Evaluation of a self-report system for assessing mood using facial expressions

Hristo Valev[1,2][0000−1111−2222−3333], Tim Leufkens[2][1111−2222−3333−4444], Corina Sas[1][2222−−3333−4444−5555]], Joyce Westerink[2,3][2222−−3333−4444−5555]], and Ron Dotsch[2][2222−−3333−4444−5555]]

[1] Philips Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
{hristo.valev,tim.leufkens,joyce.westerink,ron.dotsch}@philips.com
[2] Lancaster University, Bailrigg Lancaster LA1 4YW, United Kingdom
{h.valev,c.sas}@lancaster.ac.uk
[3] Eindhoven University of Technology Tu/e, 5600 MB Eindhoven, The Netherlands
j.h.d.m.westerink@tue.nl

Abstract. Effective and frequent sampling of mood through self-reports could enable a better understanding of the interplay between mood and events influencing it. To accomplish this, we built a mobile application featuring a sadness-happiness visual analogue scale and a facial expression-based scale. The goal is to evaluate, whether a facial expression based scale could adequately capture mood. The method and mobile application were evaluated with 11 participants. They rated the mood of characters presented in a series of vignettes, using both scales. Participants also completed a user experience survey rating the two assessment methods and the mobile interface. Findings reveal a Pearson's correlation coefficient of 0.97 between the two assessment scales and a stronger preference for the face scale. We conclude with a discussion of the implications of our findings for mood self-assessment and an outline future research.

Keywords: Mood assessment · Self-report system · User Interface

1 Introduction

Different approaches exist that can be used to measure mood, for example, using graphical discrete scales, such as Likert scale, continuous scales such as the visual analogue scale (VAS) or other abstract methods such as colors, pictures, etcetera.

Discrete scales such as the Likert scale [11] or a continuous scale such as the visual analogue scale (VAS) [3] are suitable for mood assessment as they are generally quite intuitive and have been widely used in practice. However, using such scales requires participants to transform the concept of mood onto a numerical or graphical scale. That may result in some information loss, which makes graphical scales less practical for mood assessment. Furthermore, graphical scales have no particular inherent inclination to represent mood [12].

Other approaches of measuring affect are through affective pictures [14, 19, 20], smileys [2, 15], colors [7, ?] or physiological data [17, ?]. Photographic Affect Meter (PAM), for example, is using affective pictures to measure affect. It consists of 16 images, spatially allocated in a two-dimensional space, according to their ratings valence and arousal. The authors in [2] use a discrete scale for valence and arousal represented through icons and sad/happy smileys, for the assessment of arousal and valence, respectively. [7] uses colors to span a two-dimensional emotion space. Different colors represent emotions while color shades represents the intensity of the emotions. All those approaches provide an easily accessible way of reporting mood, however, they are limited to the amount of emotional intensities they provide.

These abstract representations, while very expressive cannot translate well between people as they are highly subjective in nature. In order to have consistency in the measurements, we need a representation which is universally understood by different populations and provides enough variation to describe a broader space.

Facial expressions are inherently linked with emotions and are a visual tool for us to communicate our emotions to the surrounding world. They are embodied representations of our feelings and are as such intrinsically suitable for measuring mood. We are also well versed in using and recognizing facial expressions, which supports the universality of the representation. Research has identified distinct facial expressions, which are associated universally with a specific emotion [5, 6]. For those basic emotions there is a distinctly associated facial expression.

Lorish et al. introduced the concept of using a face scale to measure mood [12]. He argues that facial expressions are tuned to capture and represent mood, because facial feature variations are universal, valid indicators of mood [5, 6]. Kamashita et al. explored the reliability of such scales by comparing them to VAS [9]. The authors evaluated two facial expression-based scales with a VAS scale, which resulted in a 0.68-0.70 correlation between both assessments. Also, in a user experience questionnaire, participants preferred the face scales to VAS scales. This yields the insight, that there might be some interaction quality unique to such scales. Another study conducted by McKinley et al. explored the consistency of a facial expression-based scale [13]. Seven photographs of facial expressions with increasing intensities had to be positioned on a VAS scale. Six out of seven photograph placements were almost equidistant and fell within the expected intervals.

If we are to use and/or improve such a method, we need to make sure that it is reliable in the sense that assessing mood with facial expressions yield at least comparable results to established mood measurement methods and sensitive in the way that assessments provided with such a scale will effectively capture changes in the mood.

