and even lack of answers. This research aimed to show how such error analyses could be valuable in evaluating graphical symbols especially across potentially different user groups. The study tested 3 sets of icons representing 7 videophone functions. The methods involved parameters such as hits, confusions, missing values, and misses. The association tests showed similar hit rates of most symbols across the majority of the participant groups. However, exploring the error patterns helped detect differences in the graphical symbols' performances between participant groups, which otherwise seemed to have similar levels of recognition. These are very valuable not only in determining the symbols to be retained, replaced, or re-designed, but also in formulating instructions and other aids in learning to use new products faster and more satisfactorily.

symbols videophone icons empirical testing methods

1. INTRODUCTION

1.1. Advent of Graphical Symbols

Under the current globalisation, it is an increasing challenge for ergonomists to ensure the usability, comfort, and safety of new technologies

Acknowledgements to Mr. Martin Böcker for his invaluable support in acquiring the videophone symbols used in this study and in sharing his thoughts and expertise throughout, and to KFB (Swedish Communications Board) for the financial and other logistical assistance.

Correspondence and requests for reprints should be sent to Dominic Paul T. Piamonte, Industrial Ergonomics Division, Human Work Sciences Department, Luleå University of Technology, 971 87 Luleå, Sweden. E-mail: <Paul.Piamonte@arb.luth.se>.

and their products across varied user groups. Graphical symbols (icons and pictograms) play an important part in achieving this goal. From their traditional use of appearing singly or in isolation to denote locations or services, they have started to appear simultaneously in sets, as actual parts of the human-machine interfaces of modern devices. They are now also used to present the concepts, ideas or objects, and functions of the technology and products where they are used. When properly designed, tested, and taught, graphical symbols can help introduce new technologies and products to different user groups across the world regardless of culture or language groups.

1.2. Methods in Evaluating Graphical Symbols

As in all products designed for public use, graphical symbols need to be evaluated which among them are best suited across all intended user groups. The test methods to be employed are often dependent on the kind of study being pursued. Different methods have been used in evaluating symbols such as icons and pictograms (Magyar, 1990; Nolan, 1989; Vora, Helander, Swede, & Wilson, 1991; Webb, Sorenson, & Lyons, 1989). If symbols are intended as parts of interfaces of devices for international use or for standardisation, different tests are usually needed. For example, the International Organization for Standardization (ISO) has come out with ISO 9186 (ISO, 1989). It is a six-stage procedure for the development and testing of public information symbols (Zwaga, 1989). Its major portions are the comprehensibility judgement tests, comprehension test with a suggested comprehension level of at least 66%, and matching tests. The American National Standards Institute (ANSI) likewise has the ANSI Z535.3 (ANSI, 1991) describing the methods of evaluating the comprehensibility of graphical symbols. It utilises four categories, namely correct answer, wrong answer, critical confusion, and no answer. It further recommends a comprehension level of at least 85%. On the other hand, the International Telecommunications Union (ITU) has Recommendation F.910. Recommendation F.910 endorses a symbol testing procedure composed of four parts. It involves the determination of need for new symbols, the creation and evaluation of the new designs, and the selection and approval (ITU, 1995).

Another important aspect of symbols evaluation is the need to explain and test the functions (represented by the symbols) in manners reflecting typical user scenarios (Tudor, 1994). As mentioned earlier, numerous modern computer-based devices contain small graphical symbols simultaneously appearing as groups as part of the general control interface. The most appropriate way to evaluate the symbols in such cases is to present each candidate a set of symbols against one target referent at a time. For example, in a typical user scenario a prospective user intends to use a device or equipment and is confronted with controls represented by a set of graphical symbols. He or she then has to choose which among these symbols correspond to his or her desired function. One such method was developed by Böcker (1993) with the Human Factors Technical Committee of the European Telecommunications Standards Institute (ETSI). They evaluated videotelephony symbols using a multiple index approach (MIA), which can also be used in testing symbols for other commercial products. The MIA tested several sets of candidate videophone symbols by presenting each set to the participant with only one referent at a time. This approach enabled the tester to collect seven indices that could help in the final selection of symbols. The indices were hit rates, false alarm rates, missing values, subjective certainty and suitability, symbol and symbol set preferences (Böcker, 1993).

