Capturing & Measuring Emotions in UX

Sarah E. Garcia
UEGroup
1165 Lincoln Avenue, Suite 221
San Jose, CA 95125 USA
sarah@uegroup.com

Laura M. Hammond
UEGroup
1165 Lincoln Avenue, Suite 221
San Jose, CA 95125 USA
laura@uegroup.com

Abstract
Gathering data about the emotional journey of a product and user experience is on the forefront of both user and customer experience, but the question remains: What is the best way to do this? There are sloughs of solutions that claim to capture the user's emotions in various ways: via biometrics, facial analysis, vocal analysis, and more. While some of these solutions can provide you with seemingly accurate feedback, they can also be intrusive. Other solutions can be expensive, leaving a start up or other lean UX team struggling to find these answers. This case study follows UEGroup's approach to tackle the issues surrounding capturing the emotional experience of a product, with a focus on an agile self-reporting method. In this case study we attempt to answer the question-is self-reporting more or less effective than these other emotion capturing methods?

Author Keywords
User Experience; Research Methodologies; UX; Emotion

ACM Classification Keywords
Human Factors

Case Study Presenters:
Sarah Garcia is UEGroup's Senior Lead Researcher with a background in market and user experience research. Sarah is experienced with on screen and device usability, developing innovative testing methodologies and training while mentoring other UEGroup staff for the past 10 years.

Laura Hammond is a UX researcher with a background in library and information science, with a focus on information seeking behaviors. Laura is experienced with user testing and ethnographic research across different user groups, from gamers to cyber security engineers.

About UEGroup
UEGroup is a User Experience and Design company in Silicon Valley that has established long term relationships with leaders in the entertainment and digital media field, medical, consumer electronics, gaming and handheld industries.

Early Emotion Capturing Methods

Emotions, and the role they play in the customer experience, have been in the UX conversation for some time. Biometrics analysis was an obvious candidate for tackling this complex issue, but with quick product iterations and small budgets this is not always a feasible methodology. UEGroup had been using multiple metrics when conducting UX research to attempt to gauge the emotional response to a product, such as probing interview questions and ratings, but this did not get at these answers in a clear and quantifiable way. As stakeholders continued to request this type of research, a moderator interpretation capture was employed. Moderators observed participants' behavior and recorded moments of delight and frustration based off of their facial expressions, actions, and verbal cues. We quickly realized this was far too subjective and difficult to standardize. Ultimately, stakeholders wanted to understand the entire emotional journey of a product with accurate and rich qualitative data that really told the story of that journey. Was some frustration in the middle of an experience suitable if the end result was excitement? Unsure how to accurately interpret users' emotions, a self-reporting method was adopted. Participants would tell researchers how they were feeling at distinct points throughout the research study. This proved to eliminate moderator bias, standardized the methodology, and validated the user's experience as their feelings toward a product were being taken into account.

Researching Emotions

Before implementing the self reporting method, a set of emotions were chosen to aid in reflection and allow for researchers to easily map the emotion data. While multiple disciplines were consulted, emotions were adapted from the work of Robert Plutchik who classified the primary emotions as a base psychological theory. The Plutchik Emotion Wheel [1], developed by Plutchik, offered a variety of emotions that were more nuanced than the eight primary emotions. We decided this was going to be the basis of our self-reporting tool as it offered a spectrum of choices when reporting emotions. We decided to couple this with simple verbal responses to the question, "How are you feeling at this point?". However, we quickly faced the same problems as we had during earlier capture methods: how can one quantify this information? Basic UI principles also suggested that too many options would quickly overwhelm and users would abandon or offer flippant responses, so we pared down the Plutchik Emotion wheel to 8 emotion categories ranging from a positive to negative experience. While less choices was limiting to participants, we felt it was important to make the reporting process quick and easy to decipher. Terminology was also adjusted to make the emotion chart easier to understand and use, if users are judging a product experience terms like "sadness" were rarely relevant.