Increasing HCI research has focused on the impact of emotions and their awareness on emotional wellbeing and mental health [16]. Such a system would be particularly useful in the context of affective disorders, for example depression. Such conditions are characterized by disturbances in the mood as one of the main

symptoms. Being able to frequently assess a person's mood could potentially provide us with a reasonable estimate of a person's state of well-being and enable, for example, the early detection of depressive episodes.

2 Method

We developed an android application, which features a bipolar sad-to-happy facial expression scale and a VAS scale. The facial expression scale is represented through an image of a face, which can be interacted with to display happier or sadder expressions by sliding your finger vertically along the display (see Fig. 1). The middle point of the scale is the neutral expression. Navigating upwards displays increasingly happier expressions, while downwards – sadder ones. The image space features 101 images, where 50 represent happiness, 50 – sadness and one – the neutral expression. The images were taken from the male facial expressions of sadness–happiness of the dynamic visual analogue mood scales (D-VAMS) project [1]. The scale is conceptualized as a brief, nonverbal mood assessment instrument to be used for self-reporting. A slider with 101 discrete points represents the VAS scale (see Fig. 2). Text anchored on both extremes denotes the respective emotions (i.e. sadness and happiness). Both scales aim to capture the valence of the provided assessment. When providing an assessment, both scales were initialized in the neutral position, i.e. the slider positioned in the middle and the face – to a neutral expression.
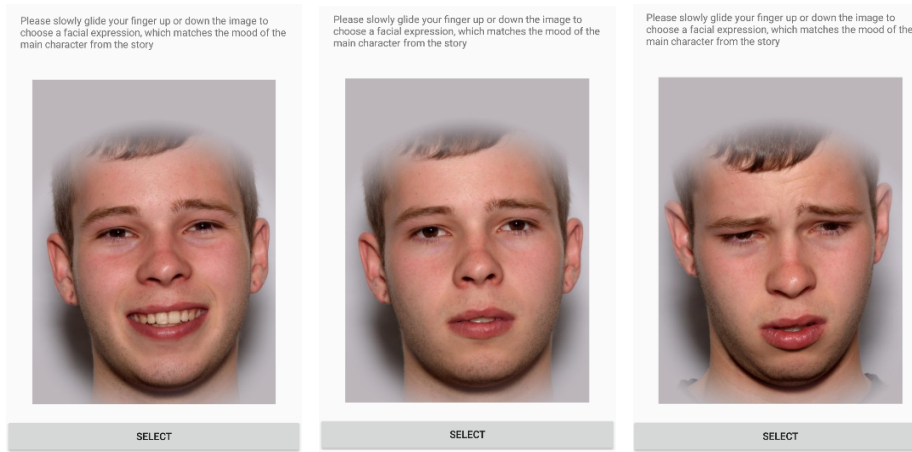

Fig. 1: Screenshots of the D-VAMS face scale assessment from the application [1].

2.1   Participants

We recruited 11 healthy participants via flyers. Eight women and three men took part in the study, with an average age of 29. The participants were recruited from a research environment. They have been handed and signed an informed consent form.

Fig. 2: Screenshot of the VAS scale assessment as taken from the application.

2.2   Assessment

The conducted experiment aimed to evaluate, whether a facial-expression based scale would yield a comparable performance to a VAS scale for mood assessment and whether the user experience between both scales would differ.

Participants were asked to read 30 vignettes and use a smartphone provided by the experimenter for the assessment. Half of the vignettes were taken from [10] and were labeled with a positive emotion. The negative vignettes were collected from various online blogs and forums. The vignettes were paraphrased to portray a story from third persons' perspective.

2.3   Procedure

Before starting the experiment, participants were presented with three training vignettes in order to be acquainted with the system. The results from the training set were omitted from the final dataset. Participants were asked to read each vignette and then use the application to assess the mood of the main actor in the vignette using both VAS and facial-expression based scales. All participants received the vignettes in the same order. The assessments were completed through both scales, presented in a randomized order for each vignette.

At the end, all participants filled out a user experience survey featuring 26 questions. The survey can be found in Appendix A. Eighteen questions evaluated the method and implementation. Those included the ease of use, suitability for mood-assessment, accuracy, satisfaction, user experience, responsiveness, intuitiveness and preference on unipolar Likert scales. Two questions evaluated the preference and speed of both implementations as bipolar Likert scales. Two yes/no questions prompted the participants if they would be able to use the interfaces without instructions. The survey also included four open-ended questions, which inquired about any potential difficulties participants might have had with the application or prompted them to share their insights as to how the assessment can be improved.

