Experimenting with an Emotions Measurement Instrument in Usability Testing

John D. Fernandez, Ph.D., Texas A&M University – Corpus Christi
Mary Alice Fernandez, LPC, Texas A&M University – Corpus Christi
H. Richard Priesmeyer, Ph.D., St. Mary's University - San Antonio

Introduction

Human computer interaction (HCI) is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of related major phenomena[5]. The preponderance of research in HCI has been focused on the components of design and implementation of interactive computing systems. Somewhat less attention has been given to the evaluation of such systems. This paper addresses methods for teaching and conducting usability evaluations, including the use of an emotions measurement instrument.

The main goal of HCI is to build interactive systems that are easy to learn, effective to use, and enjoyable from the user's perspective[5]. These characteristics are summed up in one word – usability. Usability can only be understood from the user's mind-set. Glass (as cited in Pressman[7]) contends that even the quality of a system is not as important as the user being satisfied, because if the user isn't satisfied, nothing else really matters. Therefore, usability is of utmost importance in measuring a software product's positive impact on the user.

Since the focus is on satisfying the needs and desires of the user, the development of interactive systems should follow a user-centered approach that is sensitive and responsive to those needs. This foundational principle becomes a challenge for the typical software developer because it calls for skills and abilities that are not normally part of the software engineer's portfolio[1]. HCI is an interdisciplinary methodology that involves skills from disciplines such as psychology/cognitive science, engineering, informatics, computer science/software engineering, ergonomics, human factors, and social sciences[6]. Therefore, HCI is closely aligned with user-centered development. Web-based systems are good examples of HCI-type systems because of the availability of the medium upon which they operate.

HCI and Usability Education

One of the authors, J. Fernandez, teaches two community-based HCI courses at Texas A&M University – Corpus Christi[2]. The types of projects that seem to fit best for the HCI courses are Web-based systems with interactive components. Before initiating these courses, contacts are made with city, school, and university organizations in order to find HCI-type requirements that can provide the basis for student projects. Once client organizations are identified, the principals in each are informed of the process that students will use to develop and complete the projects.

Before releasing the students to have their first client meeting, lecture material is presented on the basics of user-centered development and guidelines are given on how to gather information and conduct interviews. The students cover HCI principles throughout the course, but the basic development methodologies for designing and implementing solutions for the clients are studied in the first weeks of the semester. A key step in analyzing requirements is for the students to determine all the stakeholders associated with the system they are assigned. Students are directed to emphasize scenario-based design using Rosson and Carroll's[9] or McCracken & Wolfe's[5] textbooks as guides. After an analysis of problem scenarios, a primary persona is chosen to focus the design of the Web site. The primary persona is the principal stakeholder or user of the system.

Low fidelity and high fidelity prototyping are used extensively by the students as presentations are regularly made to report progress to the class. When the system is in semi-final form, students are required to prepare materials for usability evaluations by class members. Using a large computer lab, students evaluate each other's work and provide written feedback to the developers.

Before initiating usability testing, students are asked to prepare a task list that would be common for the primary users of the Web site. This task list serves as the test scenario to be used by students assigned to test the site. Students are also asked to use data collection forms provided by the professor. These forms are used to note the start and stop time for each task being completed by the site evaluator. There is room on the forms to annotate problems or items of interest for later study by the development team. The usability test materials also include a brief questionnaire that is used by the evaluator to give overall feedback on various aspects of the site, such as navigation, and other elements of significance. This questionnaire, essentially a usability survey, is created by the Web site developers and completed by the evaluator as soon as he or she completes the evaluation.

All the information collected during the evaluation process is reviewed by the professor immediately at the end of each evaluation and is later included in the final systems document submitted for the project. Some of the students expressed having found the evaluation process a valuable experience because they could now clearly see that their view of the system's usability may not be the same as that of the user.

User Satisfaction

Krug[3] points out that the real expert in usability is the user. Despite the best efforts of HCI experts and Web designers, the ones who ultimately determine if a Web site is useful (and therefore successful) are the users. Users, as a collective group, bring such a wide diversity of skill levels and backgrounds, that it is impossible for Web designers to anticipate and address every technical or cultural nuance. Therefore, it is worthwhile to give the users a voice in the design of Web sites through some form of a questionnaire as utilized in the evaluations noted above.