1.3. Importance of Analysing Errors

The hit rate in MIA reflected the levels of correct associations of the symbols to their intended referents, and was the main index of performance. However, it has been emphasised that hit rates alone cannot be the sole basis of determining how well a symbol is understood especially if it appears together with other symbols. Analysing the patterns or characteristics of errors such as misses, false alarms, or confusions with other symbols, and even missing values is equally important (Böcker, 1993; ETSI, 1993). Misses are instances of selecting the wrong symbol in the context of a particular function being sought. Their patterns of distribution can give some knowledge if the referents themselves are conceptually clear to the users. False alarms are instances of a symbol being wrongly selected under different non-corresponding functions or referents. Missing values are instances wherein no response or no answers were given. A hit or correct association is just one of the possible outcomes in symbol testing. Misses and other forms of errors are others that need to be closely studied as well. These considerations become more valuable when faced with potentially different user groups.

This paper aimed to show how such analyses of errors, that is, their distribution pattern and other characteristics, could be valuable in evaluating graphical symbols especially across potentially different user groups. The study was part of an international project in evaluating telecommunication icons using multiple inter-related test parameters. Some of these are the errors or the so-called non-hit parameters, which are the foci of this paper. As stated earlier, graphical symbols always have the potential of widespread usability among different user groups, but they still need to be tested when the groups in question are as diverse as in Asia. This paper focused on some countries in Southeast Asia—a region characterised by an immense diversity in language and culture. Modern products and the symbols they contain are often designed and tested in the west but targeted for worldwide use. Thus, a region such as the Southeast presents a big challenge to the designers of such products to make them usable and thus, commercially viable as well. It is then essential to evaluate products and their interfaces with these potential users in mind. It is hypothesised that simple recognition tests eliciting hit rates may fail to exhibit differences on how symbols are understood by different user groups. Other parameters, as those just discussed, would be utilised to detect other possible differences in symbol recognition by different user groups. In the design of the study, the MIA by Böcker (1993) as well as Tudor's (1994) comments and recommendations were greatly considered.

2. METHODS

2.1. Participants

Two hundred and forty university students and employees and professionals (127 males and 113 females) from small and large companies from five Asian countries participated in this study. These countries were Indonesia, Malaysia, the Philippines, Thailand, and Sri Lanka. There were 48 participants per country with equal numbers of students and employees, all of whom were computer-literate or had at least experienced using computers or computer-related products for the past 2–5 years. At the time of the study, none of the participants had ever experienced using a videophone. The mean age was 25.1 years (SD = 6.1), ranging from 17 to 52 years old. These participants came from the major cities of these countries and the tests were conducted in their schools, offices, and homes.

2.2. Materials

The stimuli were three sets of symbols representing seven referents or functions of a videophone for a total of 21 symbols (Figure 1). The referents were camera, document camera, handsfree, microphone, self-view, still picture, and videophone. Both referents and symbols were based on the ETSI study headed by Böcker (1993). The sets were grouped based on the outcome of the ETSI tests. Set 1 had the best performance based on the MIA tests (and was recommended by ETSI as the standard), followed by Set 3, and Set 2 was last. These three sets

