Relationship between Elements of the Usability and Emotions Reported after Use: A Mexican Case

Irma Cecilia Landa-Ávila[1] and Lilia Roselia Prado León[2]

[1] Maestría en Ergonomía con orientación en Diseño,
Universidad de Guadalajara, Guadalajara, MEX
clandavila@gmail.com
[2] Centro de Investigaciones en Ergonomía,
Universidad de Guadalajara, Guadalajara, MEX
aililpleon@gmail.com

Abstract. This paper presents the study of Mexican users of design software to find out if there is any relationship between any of the elements of usability and emotions reported after using said product.

A usability test was conducted with inexperienced users who had to perform three tasks in sequential order, under a spoken protocol. After each task, participants responded to the SMEQ mental effort questionnaires, the ASQ satisfaction questionnaire and used the PrEMO tool to report their emotions and feelings while using the software. Subsequently, video recordings of the tests were reviewed to find the level and quality of completed tasks and the number of errors.

We found that there is a relationship between some of the usability metrics and the emotions of the users, which is manifested more significantly between subjective usability metrics. The perception of the perceived mental effort turned out to be the element with the biggest relationship to the reporting of emotions, however, this relationship appears to affect both positive and negative emotions and some assumptions point to factors such as the sudden increase in the complexity of a task, which increases the force with which negative emotions are reported.

Keywords: Web Usability, User's Emotions, Mexican Usability Test, Software Design.

1 Introduction

The importance of emotions as an integral element of the user experience is being widely considered in current studies of Human Computer Interaction (HCI). It is, however, considered a secondary aspect. The role of the elements of usability such as efficiency, effectiveness and satisfaction still positions itself as the most important and reliable for the study of the use of products.

Emotions during product use have been considered in recent studies such as in Brave and Nass (2002) who have described some emotion metrics and their

A. Marcus (Ed.): DUXU 2014, Part I, LNCS 8517, pp. 285–295, 2014.

application at a more theoretical level. Other approaches have shown that humans are more efficient and have managed to solve creative problems when they are happy (Hirt, ER et alt, 1996) and that emotion is linked to acceptance and user satisfaction and can greatly impact purchase intentions (Erevelles, 1998 S. Martino, et alt 2006).

Many professionals in cognitive studies argue that it is impossible to act or think without entering (at least subconsciously) into our emotions. Until recently, HCI research tended to focus on the ease of use of the cognitive system, in which the topics of greatest interest were related to ease of use, efficiency, ease of learning and error handling. These cognitive factors are certainly of great importance to HCI, but the feelings of users interacting with the system are equally important, as stated Barnes and Thagard (1996) in their study. They argued that emotions interact together with knowledge in order to achieve a certain goal.

The objective of this study is to discover whether there is any relationship between the traditional elements of usability (effectiveness, efficiency and satisfaction) of a piece of design software with emotions reported by inexperienced users after using the product.

2 Background

2.1 Usability

The International Organization for Standardization (ISO) has defined usability from two perspectives. The first one is in the ISO / IEC 9126, which states: "Usability refers to the ability of software to be understood, learned, used and being attractive to the user, within specific conditions of use". This definition emphasizes the internal and external product attributes, which contribute to its functionality and efficiency. In contrast, in the ISO / IEC 9241 we find the following: "Usability is the degree to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specific context of use".

In their study Bevan and Macleond (1994) found that the nature of the scope required in the study of usability depends on the context of use of the product. Bevan said that usability is a feature of the system as a whole, that is "the quality of its use in a certain context.

In the study by Frøkjær, Hertzum and Hornbæk's (2000), correlations between the three aspects of usability are analyzed. They found that the correlation between effectiveness and efficiency it is so weak, that it cannot be used for practical. They determined that those three measurements should be considered as an independent aspect of usability, but all must be included and evaluated in the test.

Sauro and Dumas (2009) conducted a review of the most common satisfaction tests. Among the reviewed instruments, the scale of the Subjective Mental Effort Question (SMEQ) showed the best performance, it was easy to learn and highly correlated with other measurements, so their suggestions pointed at its use on other post-test scales.

In the study, Lewis (1991) concluded that the three items in the ASQ questionnaire can greatly condense the results of post-test assessments by adding up the results.

Another conclusion was that ASQ is sensitive enough to be used in usability studies, and that its concurrence validity in each of the scenarios was equally good.

For this study, a usability test was performed with voice protocol, and at the end of each task the SMEQ and ASQ questionnaires will be applied. Subsequently, the level and quality with which a task was completed will be recorded, in addition to the number of errors; allowing us to offer a comprehensive assessment of usability.