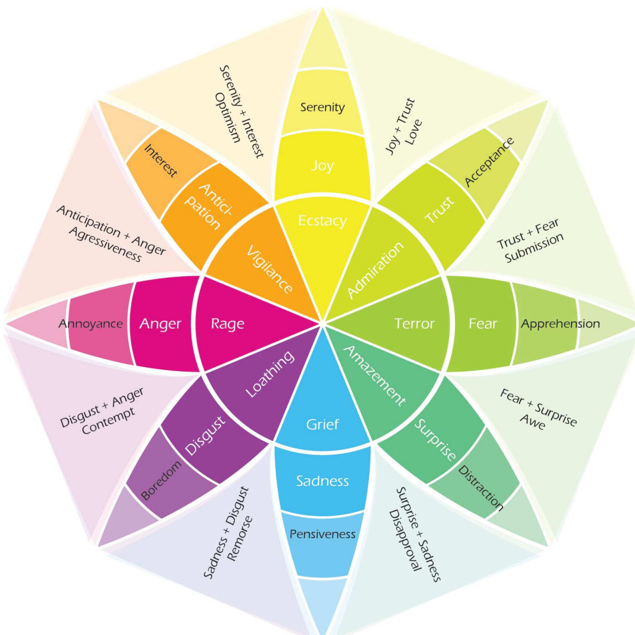

Figure 1: The Plutchik Emotion Wheel

Although we wanted to simplify the experience of selecting an emotion from the Plutchik Emotion Wheel, we did not want to lose the nuances of emotions. The Plutchik method offered a range of related emotions that were either stronger or milder, something that participants could not express if just choosing from the 8 categories we had determined. So in order to capture this we decided to have participants choose a range of emotional intensity within an emotion category. For example, participants could feel very strongly excited, moderately excited, or mildly excited. We felt that this offered the same specificity as the Plutchik Emotion Wheel, but in an easier to use and easier to analyze way.

Paper Prototype

With a list of emotions, intensity levels, and a system in place for data analysis, UEGroup used this method for an out of box experience benchmark study. In addition to gathering ratings and verbal feedback, paper print outs of the emotion chart along with numbered Post-its were given to the study participants. At distinct points in the product set up process, participants were asked how they felt. The moderator recorded their verbal response and then they were asked to place a Post-it on the emotion chart that best matched how they were feeling. We quickly learned that simply self-reported emotions without guidance led to less relevant data. Participants often responded to "How are you feeling?" with statements that did not describe their emotion at all. They would say, "that was fine" or "It was ok" versus considering their emotional state and speaking about it. When probed, many remained stumped. However, participants were reflective when they placed their Post-it on the chart. In addition to noting their emotion, they offered verbal feedback that provided context for their answer. They explained why they chose the emotion they had selected on the chart and they considered their emotion in the context of the whole set up versus the part they had just completed. We also learned that our emotion chart was oversimplified. Although we offered 8 emotion categories with different levels of intensities, this was still not enough options to capture how people feel. When reporting on the emotion chart, participants often felt that one emotion did not properly encapsulate their experience so they would place their Post-it between multiple emotion categories. In addition to participants interacting with the chart in an unexpected way, analysis brought up unforeseen complications. Although we had planned to compare the participants' emotion charts to find themes and produce a heat map of the emotional journey, it was difficult to truly convey accuracy. We knew we had to make some adjustments moving forward, but we were able to create an effective emotional journey chart that conveyed the users emotional experience with this first paper prototype.

Although there were adjustments, we were on to something and began planning chart improvements and more studies that would utilize this tool.

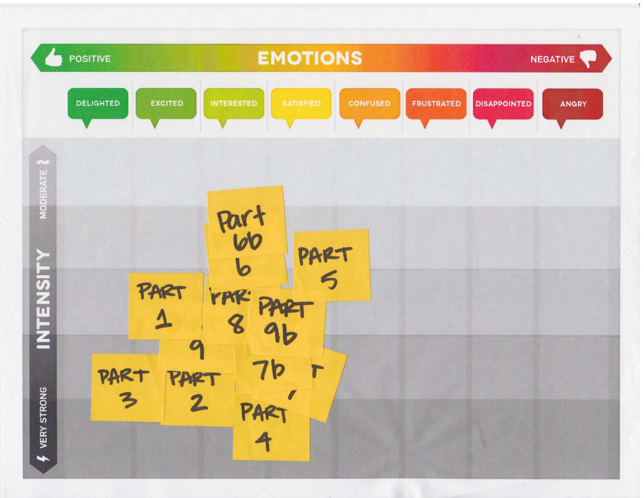

Figure 2: UEGroup's first draft and paper prototype of Emotrak™. Participants charted their emotions with Post-its (© UEGroup)