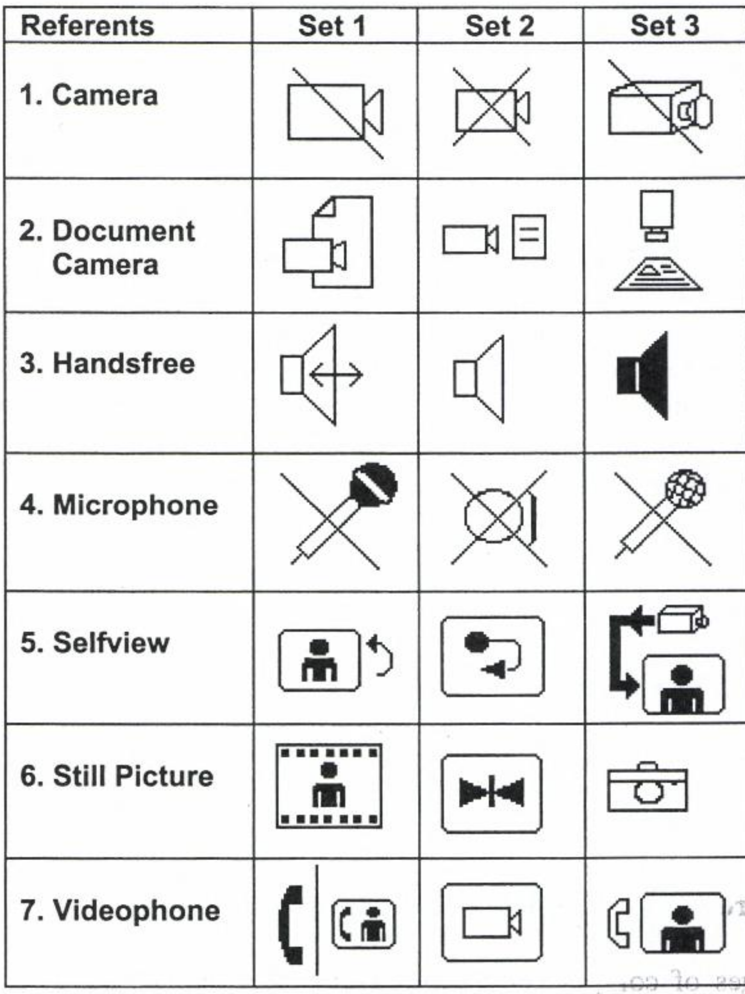

Figure 1. The videophone referents and the three sets of symbols used in the study (Böcker, 1993).

were chosen also on the basis of their symbols' general attributes such as concreteness or abstraction, another factor being studied by the project. The questionnaires were divided into three parts: symbol identification and subjective certainty tests (Part I), symbol association and subjective certainty tests (Part II), and symbol and set preferences (Parts III and IV). Three versions of the questionnaires with different orders of the symbols and sets were generated and used in the study. This paper will deal with the results of Part II.

2.3. Procedure

The tests were done in small groups in each country in schools, homes, or offices. They lasted for about 45–60 min. The participants were randomly given one of the three versions of the questionnaires. The versions differed only in symbol and referent order. An illustration of a videophone was shown and its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the tests. Emphasis was given on avoiding skips in trials in order to get back to them later. The order of the tests was also strictly followed, that is, Part I was followed by Part II, then Parts III and IV.

In the cued response test (Part II), the participants first read a referent and its description. Then they were asked to select, by putting a mark such as a circle or an x, one symbol from a set of seven symbols they thought best represented the referent in question. Each page contained one referent description and one set of symbols. Subjective certainties for their answers using 7-point rating scales (from *very certain* to *very uncertain*) were likewise done. There were seven videophone referents tested on three sets of symbols rendering a total of 21 trials.

3. RESULTS AND DISCUSSION

3.1. Correct Associations and Subjective Certainties

Percentages of correct associations (hit rates) and subjective certainties in the cued-response tests are shown in Table 1. Correct association resulted when a participant correctly matched the referent in question to its right symbol. Correct responses were labelled as 1 (hits) and wrong responses as 0 (misses). Missing data or no responses were initially treated as 0 also during this stage of the analysis. Thus, the response data were binomial percentages and arcsine transformations were warranted to ensure that assumptions in normality were met. Such transformations of percentages eliminate participants as a variable leaving only the independent variables and lowering the degrees of freedom. A repeated measures factorial design (multifactor ANOVA) was then used to determine significant differences in the hit rates. Games-Howell post hoc tests were done to analyse significant main effects.

The hypotheses tested in this part were that hit rates would differ between sets and between countries. The data in Table 1 showed that between the three symbol sets (main effect) and except for the handsfree symbols, Set 1 symbols had significantly higher hit rates than Sets 2 and 3 in the seven referents tested ($F(2, 30)$, $p < .05$). These results were expected partly because Sets 1 and 3 symbols were more concrete and representative of the referents they represented. With country as another main effect, Indonesia, Malaysia, the Philippines, and Thailand had comparably similar levels of hit rates. Only Sri Lanka showed significantly lower hit rates compared to the other countries in the symbols for camera, document camera, and handsfree ($F(4, 15)$, $p < .05$)—results that would be discussed in detail later. Regarding the certainty ratings, they were generally higher for Sets 1 and 3 compared to Set 2. The certainty scores of the symbols for microphone and videophone were either very high (Sets 1 and 3) or very low (Set 2). However, Kruskal-Wallis tests showed only a few significant differences between countries regarding their certainty ratings (Table 2).