2.2 Emotions

Emotion is a psychophysiological reaction to certain stimuli related to needs, goals or individual concerns. Emotions are affected by innate and learned influences, with invariant features and others that depend on the group or culture to which they belong (Levenson, 1994).

While evoked emotions are idiosyncratic, there are universal patterns that can be identified in the essential process of how these emotions are evoked. Usually, the duration of an emotion lasts merely seconds or minutes (Ekman, 1994), so reporting of emotions must be performed immediately after they are generated.

In the field of HCI, the interpretation of emotions focuses on emotional responses (feelings) that are assigned to an interface during and after their use. It is believed that emotions are intentional, because they comprise and involve a relationship between the person experiencing it and a particular object (Frijda, 1994).Also, people are able to identify the object causing the emotion (Ekman 1992).

2.3 Usability and Emotions

The role of emotions in HCI studies begins to establish the need for instruments capable of measuring with the same rigour as other mechanisms from which traditional usability metrics are obtained. It is said that an emotion always includes evaluating how the use of an object can harm or benefit a person (Arnold, 1960) and that evaluation is always immediate and direct in a positive or negative sense.

In the study by Agarwal and Meyer (2009) compared the difference that existed between the tools to measure emotions verbally and nonverbally. They established beforehand that verbal tools had some limitations, including the fact of dependence on language, and users having to remember how they felt and ascribe a term to it. Meanwhile nonverbal methods attempt to capture unconscious emotional responses and incorporate a certain amount of vagueness appropriate for the study of emotions. They concluded that a positive user experience cannot be expressed only in terms of usability metrics and that it is valuable to study emotions, because they supplement the evaluation of interfaces, since it allowed them to know whether there were differences with standard usability metrics that they would not notice.

2.4 Methods for Measuring Emotions

Emotions are entities of multiple components. In relation to that, tools for reporting emotions can be classified into one of four classes with respect to the measurement components. (Table 1)

Table 1. Measuring Instruments and recorded components

Requirements	Measuring Instruments			
	Emotional expression	Psychological Reactions	Feelings	Subjective
Set product emotions	-	-	+	+
Found emotions	+	+	+	
Transcultural	-		-	+
Equipment / experience			+	+

It is observed that no instrument is capable of measuring all four components. Despite this shared drawback there is a kind of instrument that has some important advantages over the others: the instruments of self-reporting (subjective), they are capable of measuring a mixture of emotions beyond the basic emotions, such tools are in turn subdivided into verbal and nonverbal methods.

PrEMO. The acronym of Product Emotion Measurement is an instrument specifically designed to measure the emotions of any product. Developed by Pieter Desmet (2005), it is based on 14 animations of a cartoon character, seven of them pleasant (desire, pleasant surprise, inspiration, fun, admiration, contentment and fascination), and seven unpleasant (anger, contempt, disgust, unpleasant surprise, dissatisfaction, disappointment and boredom).In the validation process, two emotions failed the test, so they were removed from the instrument.

It measures various and mixed emotions and does not require participants to verbalize their emotions, avoiding excessive mental effort, since it was designed to be fast and intuitive to use. The visual display is accompanied by a body animation of a character and a sound that lasts approximately one second. Participants are required to report their emotional response to the twelve animations by interacting and placing each of them on a scale of five values ranging from 0 for no feeling to 4 for a strong feeling.

3 Methodology

3.1 Participants

Recruitment was used to obtain participants for convenience. The exclusion criteria were being older than 18 and not having previous experience with design software.

It featured a sample of 32 participants (11 men and 21 women, mean age of 21 years and a standard deviation of 3.9).

3.2 Materials

The test was conducted in a room with two computers, one for the use of the software, in which the user performed the tasks of the test, and a program that allowed videotaping the test for later review; in the computer the PrEMO tool for reporting emotions was installed.

Reporting PrEMO Emotions
To report emotions, the PrEMO system was installed on a computer with internet connection. The tool consists of a character showing 12 different emotions, which should be evaluated through a 5-point scale by the participants after each task, the range was from 0 if no emotion was felt and it goes up to 4 if they felt it strongly.

SMEQ Scale
It is also known as the effort grading scale. After each task the moderator conducted the Spanish version of the SMEQ questionnaire. This scale has a range of 1 to 150 and nine labels from "not at all difficult to do" to "extremely difficult to do."