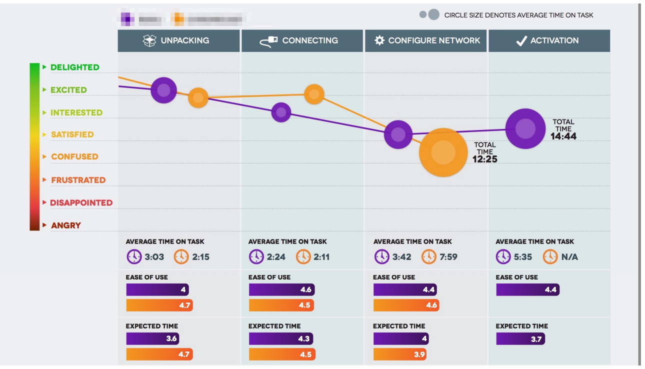

Figure 3 and 4: UEGroup's first emotion chart output. Although the ease of use ratings were similar, the emotional journey chart displayed a much richer story that identified which steps of the process needed to be improved

Paper Prototype 2.0

Although the result began telling the story of the user's emotional journey, the inability to easily and accurately quantify this data was problematic. It was necessary to assign scale values to each emotion, but important to not let these numbers bias the participants when self reporting. Numbers were assigned on the back end and a 9-point scale was adopted, each emotion was given a number value from negative 4 to positive 4. An additional emotion column was added to offer a neutral option for users and a zero column for data analysis. Participants were also instructed to solely select one emotion when self reporting so that the data could be standardized and quantified. In addition to adding another emotion column, we also increased the use of color associated with each emotion, as influenced by the color used within the Plutchik Emotion Wheel. Colors helped participants quickly identify the area where they would eventually place their Post-it and self-report.

We ran another study, this time using 4 different products within a single session. Participants self reported as before, first offering their verbal response then charting it with a Post-it. The addition of numerical values made recording and analysis easier and much more accurate by removing the bias of researcher interpretation. Now we had qualitative data that was easy to quantify and report. However the paper prototype started to bring its own set of problems. It was messy, especially when using it for a study that had multiple products and therefore needed multiple charts. Post-its would not stay adhered to the paper and we would have to review session footage to recall what was reported, sometimes it was indiscernible and the rating had to be dropped.

The paper format also led to influenced responses, as participants were able to see what they had reported in previous parts of the user study simply by looking at the Post-its they had placed on the chart. While this was useful for some studies where they were asked to consider the entire session as part of the same journey, this created bias in projects where they needed to approach unrelated tasks or products with a new outlook. Participants were quickly influenced by their previous answers and would consult what they had charted before reporting how they were feeling at that moment.

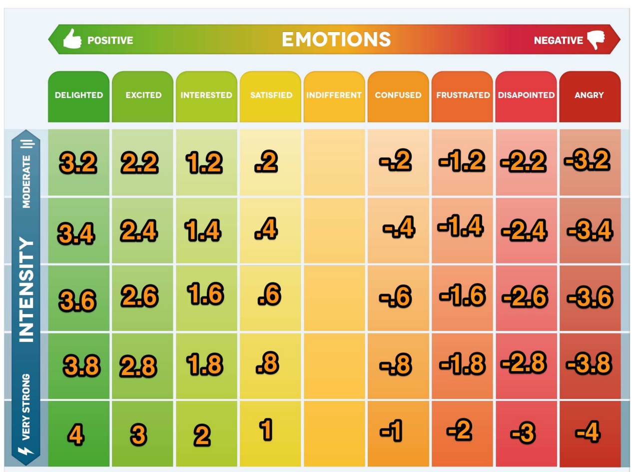

Figure 4: Numeric values were assigned to each emotion category and intensity to improve the accuracy of the output, the emotional journey chart (© UEGroup)

Emotrak™ First Iteration

After several projects with the paper prototype and the same problems, we searched for a way to make this emotion chart digital. The prototype was given a name, Emotrak, and prototyping tools were investigated as possible platforms. After experimenting with open source survey and interaction tools along with standard prototyping tools such as Axure, a development team was brought on to digitize the chart. Since we did not want the emotion reporting to be too intrusive or complex we chose to implement it on a tablet, specifically on an iPad. This would allow for flexibility when testing off site and would not take up much space in the lab.

Since this was the first iteration of the mobile version of the emotional chart, we wanted to ensure that it was easy to use and not overburdened with features. Basically, we wanted a replication of the paper format that solved our problems without adding any new constraints. Vital features included the ability to create and save multiple projects, label each question/step for emotion reporting, define how many times each participant would be asked to chart, and stop participants from being able to see their previous answers.