Since developers are pursuing the user's approval, they are interested in the user's bottom-line satisfaction with the Web site. This phenomenon begins to enter into the world of users' emotional reactions to a Web site. According to Marcus[4], the HCI community's interest in emotions is heating up. Many questions arise about which emotions measure user satisfaction and Marcus includes a good discussion of the history of emotions research and states that this area of study is growing in scope for HCI professionals.

One can also approach the study of user satisfaction from a more traditional product evaluation perspective. Priesmeyer et al.[8] present an excellent case for measuring emotions to understand people's purchase intent of products and services. This work is described in some detail in the next section.

Measuring Emotions To Detect User Satisfaction

Priesmeyer et al.[8] developed a computer-based system that measures and interprets eleven human emotions. Although Marcus[4] presents a variety of studies that utilize a different number of emotions, there is no universal agreement as to the correct set or number of emotions that should be considered. Priesmeyer et al.[8] used the computer program, called the Emogram to determine the emotional quality of products as evaluated by a user.

Darwin made the first major contribution to the study of emotions and provided an initial list of separate emotions, highlighting the way in which emotions are expressed[8]. Paul Ekman (as cited in Priesmeyer et al.[8]) advanced the study of emotions by focusing on facial expressions. Ekman identified specific muscle groups that relate to specific emotions and developed a coding system (Facial Action Coding System) that allows one to identify an emotion from the combination of facial muscles used to express it.

Based on substantial research and cited studies, Priesmeyer et al.[8] decided to include the following list of basic emotions in the Emogram system: happiness, interest, surprise, contempt, disgust, shame, fear, anger, distress, sadness, and anxiety. Richard Lazarus (as cited in Priesmeyer et al.[8]) confirmed many of the basic emotions in this list through his own research and provided much discussion about the interaction of emotions that directly contributed to the interpretation and analysis of emotional dynamics in the Emogram system. An important step in the development of Emogram was the creation of precise photographs depicting varying degrees of the eleven basic emotions.

The Emogram system provides measures of each of the emotions by combining responses to low, medium, and extreme expressions of each. It also computes an overall Emotional Quality (E-Quality) score that reflects the overall emotional state of the individual at the time of the assessment. This E-Quality score is computed as the difference between the average of the pleasant emotions (happiness, interest and surprise) and the average of the unpleasant emotions (contempt, disgust, shame, fear, anger, distress, sadness and anxiety). The difference is then recalibrated to range from +100 to -100. The result is a

measure that reflects a more satisfying overall emotional state as the measure approaches +100 and an increasingly uncomfortable state as the score approaches -100.

To measure the emotional qualities of a product using the Emogram, individuals take a baseline Emogram prior to being exposed to a product. The individual is then exposed to the product and then takes the Emogram a second time. Each Emogram assessment takes approximately six to eight minutes. The Emogram program then computes the strength of each emotion based on the reaction of the consumer and analyzes the changes in the emotions from the baseline to the second assessment. During the assessments the consumer is not asked to think about how he feels, but rather to respond to each photograph at an emotional level to more accurately reflect his or her emotional state at the time of the assessment. Specifically, the question posed to the individual is, "To what extent do you feel the way the person in the photo feels?" The Emogram assessment screen and the available responses to this question are shown in Figure 1 below[8]. This protocol effectively bypasses many of the problems associated with attempts to measure emotions with words. It was followed by Priesmeyer et al.[8] to successfully evaluate consumers' reactions to a unique stress-relieving recliner called the Dondolo.