ASQ Scale
Self-reported satisfaction questionnaire, it was performed at the end of each of the tasks comprising 3 statements regarding satisfaction with the time taken to perform the task, the ease with which it is performed and help to do so. Here, the participant could respond by using a scale that ranges from 1 for strongly agree until 7 for strongly disagrees.

Time
Each of the tasks was assigned a maximum time for completion, a stopwatch signaled when the time was up.

Number of Errors
After the test, a review of each of the usability tests was performed and a count of detected errors was made. An error was considered as any action that did not respond to the user's intention. It was not considered wrong to follow an order, to conduct a search or to visually scan the interface.

Level and Quality Completed Tasks
A checklist for each of the tasks was made. This list was made according to a subdivision by individual actions, which were assigned a score relative to the complexity they had. Each task could have a score 100 if it had been fully completed with quality and accuracy.

Support for Tasks
Aid was provided for each item, in which some instructions concerning the tasks are listed, participants were delivered a full-scale model of the design that they had to make.

3.3 Tasks

The test was made of three tasks, which should always be made in the same order, due to the learning that participants could obtain from the preceding task.

- Task 1: In the first task the participant had to make a red square of 8 cm and a blue circle of 5 cm in diameter. The maximum time to complete the task was 3 minutes.
- Task 2: The second task was to make a business card, this design required to make rectangles, enter text in specific sizes and place some elements that were provided as input. The maximum task time was 6 minutes.
- Task 3: For the third task the participant was asked to do a poster. Some of the specific actions that had to be done were making boxes, changing colours, copying items, changing size, writing text, and making blueprints. The maximum task time was 10 minutes.

3.4 Procedure

At the beginning of the session participants were informed about the dynamics of the test, permission to videotape the test was requested and an initial questionnaire that collected demographic data was applied, also one for similar design software.

Each of the tasks was explained just before beginning, task aids were given and questions concerning the task were answered. Questions regarding how to do it using software were not answered.

The participants that were performing the task had to talk out loud at all times and they had to mention the actions performed or intended to perform. If participants showed little verbal activity, the moderator made some questions during the course of the tests.

At the end of the task or the maximum end time the ASQ and SMEQ questionnaires were delivered. Immediately thereafter, the PrEMO tool was used for reporting the emotions of each of the tasks, the participants had unlimited time to assign a value to each of the characters used.

Once task 3 concluded, a final questionnaire was administered in which participants were asked about their overall perception and satisfaction with the software. The test had a maximum duration of 45 minutes.

Subsequently, a review of each of usability test for a count of errors per task was conducted. Also, a complete checklist regarding the level and quality of completed tasks was made.

4 Results

The results of the usability tests showed the following results in each of the following areas:

4.1 Efficiency

The metrics considered for efficiency were time and perception of perceived mental effort. Due to the characteristics of the test, each task was analyzed independently. The results are shown in Table 2.

Table 2. Time results and SMEQ by task

Test Tasks	Time			SMEQ		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Task 1	79 seg.	300 seg	155 seg. SD 35.9	0	15	28 SD 25.9
Task 2	360 seg.	360 seg.	360 seg. SD 0	10	142	58 SD 35.95
Task 3	298 seg.**	600 seg.	578 seg SD 74.61	20	150	79 SD 34

**All minimum results of task 3 were obtained with participants who decided to abandon the task.

4.2 Effectiveness

For the effectiveness metrics, two metrics were considered, the number of errors and the level and quality of completed tasks.

Table 3. Results of errors, level and quality of completed tasks

Test Tasks	Mistakes			Level and quality tasks		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Task 1	0	8	3.2. SD 2.12	0	100	55.4 SD 29.9
Task 2	1	13	6.1 SD 2.9	0	60	30.5 SD 17.2
Task 3	2	21	10.1 SD 4.9	0	48	18.9 SD 12.9

4.3 Satisfaction

Satisfaction was reported by users through the ASQ scale, the scale value of 1 was assigned to "strongly agree" for satisfaction, and 7 stood for "strongly disagree." (table 4).

Table 4. Percived satisfaction results (ASQ)

2	Usability Satisfaction			Time satisfaction			Material support satisfaction		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Task 1	1	7	3.88 SD 1.8	1	7	4.97 SD 1.8	1	7	5.41 SD 1.8
Task 2	1	7	3.38 SD 1.9	1	7	4.84 SD 1.7	1	7	5.31 SD 1.8
Task3	1	7	2.56 SD 1.9	1	7	2.59 SD 2	1	7	3.03 SD 2.1

4.4 Relationship between the Elements of Usability

Data was analyzed to find whether there is a relationship between the various metrics that make up usability. The results obtained from the tasks are presented using a bivariate correlation analysis in Table 5, where the three most significant relationships are shown (significant at the 0.01 level correlations).