3   Results

The data was analyzed using python 3.6 with the numpy and pandas libraries. The plots were created using the seaborn library.
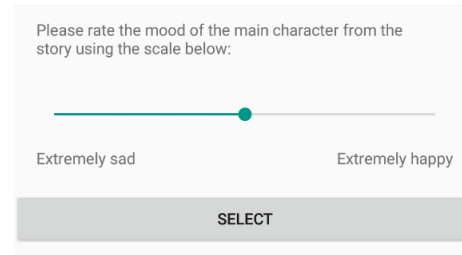
A Pearson's correlation coefficient was calculated between VAS and the facial expression scale assessments, which yielded a 0.97 correlation for all participants. Figure 3 displays the results as a scatterplot, where the assessments obtained from the VAS and facial expression scale are plotted respectively on the Y- and X-axis. The lack of 'neutral' vignettes in the stimulus set explains the sparsity of assessments in the central region of the plot.
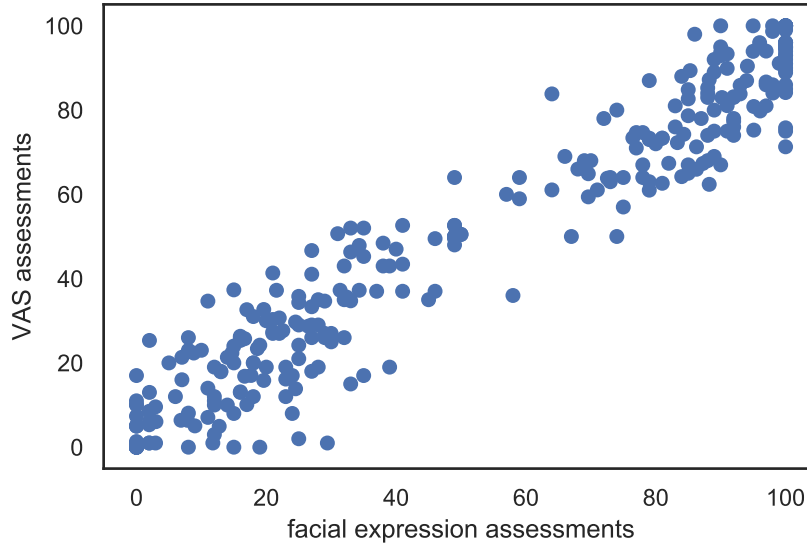

Fig. 3: Scatterplot of the assessments for each vignette and participant for the face scale and VAS

The average time to complete an assessment with the VAS scale was 4.2 seconds, while using the face scale took 5.6 seconds. Figure 4 depicts the relationship between assessment values provided with each interface and the respective duration.

Table 1 features the part of results obtained from the user experience survey, which rated the method and implementation of each scale individually. The questions were represented through a five point Likert scale, where 1 was designated as a low/negative score and 5 – a high/positive one.

Albeit none of the results was statistically significant, due to the relatively low participant count, they still show consistent preference for the face scale on most aspects. Particularly interesting are the noticeable differences in the scores for satisfaction in the method section and user experience in the application section. On both accounts the face scale was preferred to VAS, with only two participants favoring the VAS on both accounts. Both participants left the open-ended questions blank. Four participants found the slider more unresponsive, as
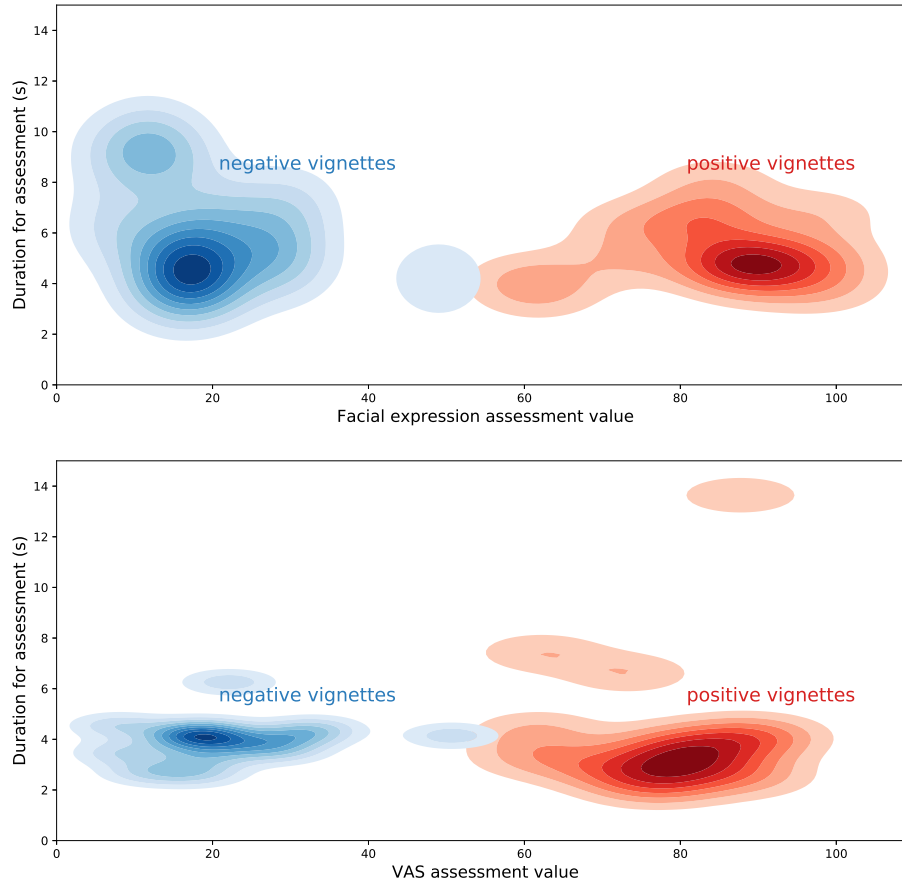
Fig. 4: KDEplot depicting the relationship between assessment values and duration with both interfaces[4]