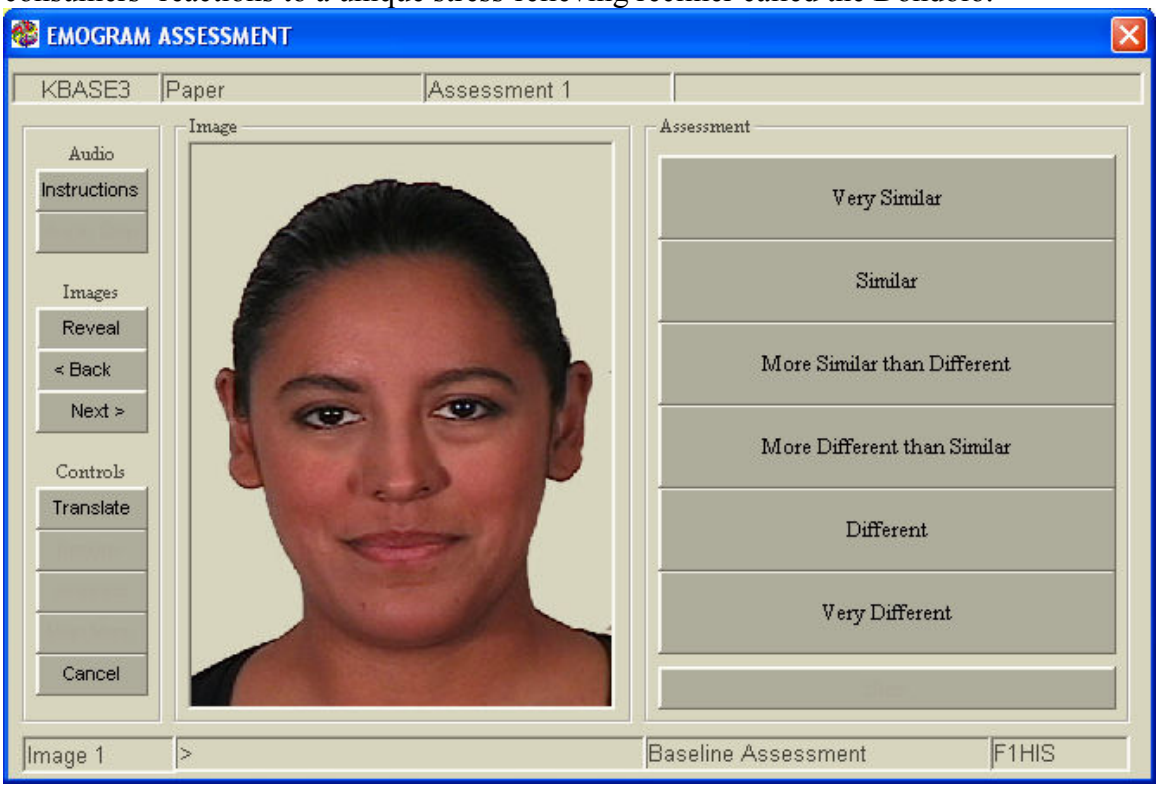

Figure 1: Sample Emogram Evaluation Screen

Both J. Fernandez and M. Fernandez were certified by H. R. Priesmeyer as Emogram clinicians after extensive training and practice with the instrument. The training reflects the fact that the Emogram assessment is considered a psychoanalytic instrument and is therefore governed by the rules of conduct and ethics imposed on mental health

professionals. Administrators of the Emogram must be trained by certified Emogram trainers and be awarded a license to use the instrument.

Preliminary Experiment Relating Emotions Related to Web Site Usability

In order to test the use of Emogram for Web site assessments, a preliminary experiment was conducted using the A&M-CC Web site and three undergraduate students. The test scenario involved locating the Institutional Review Board (IRB) form on the University Web site. The students were initially administered the Emogram while targeting a state of being relaxed, making sure they understood that the exercise was not a test. While relaxed, the students' emotions were assessed using the Emogram and following its test procedure This assessment constituted the baseline and provided scores for each of the eleven basic emotions and an overall E-Quality score. Students were then asked to complete the test scenario. During the execution of the test scenario, students were directed towards a solution if they struggled with the test for more than two minutes. Upon completion of the scenario, each student was assessed again by taking a second Emogram while targeting their experience with the Web site's usability.

The results of the experiment were very interesting because it was quite clear that the E-Quality score of each student declined considerably after only a two-minute struggle with the test scenario. Figure 2 shows the results of the preliminary experiment using Emogram for Web usability assessment. Each of the students had to be redirected to find the solution after two minutes of struggling with the assigned task and this frustration is clearly seen in the results achieved. The individual emotional scores were not analyzed in this preliminary assessment; the E-Quality score was the only measure used to assess the success of the instrument in showing the emotional response of the evaluators to the Web site.