Overall, when ISO 9186 (ISO, 1989) and ANSI Z535.3 (ANSI, 1991) comprehension levels were considered through hit rates, most of the tested videophone symbols performed poorly. Combining country results, only 7 of the 21 symbols reached ISO's required comprehension level of at least 66%. These were Set 1 symbols for camera (71%), document camera (86%), microphone (84%), and videophone (90%); Set 2 symbol for document camera (68%); and Set 3 symbols for microphone (83%) and videophone (81%). If ANSI Z535.3 was the basis, only two of the symbols would have met its 85% comprehension level (Set 1 symbols for document camera, 86%, and videophone, 90%). Thus, by looking merely at the hit rates, the following could be stated. Very few of the symbols performed well regardless of country, and most of these were from Set 1. The symbols most poorly understood were basically abstract in their general attributes (Set 2). Moreover, when comparing countries,

TABLE 1. Hit Rates in the Cued-Response Tests and Subjective Certainties Among the Three Sets of Graphical Symbols Representing Seven Videophone Functions ($n = 48$ per Country)

HITS	Camera			Document Camera			Handsfree			Microphone			Selfview			Still Picture			Videophone		
Countries	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Indonesia	73	54	69	92	79	52	40	42	35	83	21	77	60	17	65	44	25	27	88	35	85
Malaysia	71	63	65	92	77	73	38	27	38	88	33	85	73	21	63	77	42	35	94	17	88
The Philippines	83	63	71	81	69	69	35	31	33	90	40	85	81	31	79	67	25	46	94	31	79
Thailand	83	63	60	88	79	77	56	46	48	90	35	90	60	31	69	50	52	35	92	33	77
Sri Lanka	42	42	54	75	35	46	19	15	25	67	29	75	40	21	40	38	48	8	81	13	75
Means	71	57	64	86	68	64	38	33	36	84	32	83	63	25	64	55	39	31	90	26	81

CERTAINTIES	Camera			Document Camera			Handsfree			Microphone			Selfview			Still Picture			Videophone		
Countries	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Indonesia	4.2	3.8	4.7	4.6	4.3	4.1	4.1	3.1	4.0	5.3	3.7	5.1	4.0	2.6	4.4	4.3	3.9	4.4	5.7	3.2	5.1
Malaysia	4.5	4.2	4.6	4.6	4.4	4.3	3.7	3.6	3.7	5.8	3.2	5.2	4.3	3.5	4.1	4.6	4.1	3.8	5.9	3.1	6.1
The Philippines	4.4	4.3	4.6	4.2	4.6	4.8	4.1	3.1	3.8	6.1	3.3	5.3	3.8	3.7	4.5	4.4	3.9	4.6	6.2	3.0	5.9
Thailand	5.1	4.4	5.0	4.7	4.6	4.6	4.1	3.4	3.5	6.0	3.7	5.5	4.6	3.1	4.9	3.8	3.9	4.3	5.8	3.3	5.1
Sri Lanka	5.3	4.4	5.6	4.2	4.4	4.7	4.6	4.1	4.3	5.1	4.2	4.7	4.8	4.0	4.2	5.1	4.5	4.6	5.6	3.8	4.8
Means	4.9	4.6	5.5	4.7	4.9	5.1	4.3	3.9	4.5	5.9	4.0	5.8	4.5	3.8	5.0	4.6	4.5	4.9	6.0	3.7	6.0

TABLE 2. Kruskal-Wallis Tests for Subjective Certainties for Each Symbol (n = 48 per Country, df = 4, $p < .05$)

Symbol Sets	Chi-Square	Asymptotic Significance
Set 1		
Camera	12.12	.017
Document Camera	1.51	.826
Handsfree	4.95	.292
Microphone	6.96	.138
Selfview	8.15	.086
Still Picture	12.87	.012
Videophone	1.03	.906
Set 2		
Camera	3.57	.468
Document Camera	6.71	.348
Handsfree	7.26	.292
Microphone	5.08	.279
Selfview	16.51	.002
Still Picture	3.95	.413
Videophone	3.78	.437
Set 3		
Camera	10.35	.035
Document Camera	3.87	.424
Handsfree	4.01	.405
Microphone	2.10	.718
Selfview	3.25	.517
Still Picture	5.83	.212
Videophone	7.47	.113

only Sri Lanka seemed to differ significantly compared to the other countries. Notwithstanding that some other cultural differences also exist and except for Sri Lanka, the four other countries lie fairly close to each other, with standards of living and education quite similar as well. This can thus partially explain the results but admittedly will need more studies to be confirmed. However, by analysing the errors or non-hit parameters of the test procedures, more differences among the symbols and countries were discovered as discussed further in the paper.