they would have liked. This would have partially influenced the user experience scores and the speed of assessment for the VAS scale.

Only one participant pointed out, that they would need instructions before using the face scale.

Table 2 features the preferred method of assessment as well as which interface was considered faster for assessing. Those were assessed on a bipolar Likert scale, where 1 favored the face scale and 5 – VAS. The results show that most participants found the VAS scale slightly faster than the face scale. This is also

---

[4] The facial expressions values spread over the maximum value of 100, due to the gaussian kernel estimate used to model the data. The input is the assessment values and their respective durations. For the face scale assessments more often the maximum value was selected (see Figure 3), which causes this effect.

coherent with the results from Figure 4, which established a 1.4 second difference on average for assessments between the VAS and the facial-expression scales.

However, most participants preferred the face scale for mood assessment. Two participants, which preferred the VAS scale in the previous section consistently preferred the VAS scale here as well.

Table 1: Mean (standard deviation), t-value and p-value scores on aspects of the method and application. Rated on a five point Likert scale for negative-to-positive responses.

Method	Face (std)	VAS (std)	t-value(df=10)	p-value
Ease of use	4.09 (1.0)	3.73 (0.96)	-0.83	0.42
Suitability for mood	3.73 (1.29)	3.64 (0.88)	-0.18	0.86
Accuracy	3.73 (0.96)	3.73 (0.62)	0.0	1
Satisfaction	4.0 (1.28)	3.18 (0.72)	-1.77	0.1
Application	Face	VAS	t-value(df=10)	p-value
User experience	4.45 (0.89)	3.73 (1.19)	1.85	0.08
Ease of use	4.18 (1.19)	3.73 (0.75)	1.02	0.32
Responsiveness	4.36 (0.88)	3.6 (0.92)	1.84	0.08
Intuitiveness	4.0 (1.04)	4.09 (0.67)	-0.23	0.82
Assessment preference	3.91 (1.44)	3.73 (0.86)	0.34	0.73

Table 2: Ratings on a 1 to 5 Likert-scale for preference- and speed of assessment. 1 is the maximum value for the facial expression scale and 5 – that for the VAS scale.

Methodology	Mean
Preferred method	2.0
Speed	2.9

Several participants revealed in the open-ended questions section that a simple sadness-happiness scale is insufficient to capture mood for the presented vignettes. One participant shared – *"I think there is more to the emotional spectrum than just happiness or sadness. Other emotions might be relevant to depression as well. Such as fear, disgust, anger, disappointment, frustration, satisfied, grateful, relaxed, nervous, challenged."* Interestingly, one participant pointed out that they liked that the face scale featured a real face instead of a cartoon-like character - *"I like the use of a real person and not a cartoon or smiley-type of representation."*

4 Discussion

First, we would like to acknowledge that the study was conducted as a pilot and is aimed to give us some insight on the proposed assessment method. As several participants pointed out, such an approach featuring only sad and happy facial expressions are not sufficient for true mood assessment. The study was set up to

assess only based on a sadness-happiness scale. An open question remains, how scales featuring multiple mood dimensions would perform. Future research will aim to assess interfaces featuring multiple facial expression and produce a more comprehensive tool for mood-assessment.