We ran internal studies on Emotrak and quickly determined that more features were needed. With the paper prototype, if we ran out of time to do a task or had to skip a task due to technical issues we just had participants use a different Post-it. However, with the digital version of Emotrak the emotion chart was too linear and participants could not skip a step. This led to reporting a neutral emotion and the researcher manually removing that data during analysis. With studies that had large data sets and many participants, this simply was not feasible so we needed to add the ability to skip questions. We also learned that participants would accidentally select the wrong emotion, which would result in incorrect data so we needed a confirmation when a selection was made.

We also learned that participants tended to consider the left side of the scale as the negative end as our other ratings scales that were being used in conjunction with Emotrak were set up in that way. So we flipped the emotion chart to reflect this and matched the rating scales. When we were using the paper prototype, these aspects were not problems and were not considered as potential affordances in an app version. We continued to do multiple rounds of internal research on the usability of the newly digitized Emotrak. Once we had identified and confirmed the base features that allowed us the same flexibility as the paper prototype, we continued to refine the content within the tool and the notion of self-reporting as a viable method for emotion capturing.

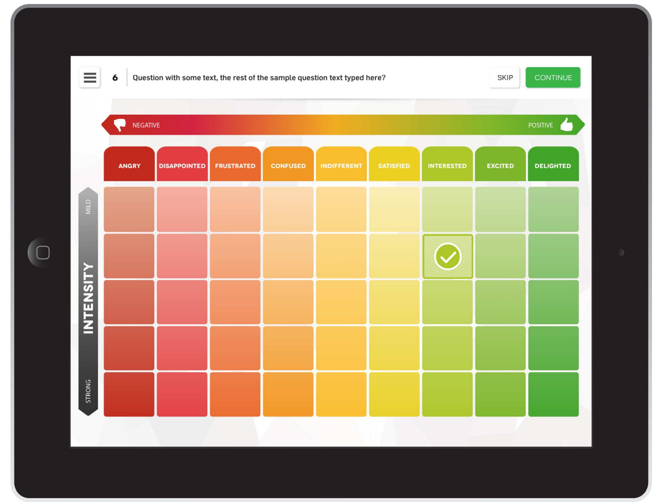

Figure 5: A digital version of Emotrak™, it continued to be refined through multiple rounds of internal testing (© UEGroup)

Validating Self Reporting

After 2 years of using the Emotrak self-reporting method via paper prototypes and exploring the idea of a digital version internally, we found it added incredible value to each research study. Although we still captured user ratings and verbal feedback, the self-reported emotions told a different story and offered a deeper understanding of the product experience. For several studies, we found that although the average user ratings were very similar and did not express nuances that occurred during the study, the emotion data showed which steps of the process or parts of the product elicited frustration or delight.

Figure 6: The emotional journey chart coupled with the participant ratings told a much more nuanced story of the user experience

While we remained confident that the self-reporting method was offering greater insight into user emotions, the big question remained- **is this more or less effective than other emotion capturing methods?** We knew that as an agile and lean UX company, we were not interested in comparing self-reporting to methods that required major financial or time investment. We wanted to compare this methodology to existing products that were being marketed as leading emotion-capturing tools. Since time was of the essence, we were not interested in a longitudinal or quantitative study so we planned a user study with multiple rounds of testing that could be done quickly.

A validation research study was planned, 9 participants were shown emotional stimuli video to determine how they self-reported emotion verbally via the prompt we had been using, "How are you feeling?" and if they felt that they could properly self-report their emotion on the Emotrak tool. User testing sessions were recorded and participant video captures were sent to an emotion video analytics company, Emotient. Since UEGroup was able to record different camera angles and have 2 recordings of the same session from different camera views, both were sent for analysis. Audio was captured and put through a voice analytics service, Moodies Emotion Analytics. We then analyzed this data and compared what participants had self-reported and what the analytic tools reported. We found disparity between these tools, both analytic tools reported vastly different emotions than participants had. Since we uploaded 2 videos of the same participant from different camera angles, we were surprised to see that even these varied from one another with vastly different emotions being reported within the same video clip. The video analysis was also inconclusive for participants who wore a hat or rested their head on their hand. Some participants had to be eliminated from the next round of testing because of these factors.