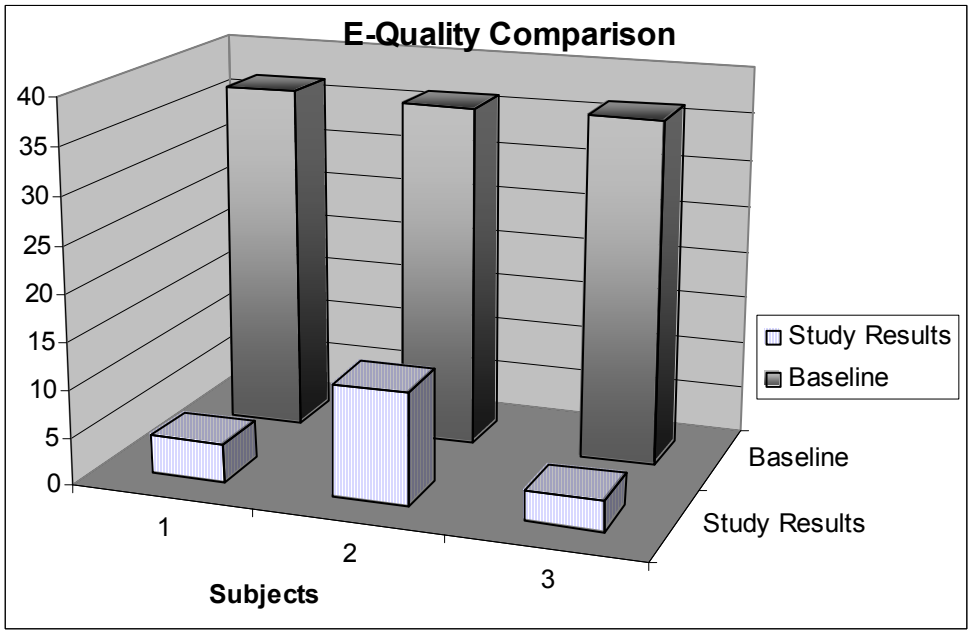

Figure 2: Comparison of Emotional Quality Scores

Multiple Web Sites Assessed with Emotions Measurement

The results obtained in the preliminary test suggested a more comprehensive test with students in the HCI course. As before, students in an HCI course were assigned Web site development projects for community organizations. A total of five teams of students developed Web sites following the structured process presented in the class. When it was time for the normal in-class usability evaluations, students were again assigned to evaluate the projects of other students to provide feedback to enhance the Web site for the community clients. Each team prepared a test scenario based on the main tasks that the primary Web site users were supposed to be able to accomplish. Five to six student evaluators were assigned to test each Web site using the test scenario. The Web developers followed the normal protocol of tracking the evaluation by noting the start and stop times of each individual sub-task and any problems encountered by the evaluators.

To incorporate a process for measuring emotional responses to the Web sites, the protocol was amended to include the following. The professor (J. Fernandez) provided a questionnaire (usability survey) to be used by each evaluator after the completion of the test scenario. This generalized questionnaire was standardized for use by all evaluators so that it could be compared with the Emogram assessments. Once the student completed the usability questionnaire, he or she was immediately dispatched to another computer lab where M. Fernandez administered the Emogram while students were targeting their experience with the Web site they evaluated. Students evaluated multiple Web sites and all the data were collected for analysis.

Analysis of Results of Usability Tests With Emotions Measurement

Although baseline assessments were administered to evaluators, it was determined that they were not useful in this case because each student evaluated two or three Web sites after the initial baseline. However, the usability questionnaire was used to develop a hypothesis for the experiment. After quantifying the responses on the questionnaire, a null hypothesis was proposed as follows: there is no relationship between the results of the usability questionnaire and the results of the Emogram assessments.

Thirty evaluations were conducted by students, so each Web site received an average of six evaluations. This resulted in 30 evaluations and usability surveys that were quantified. A perfect score on the usability questionnaire was 25 reflecting a highly positive experience with a Web site. The lowest score received by a Web site was that of 10, which reflected a negative reaction to a Web site. Thirty Emogram assessments of Web site reactions were conducted after the usability surveys were completed.