The high correlation obtained from both assessments points to a high consistency of results with an already established mood measurement method such as VAS. Surprisingly, this is despite the fact complex emotions, such as awe or compassion were present in the vignettes. We acknowledge that the vignettes were presented in the same order for all participants, which might have introduced a carry-over effect. This effect, however, would be consistently present in both assessments. The randomized order in which both scales were presented after each vignette ensured that participants would not be able to 'seek out' the corresponding value on the latter. Furthermore, the facial expression scale provided no numerical reference as to what value is currently selected. This made it more difficult to simply carry over values from one scale to another. The design, unfortunately, does not allow to establish whether either scale 'outperforms' the other. This is due to the mismatch of the emotions portrayed by the vignettes and the dimensions available on the scales. Furthermore, the negative vignettes have not been rated. It will be interesting, however, to evaluate a multidimensional facial-expression based scale with a validated set of stimuli. Such an approach could provide some insights as to how sensitive and accurate a facial expression-based scale is in capturing mood.

The slightly faster average time it took for each VAS assessment can be attributed to the scale space being completely visible. The participants could immediately select a value lying on the extremes, while the face scale needed to be 'browsed'. As the provided stimuli were emotionally charged, most of the assessments veered away from the neutral expression. Figure 4 visualizes the average time per vignette it took to complete an assessment with each scale with respect to the duration. Despite the fact that the facial expression scale had to be navigated, this didn't influence assessment time as there is no pronounced relationship, which links longer assessment times with assessments lying on the extremities of the scale. This means that the interface could be easily navigated, yielded negligible slowdown and hints that the scale can be used for frequent assessments. A potential application for this method would be as an ecological momentary assessment (EMA) tool [18]. A longitudinal approach employing such a scale might reveal if such a scale would be viable if it is to be used as frequently as multiple times per day.

Most participants preferred the face scale, despite the slightly longer time required to provide an assessment; however, some still found the VAS scale to be more adequate for mood-assessment. The face scale was preferred to VAS on most accounts. This could be due to the scale providing a better interaction experience or due to a 'novelty' factor. A real-world application would reveal if the preference for such a scale would remain if it is used daily.

It would also be interesting how such a scale would perform in a clinical population. It is known that clinical populations have an attentional bias towards

sadder-looking faces and perceive more negative expression in ambiguous faces [8, 4]. The implications of such a use case could result in more frequent and reliable mood-tracking, which could open up opportunities for the design of intervention systems. Such an approach could be further augmented by sensor data and enable a more comprehensive monitoring of patients.

5   Conclusion

This pilot study shows that assessing mood with a face scale provides similar results as assessing mood with a visual analogue scale. Additionally, most participants indicated to prefer a face scale to a visual analogue scale. The way the user interface was conceptualized resulted in slightly longer times required for assessment with a facial-expression based scale. However, most participants preferred such a scale in terms of ease of use, user experience and satisfaction.

6   Acknowledgment

References

1. Barrows, P.D., Thomas, S.A.: Assessment of mood in aphasia following stroke: validation of the Dynamic Visual Analogue Mood Scales (D-VAMS). Clinical Rehabilitation **32**(1), 94–102 (2018). https://doi.org/10.1177/0269215517714590
2. Broekens, J., Brinkman, W.P.: AffectButton: A method for reliable and valid affective self-report. International Journal of Human Computer Studies **71**(6), 641–667 (2013). https://doi.org/10.1016/j.ijhcs.2013.02.003, http://dx.doi.org/10.1016/j.ijhcs.2013.02.003
3. DD, R., PA, M., A, R., B., B.: The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain **17**(1), 45–56 (1983)
4. Duque, A., Vázquez, C.: Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study. Journal of Behavior Therapy and Experimental Psychiatry **46**, 107–114 (2015)
5. Ekman, P.: Universal Facial Expressions of Emotions (1970)
6. Ekman, P.: Facial expression and emotion. American Psychologist **48**(4), 384–392 (1993). https://doi.org/10.1037/0003-066X.48.4.384, http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.48.4.384
7. Huang, S.T.Y., Kwan, C.M.Y., Sano, A.: The moment. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct pp. 235–238 (2014). https://doi.org/10.1145/2638728.2638784, http://dl.acm.org/citation.cfm?doid=2638728.2638784
8. Joormann, J., Gotlib, I.H.: Is this happiness i see? Biases in the identification of emotional facial expressions in depression and social phobia. Journal of Abnormal Psychology **115**(4), 705–714 (2006). https://doi.org/10.1037/0021-843X.115.4.705