Table 5. Obtained Correlations with the usability elements

Test Tasks	Related elements		
	Element	In relation with	Coefficient
Task 1	Mistakes	Time	0.764
	Quality and completed tasks level	Time	-0.723
	Satisfaction usability	Perceived mental effort	0.672
Task 2	Satisfaction usability	Perceived mental effort	0.790
	Satisfaction time	Satisfaction usability	0.611
	Satisfaction time	Perceived mental effort	0.495
Task 3	Satisfaction usability	Perceived mental effort	0.600
	Satisfaction time	Satisfaction usability	0.490
	Satisfaction time	Perceived mental effort	0.440

4.5 Emotions

The PrEMO study yielded a report of the emotions per task. In Table 6 the 3 main emotions expressed by participants are listed.

Table 6. Main task emotions expressed by participants

Test Tasks	More strongly emotions factors reported by users		
	Emotions	Scale Value (Frecuency)	Average
Task 1	Fascination	0 (2) + 1 (5) + 2 (6) + 3 (11) + 4 (8)	2.59
	Joy	0 (7) + 1 (7) + 2 (1) + 3 (10) + 4 (7)	2.09
	Satisfaction	0 (6) + 1 (6) + 2 (8) + 3 (6) + 4 (6)	2
Task 2	Fascination	0 (4) + 1 (5) + 2 (4) + 3 (13) + 4 (6)	2.38
	Dissatisfaction	0 (6) + 1 (4) + 2 (10) + 3 (5) + 4 (7)	2.09
	Hope	0 (11) + 1 (2) + 2 (8) + 3 (5) + 4 (6)	1.78
Task 3	Dissatisfaction	0 (6) + 1 (4) + 2 (5) + 3 (7) + 4 (10)	2.34
	Fascination	0 (2) + 1 (6) + 2 (11) + 3 (7) + 4 (6)	2.28
	Hope	0 (6) + 1 (6) + 2 (5) + 3 (9) + 4 (6)	2.09

4.6 Value of the Elements of Usability and Emotions

The analysis of the correlation data, detected the following relationships between the elements of usability and emotions. Table 7 shows the three main task relationships between the elements of usability and emotions reported by users.

Table 7. Primary relationships of usability and emotions elements

Test Tasks	Related elements		
	Element	Emotions	Coefficient
Task 1	Level and quality tasks	Satisfaction (+)	0.660
	Satisfaction usability	Disgust (-)	0.510
	Level and quality tasks	Dissatisfaction (-)	-0.468
	Level and quality tasks	Pride (+)	446
	Satisfaction of time	Disgust (-)	0.444
Task 2	Satisfaction usability	Sadness (-)	0.672
	Perceived mental effort	Sadness (-)	0.666
	Perceived mental effort	Shame (-)	0.650
	Perceived mental effort	Disgust (-)	0.599
	Satisfaction usability	Dissatisfaction (-)	0.589
Task 3 **	Number of mistakes	Pride (+)	-0.457
	Perceived mental effort	Satisfaction (+)	-0.445
	Number of mistakes	Joy (+)	-0.435 *
	Number of mistakes	Satisfaction (+)	- 0.372 *

* Significant correlation at the level 0.05 ** In task 3 were detected only four relations

5 Conclusion

This study allowed us to analyze the results of a usability test and report emotions of users towards a piece of design software.

The first part of the analysis of the results shows that there is a rather weak relationship between some usability metrics such as satisfaction with the ease of use and time (ASQ Questionnaire) and perceived mental effort (SMEQ questionnaire).

The relationship between the number of errors and time, as well as the one concerning the level and quality of task completion time, is worth noting. These two relationships are the only ones considered objective metrics such as time and error count, which suggests that relationships between any methods of self-reporting made after using the product are more frequent.

Similarly, we can assert that the relations of objective metrics were only detected in the first task, in which the time for completion is less, probably because with an increasing time of task completion, the participant loses track of time and that stops it from being a crucial factor that is reflected in the methods that use self-reporting.

Also, the relationship between traditional usability metrics and emotions yields little conclusive results that would allow us to establish the anatomy of the relationship.

Again we find that over 71% of the relationships found were produced with usability metrics, such as the self-reported perceived mental effort and perceived satisfaction that it takes to complete a task, it may be that once the participant externalized some aspects of the test, s/he makes a general assessment of the experience with the use of the product, which is replicated in the other self-reporting tools.