3.2. Confusion Matrices and Misses

When symbols appear simultaneously or as a set and the main task is to pick out which among these symbols correspond to a desired function or object, there are other possible outcomes aside from a hit. These are the so-called non-hit outcomes or parameters such as misses, confusions, and missing values (failure to respond or no answer). The wrong answers during the cued response tests were analysed using confusion matrices (Tables 4 to 8). In this particular part of the study, tabulations were made on the rates of confusions—the rate wherein each symbol was selected in the context of a different referent. Confusions can be classified into two types: symmetric and asymmetric (Nolan, 1989). The former occurs when participants chose symbol x when presented with referent y, and symbol y when presented with referent x. Symmetric confusions usually suggest visual or conceptual similarities. These were exemplified by the selfview and still picture (visual similarity), and handsfree and microphone symbols of Set 1 (conceptual similarity). Another case of symmetric confusion due to visual similarities would be that of the camera and document camera symbols of Set 3. These symmetric confusions were present in all countries. Thus, the problem may not be culturally-linked but a design problem—that the symbols concerned were visually too similar. Moreover, according to basic ergonomic design principles, these symbols would then be prime considerations for re-design or replacement to make them more distinct from the other ones.

Asymmetric confusions occur when participants simply chose the wrong symbol for a given referent. For example, Table 4 showed that

TABLE 3. Confusion Matrices, Indonesia (n = 48)

Referent Presented	Symbol Selected																				
	Set 1							Set 2							Set 3						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Camera	—	10			2	4	4	—	15		2	2	6	21	—				13	15	4
2. Document Camera	4	—				2		6	—		4	4		6	10	—	2		13	13	8
3. Handsfree	4		—	27	2		21	4		—	15	2	15	8	8	2	—	13	6		33
4. Microphone			15	—			2			38	—	17	6	4	2		15	—			4
5. Selfview	4				—	17	10	6	15	2	2	—	17	17	2	4	6		—	2	21
6. Still Picture	10	8			35	—	2	4	13			8	—	44	4	15		2	21	—	31
7. Videophone	6	2			4		—	15	8	4		6	15	—	4		2		8		—

Notes. Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

TABLE 4. Confusion Matrices, Malaysia (n = 48)

Referent Presented	Symbol Selected																				
	Set 1							Set 2							Set 3						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Camera	—	15	2		6	2	4	—	4	2	6		4	21	—	2	2		13	19	
2. Document Camera	4	—	0		4			13	—		2		4	4	13	—			13	2	
3. Handsfree	4	2	—	25	10	6	15	6	2	—	31	19	4	8	13	2	—	8	2		35
4. Microphone			8	—			4	6	2	31	—	23	2				10	—			4
5. Selfview	2		2		—	15	8	10	17	4	2	—	8	38			2		—		35
6. Still Picture	2	2	2		17	—		4	15		2	6	—	31	4	6	2		35	—	17
7. Videophone	4				2		—	8	21	8	13	25	8	—	2	4		2	4		—

Notes. Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

TABLE 5. Confusion Matrices, the Philippines (n = 48)

Referent Presented	Symbol Selected																				
	Set 1							Set 2							Set 3						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Camera	—	13			2	2		—	8		2	6	6	13	—	6	2		8	10	2
2. Document Camera	10	—	2			4	2	8	—	2	4		2	15	8	—			10	10	2
3. Handsfree		8	—	17	8	6	25		4	—	23	10	15	15	4	2	—	25	6	6	23
4. Microphone			8	—			2			33	—	21	6				6	—	4		2
5. Selfview			4		—	4	10	4	2	6		—	17	38	2				—	2	17
6. Still Picture	2	6			23	—	2	10	15		4	21	—	25	8	4			35	—	6
7. Videophone	2	2			2		—	4	10	10	4	25	15	—	4	10			6		—

Notes. Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

Set 1's symbol for document camera (1) was wrongly selected by respondents from Indonesia when the referents camera (10.4%), still picture (8.3%), and videophone (2.1%) were presented. Between sets, Set 2 clearly had the most number of confusions above this level and Set 1 the least in all five countries. Across symbol sets, selfview and videophone symbols had the highest instances of being mistaken as representing other functions. More differences were noted when the countries were compared based on

TABLE 6. Confusion Matrices, Thailand (n = 48)