The Emogram results for each emotion can range from a low value of one (1) to a high score of six (6). Some of the emotions are combined to six other measures computed from the emotional scores: openness, internal accountability, external accountability, congruence, incongruence, and relevance. Each of these additional measures can also

range from a high of six to a low of one. The Emotional Quality score has a range from +100 to -100 as described above.

SPSS was used to analyze the data. Specifically, Spearman's rho was used to calculate the correlation coefficient between the Emogram scores and the overall scores on the usability questionnaire. The following results were found. Distress was found to have a significant negative correlation with the usability score at the 0.05 level. Sadness was found to have a significant negative correlation with the usability score at the 0.05 level. Happiness was found to have a significant positive correlation with the usability score at the 0.01 level. Finally, the overall E-Quality Score summarizing all of the emotions collectively was found to have a significant positive correlation to usability at the 0.05 level.

Conclusions and Future Work

One can reject the null hypothesis and indicate that there is a relationship between how a student cognitively rates a Web site on a questionnaire and how the student emotionally assesses the same Web site. However, more data analysis needs to be conducted to substantiate the findings of this study.

More research is needed with the Emogram instrument to assess its use in usability assessments. A large pool of evaluators is needed in order to take advantage of the baseline assessment for each evaluator. The protocol of baseline, treatment, and assessment should be followed for future work. Because user satisfaction is likely driven by both cognitive and emotional responses, the use of a heuristic evaluation questionnaire should be continued in the future work, but such measures should be complemented with measures of specific emotional responses. This study shows that Emogram can be used to assess a user's emotional satisfaction with a Web site. To the extent that satisfaction is based on emotions, the Emogram scores can supplement traditional cognitive measures and lead to improved Web site designs.

References

1. Fernandez, J.D., Human-computer interaction closes the software engineering gap, *Proceedings of the 2004 American Society of Engineering Education Annual Conference*, June 2004.

2. Fernandez, J.D., Engaging students with community organizations by using computer technology, *SIGITE 2004 Conference,* October 2004.

3. Krug, S., *Don't Make Me Think: A Common Sense Approach to Web Usability*, New Rider Press, Indianapolis, IN, 2000.

4. Marcur, A., The emotion connection, *Interactions,* November-December, 2003, 28-34.

5. McCracken, D. and Wolfe, R., *User-Centered Website Development: A Human-Computer Interaction Approach,* Pearson Education Inc., Upper Saddle River, NJ, 2004.

6. Preece, J., Rogers, Y., and Sharp, H., *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons, Inc., New York, NY, 2002.

7. Pressman, R. S., *Software Engineering: A Practitioner's Approach*, *6th Ed.,* McGraw Hill, New York, NY, 2005.

8. Priesmeyer, H. R., Axiomakaros, P., & Murria, M. A., Human emotions and purchase intent: How to measure the emotional quality shift or products, services, and advertising, from proprietary workshop titled *Marketing Emotions,* a product of Chaotics International, 2003.

9. Rosson, M.B., and Carroll, J.M., *Usability Engineering: Scenario-Based Development of Human-Computer Interaction,* Morgan Kaufmann Publishers, San Diego, CA, 2002.

Bibliographical Information

JOHN D. FERNANDEZ
Dr. Fernandez is Assistant Professor of Computer Science in the Department of Computing and Mathematical Sciences. Having served 20 years in the U.S. Air Force and 10 years in private industry, Dr. Fernandez brings real-world experiences into the classroom for his students. His research interests are in HCI, information assurance, and software engineering.

MARY ALICE FERNANDEZ
Ms. Fernandez is a Licensed Professional Counselor and program coordinator for the strengthening families program at the Garcia Arts and Education Center. She is a doctoral student in the Counselor Education program at Texas A&M University – Corpus Christi. Ms. Fernandez has extensive teaching and counseling experience with at-risk families.

H. RICHARD PRIESMEYER
Dr. Priesmeyer is Professor of Business Management in the School of Business and Administration at St. Mary's University and founder of Chaotics International, the creator and distributor of the Emogram system. Dr. Priesmeyer has published extensively in the field of chaos theory and its application to several domains. His Emogram system is used in various counseling practices to assist people understand their emotions.