In the second task we can note that relations with greater significance are expressed in negative emotions, although usability metrics were very similar to the ones in task number three, but varied considerably in comparison with the first task, especially concerning the mental effort perceived. This sharp increase in the complexity of the task seems to have had a strong negative impact on the emotions of the users, a phenomenon that is not seen in the third task, because despite being considered an extremely difficult task to do, participants reported emotions such as pride or satisfaction, emotions likely to be associated with the fact that they had completed the usability test in full.

We can state that if there is a relationship between some of the metrics that make up usability and emotions of the users, this is observed more significantly in subjective measurements belonging to self-reporting, where the perception of mental effort turned out to be one of the elements that bears a closer relation to the report of emotions. This relationship appears to affect both emotions, positive and negative, and some assumptions suggest that other factors are the ones that increase negative emotions, such as the sudden increase in the complexity of the task.

6 Discussion

The relationship of the elements of usability with emotional factors has mixed and somewhat ambiguous results that prevent us from establishing its invariance.

The tendency of emotions to show a more pronounced relationship to subjective measurements scales as the ASQ or SMEQ allows us to raise the possibility that the relationship occurs not because of the use of the product. Rather, the externalization of users of any judgment, under any method of self-reporting, is the variable that we must pay more attention to.

There is a need for studies in which self-reporting instruments are controlled. A possible effect derived from the order of delivery of the tools must be considered. In this study, the SMEQ questionnaire recorded the strongest relationships with emotions, and it was the first method used. The variable of the order of delivery of the questionnaires must be considered and one must rethink the relationships found as a result of that.

Similarly, the context of the users must be considered in order to enhance the contribution of the study's findings. Testing expert users, who have used software as a

tool would help to determine the evolving relationship between emotions and usability or, alternatively, whether the relationships are fleeting and random.

The relationship between the elements of software usability and emotions reported by users should be explored in order to clarify these questions.

References

1. Arnold, M.B.: Emotion and personality. Colombia University Press, New York (1960)
2. Agarwal, A., Meyer, A.: Beyond usability: evaluating emotional response as an integral part of the user experience. In: CHI 2009, Extended Abstracts on Human Factors in Computing Systems, pp. 2919–2930 (2009)
3. Bevan, N.: Measuring usability as quality of use. Software Quality Journal 4, 115–150 (1995)
4. Barnes, A., Thagard, P.: Emotional decisions. In: Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, University of California, pp. 426–429 (1996)
5. Bevan, N., Macleod, M.: Usability measurement in context. Behaviour and Information Technology 13, 132–145 (1994)
6. Brave, S., Nass, C.: Emotion in human-computer interaction. In: Jacko, J., Sears, A. (eds.) Handbook of Human-computer Interaction, pp. 251–271. Lawrence Erlbaum (2002)
7. Desmet, P.: Measuring emotion: development and application of an instrument to measure emotional responses to products. Funology, 111–123 (2005)
8. Desmet, P.M.A.: Faces of Product Pleasure: 25 Positive Emotions in Human-Product Interactions. International Journal of Design 6(2), 1–29 (2012)
9. Erevelles, S.: The role of affect in marketing. Journal of Business Research 42, 199–215 (1998)
10. Ekman, P.: An Argument for Basic Emotions. Cognition and Emotion 6, 169–200 (1992)
11. Ekman, P.: Moods, emotions, and traits. The nature of emotions, fundamental questions, pp. 56–58. Oxford University Press, Oxford (1994)
12. Frijda, N.H.: Varieties of affect: emotions and episodes, moods, and sentiments. The nature of emotion, fundamental questions, pp. 59–67. Oxford University Press, Oxford (1994)
13. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems, pp. 345–352. The Hague, The Netherlands Press, New York (2000) (preprint version)
14. Hirt, E.R., Melton, R.J., McDonals, H.E., Harackiewicz, J.M.: Processing goals, task interest, and the mood-performance relationship: A mediational analysis. Journal of Personality and Social Psychology (1996)
15. Levenson, R.: Human emotion: a functional view. In: The Nature of Emotion, pp. 123–126. Oxford University Press (1994)
16. Lewis, J.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. SIGCHI Bulletin 23(1), 78–81 (1991)
17. Sauro, J., Dumas, J.: Comparison of Three One-Question, Post-Task Usability Questionnaires. In: CHI 2009 (2009)
18. Sauro, J., Lewis, J.: Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. In: CHI 2009 (2009)