Referent Presented	Symbol Selected																				
	Set 1							Set 2							Set 3						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Camera	—	8				4	4	—	10			2	6	19	—		2		8	25	2
2. Document Camera	10	—				2		15	—			2	4		19	—				4	
3. Handsfree	6		—	17	6	2	8	2	2	—	33	6	2	4	4		—	23	4	2	17
4. Microphone	2		8	—						38	—	15		6			8	—			2
5. Selfview	8	2			—	19	10	6	8		4	—	2	42	6	2			—		23
6. Still Picture	8	6			31	—	4	2	8	4		6	—	27	15		4		33	—	13
7. Videophone	6	2					—	13	6	2	13	8	21	—	13	4			6		—

Notes. Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

TABLE 7. Confusion Matrices, Sri Lanka (n = 48)

Referent Presented	Symbol Selected																				
	Set 1							Set 2							Set 3						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Camera	—	10	8		17	8	15	—	19		6	8	6	19	—	2	2		13	21	8
2. Document Camera	8	—	4	2	2	4	4	19	—	4	17		6	19	31	—	4		8	6	4
3. Handsfree	13	8	—	21	15	4	21	4	10	—	21	25	19	6	8	13	—	10	15		29
4. Microphone	2	2	17	—	2	2	8		10	19	—	31	6	4	6	4	4	—	10		
5. Selfview	4	8	4		—	31	13	8	10	2	8	—	19	31	6	15	6	2	—		31
6. Still Picture	6	6	4		44	—	2	4	17	4	2	2	—	23	4	15	4		42	—	25
7. Videophone	2	8	2		6		—	8	40	4	8	17	10	—	2	15	2	2	4		—

Notes. Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

symbol confusions. For example, whereas Thailand had more instances of confusing Set 1's symbol for camera than the other four countries, it had the lowest instances of confusing Set 1's videophone symbol for other functions. In asymmetric confusions, the problem may lie in the vagueness or too much generality of the symbol or symbols in question with regards to the other referents. Thus, they can easily be associated with several referents. The aforementioned results showed the countries

TABLE 8. Missing Values or No Answers Given in Cued Response Tests per Set (n = 48 per Country)

HITS	Camera			Document Camera			Handsfree			Microphone			Selfview			Still Picture			Videophone		
Countries	/1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Indonesia	6	—	—	2	—	2	6	15	2	—	15	2	8	23	—	—	6	—	—	17	—
Malaysia	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—
The Philippines	—	2	—	—	—	—	—	2	—	—	—	2	—	2	—	—	—	—	—	—	—
Thailand	—	—	2	—	—	—	4	2	2	—	6	—	—	6	—	—	—	—	—	4	—
Sri Lanka	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

with varying patterns of asymmetric confusions. These could be useful in deciding which symbol (or referent) in each country may need more explanations and tests to avoid vagueness and misconceptions of the functions being represented.

Confusions are also very useful indicators of the suitability of the symbol and may even complement the hit rates. For example, the videophone symbols of Sets 1 and 3 had very similar hit rates among the countries. Nevertheless, when symbol confusion was also considered, Set 1's version was better because of its lesser number of confusions. With regards to the microphone symbols wherein participants from the five countries had similar hit rates also, Set 3's version performed best for Indonesia, Malaysia, and Sri Lanka (lower symbol confusions). In the case of the Philippines and Thailand, Set 1's microphone would be the better choice.

Misses are the opposite of hits. These occur when the participant selected the wrong symbol for the referent being presented. Tables 3 to 7 illustrate the misses (in percentage) in rows in accordance to the referents it was wrongly associated with. The essential thing with misses is that their pattern of distribution under each referent can give an idea of the understandability of the referent or function itself. For example, the referent handsfree was associated with almost all the symbols of in all sets in quite high levels. This could suggest that the concept of handsfree might not be entirely clear to most of the participants in all countries.

3.3. Missing Values

Missing values were instances when some of the participants gave no response or answers during the cued response tests. Table 8 illustrates the distribution of the missing values across countries for each set. Between sets, Set 2 had the highest instances of missing values. Between countries, Thailand had the most instances of missing values among all countries. Its missing values ranged from 2 to 23% across all 21 symbols. Sri Lanka was the opposite with no missing values in the cued response test. When arrayed with the other non-hit parameters or errors, they are also very important as they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). In actual user scenarios, they can be akin to non-use or under-use of the device.