All the data output from the remaining participants was gathered and 5 participants were asked to return 1 month after the initial study. The emotions that were reported from the video analytics, voice analytics, verbal self-reporting, and Emotrak self-reporting, were collected and randomized. Participants then watched the videos of themselves watching the emotional stimuli from the first round of research. Videos were pared down into small clips so that participants would not be overwhelmed with information and so that they were unable to hear their comments from the first round of research. Participants were then presented with the list of randomized emotion data and chose on a 5 point scale how much they believed that emotion represented how they were feeling during the video from the first round of research.

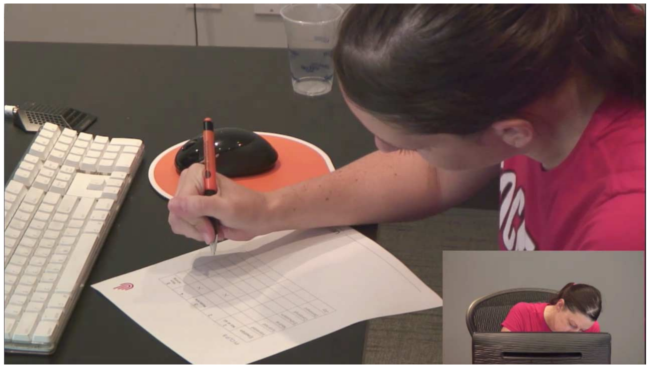

Figure 7: Participants rate how accurate they consider the emotion video analytics, voice analytics, Emotrak, and their verbal response to be

Self Reporting Is More Accurate

Once participants had rated how much the various emotion data represented how they felt during their research, we analyzed the data and averaged the ratings across all 5 participants. Every participant rated the emotions that they had verbally reported and the emotions they had reported via Emotrak higher than the analytic tools. The verbal self-reported emotions ranked slightly higher than the Emotrak application. Participants employed a variety of strategies for determining how they had been feeling during the first round of research, but most noted that they simply recalled how they had been feeling.

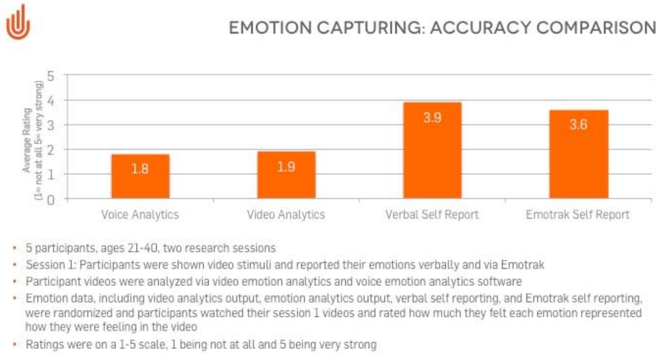

Figure 8: Average user ratings comparing the accuracy of Emotient video analytics, Moodie voice analytics, verbal self reporting, and Emotrak self reporting

While the argument can be made that participants do not necessarily understand the complexities of their own emotions and can not properly express that during the session, our research shows that users are confident when given some parameters for reporting emotion. The reality is that people are making daily choices based on how they interpret their own emotions; whether they interpret their emotion incorrectly it still affects the decision that they make. While biometric data may offer greater insight into the actual emotions of a person, the user's own brain and feelings are what actually drives what they do and how they do it. Through self-reporting via Emotrak, research participants are able to quickly and accurately convey how they feel about a product experience in a way that is easy to quantify and easy to understand. The end result allows researchers and product teams to show a multifaceted emotional journey in a visual and meaningful way through self-reported emotional data.

Next Steps

We are not finished perfecting this tool. We plan on continuing research on the content within it to determine the best terminology for the emotions used within the emotion chart. We plan on testing Emotrak within multiple industries to determine different sets of emotions based on the type of product or research. We plan to conduct a cross-industry card sorting study to accomplish this. We are also planning a beta group to determine best features and have a prioritized list of wanted features to develop and test. We will continue using Emotrak in our research studies and telling the more complex story of a product experience in a simple and inexpensive way.

Acknowledgements

We would like to thank all the moderators and researchers who worked on this tool, the authors Sarah Garcia and Laura Hammond, Denise Su PhD, and Megan McCracken. We would also like to thank all on the design and development teams, Tony Fernandes, Anna Soederstroem, Blair Koeneman, Cygny Malvar, and Fernando Fonseca. Lastly, we want to thank all who participated in user testing sessions, as without the user, user research is not possible.

Reference

[1] Robert Plutchik. 2001. The Nature of Emotions. *American Scientist.* 89, 4. (July-Aug 2001), 344-350. doi: 10.1511/2001.4.344