4. CONCLUSIONS

In this study, three sets of videophone symbols designed and tested in the west were tested among participant groups from the east (Southeast Asia). And when symbols are intended as parts of interfaces of devices for international use, different tests are usually needed. Usually, the different parts are designed to measure initial (a priori) meaning and appropriateness, ease of learning and remembering, as well as the probability of confusion with other symbols. The cued response test simulates (though in a limited manner) real videophone call scenarios wherein the symbols are placed on the terminal with one function to be performed at a time. It depicts situations of using the videophone the first few times with bits of information and instructions given. Determination of the hit rates gives the level of association of the symbol to its intended referent or function. However, when symbols appear simultaneously, the hit rates cannot reveal the dynamics of how each symbol can affect the other's level comprehension and association to its referent. Thus, analyses of the errors or non-hit parameters are crucial. Confusions are instances when symbols are selected under the wrong referent contexts. Together with hit rates, they can give an idea on a symbol's distinctness from other symbols under different referent contexts as well as the different patterns of their interactions in different groups. Subjective certainties are also important. They reflect the participants' degree of confidence or confusion with their answers. Other parameters also exist. They are not discussed but are also helpful in evaluating graphical symbols. For example, preferences generally indicate the aesthetics. They are useful especially when considering which among symbols with similar hit rates are more appealing for the participants. Lastly, this study was just one part of an international project. It attempted to demonstrate how multiple indices, including those that analyse non-hits can be helpful in evaluating symbols across different user groups. Different patterns in errors between countries were shown, but admittedly, deeper analyses are still needed to convert such observations into more practical guides towards designing symbols across diverse user groups—issues that will be addressed by the succeeding parts of the project.

Altogether, the results of this study supports the view that analyses of errors or non-hit parameters such as confusions, misses, and missing values are invaluable in properly evaluating graphical symbols when they appear simultaneously or in groups. They enable the tester to see other often subtle but important differences on how different users perceive and understand symbols. These are important not only in properly interpreting test results of different user groups but also in formulating instructions and other aids in learning to use new products faster and more satisfactorily.

REFERENCES

American National Standards Institute (ANSI). (1991). *Accredited standard on safety colors, signs symbols, labels and tags* (Standard No. ANSI Z535.1–5). Washington DC, USA: National Electrical Manufacturers Association.

Böcker, M. (1993). A multiple index approach for the evaluation of pictograms. In *Proceedings of the 14th International Symposium on Human Factors in Telecommunications, Darmstadt, Germany* (pp. 73–85). Heidelberg, Germany: R. v. Decker's Verlag, G. Schenck GmbH.

European Telecommunications Standards Institute (ETSI). (1993). *Technical report—Human Factors (HF): Results of an evaluation study of pictograms for point-to-point videotelephony*. Sophia Antipolis, France: Author.

International Telecommunications Union (ITU). (1995). *Procedures for designing, evaluating and selecting symbols, pictograms and icons* (Recommendation F.910). Geneva, Switzerland: Author.

International Organization for Standardization (ISO). (1989). *Procedures for the development and testing of public information symbols* (Standard No. ISO 9186:1989). Geneva, Switzerland: Author.

Magyar, R.L. (1990). Assessing icon appropriateness and icon discriminability with a paired comparison testing procedure. In *Countdown to the 21st Century. Proceedings of the Human Factors Society 34th Annual Meeting, Orlando, Florida* (pp. 1204–1208). Santa Monica, CA, USA: Human Factors Society.

Nolan, P.R. (1989). Designing screen icons: Ranking and matching studies. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 380–384). Santa Monica, CA, USA: Human Factors.

Tudor, L.G. (1994). The generation and testing of candidate telecommunications icons. In B. Blumenthal, J. Gornostaev, & C. Unger (Eds.), *Proceedings of the East-West International Conference in HCI, St. Petersburg, Russian Federation* (pp. 11–62). Moscow, Russian Federation: International Centre for Scientific and Technical Information (ICSTI).

Vora, P., Helander, M., Swede, H., & Wilson, J. (1991). Developing guidelines for symbol design: A comparison of methodologies. In *Proceedings of Interface 1991* (pp. 6–11). Dayton, OH, USA: University of Dayton.

Webb, J.M., Sorenson, P.F., & Lyons, N.P. (1989). An empirical approach to the evaluation of icons. *SIGCHI Bulletin, 21*, 87–90.

Zwaga, H.L.G. (1989). Predicting the public's comprehension of graphic symbols. In E.D. Megaw (Ed.), *Contemporary ergonomics 1989* (pp. 431–436). London, UK: Taylor & Francis.