9. Kamashita, Y., Sonoda, T., Kamada, Y., Nishi, Y., Nagaoka, E.: Reliability , Validity , and Preference of an Original Faces Scale for Assessing the Mood of Patients with Dentures pp. 93–98 (2007)
10. Lagotte, A.: Eliciting Discrete Positive Emotions with Vignettes and Films: A Validation Study (2014), http://etd.library.vanderbilt.edu/available/etd-07172014-202956/unrestricted/LagotteAE_MastersThesis.pdf
11. Likert, R.: A technique for the measurement of attitudes. Archives of psychology (1932)
12. Lorish, C.D., Maisiak, R.: The face scale: A brief, nonverbal method for assessing patient mood. Arthritis & Rheumatism **29**(7), 906–909 (1986). https://doi.org/10.1002/art.1780290714
13. McKinley, S., Coote, K., J, S.P.: Development and Testing of a Faces Scale for the Assessment of Anxiety in Critical Ill Patients. J Adv Nurs **41**(1), 73–9 (2003)
14. Pollak, J., Adams, P., Gay, G.: PAM: a photographic affect meter for frequent, in situ measurement of affect. Proceedings of the SIGCHI Conference on . . . pp. 725–734 (2011). https://doi.org/10.1145/1978942.1979047
15. Rodriguez, I., Herskovic, V., Fuentes, C., Campos, M.: B-ePain: a wearable interface to self-report pain and emotions. UbiComp Adjunct pp. 1120–1125 (2016). https://doi.org/10.1145/2968219.2972719, http://dblp.uni-trier.de/db/conf/huc/ubicomp2016ap.html#RodriguezHFC16
16. Sanches, P., Janson, A., Karpashevich, P., Nadal, C., Qu, C., Dauden Roquet, C., Umair, M., Windlin, C., Doherty, G., Hook, K., Others: HCI and Affective Health: Taking stock of a decade of studies and charting future research directions (2018)
17. Sas, C., Rees, M.: AffectCam : Arousal – Augmented SenseCam for Richer Recall of Episodic Memories. Chi 2013 pp. 1041–1046 (2013). https://doi.org/10.1145/2468356.2468542, http://dl.acm.org/citation.cfm?doid=2468356.2468542
18. Stone, A.A., Shiffman, S.: Ecological Momentary Assessment (Ema) in Behavioral Medicine. Annals of Behavioral Medicine **16**(3), 199–202 (1994). https://doi.org/10.1093/abm/16.3.199, https://dx.doi.org/10.1093/abm/16.3.199
19. Watson, D., Clark, L.A.: PANAS-X Manual pp. 1–27 (1999). https://doi.org/10.1111/j.1742-4658.2010.07754.x
20. Watson, D., Clark, L.A.: JPSP Watson Clark Tellegen 1988 **54**(6), 1–8 (2004). https://doi.org/10.1051/epjconf/201714006017, papers3://publication/uuid/CA1FC7DA-0BA7-4563-8673-5E3942A0983F

A   Appendix A

UX Survey, Participant-ID: _____

Participant information						
What is your age?						
What is your gender?	(M)	☐			☐	(F)

Method-related questions						
How would you rate the ease of use of the method working with the slider?	(difficult)	☐ ☐ ☐ ☐ ☐	(easy)
How would you rate the ease of use of the method working with the image?	(difficult)	☐ ☐ ☐ ☐ ☐	(easy)
How suitable was the slider for capturing mood?	(not suitable)	☐ ☐ ☐ ☐ ☐	(suitable)
How suitable was the image for capturing mood?	(not suitable)	☐ ☐ ☐ ☐ ☐	(suitable)
How accurate do you consider the slider is in capturing the mood?	(not accurate)	☐ ☐ ☐ ☐ ☐	(accurate)
How accurate do you consider the image is in capturing the mood?	(not accurate)	☐ ☐ ☐ ☐ ☐	(accurate)
How satisfying was the slider to work with?	(very dissatisfying)	☐ ☐ ☐ ☐ ☐	(very satisfying)
How satisfying was the image to work with?	(very dissatisfying)	☐ ☐ ☐ ☐ ☐	(very satisfying)
Which assessment method would you personally prefer?	(image)	☐ ☐ ☐ ☐ ☐	(slider)
Which method, according to you, was faster to use for mood assessment?	(image)	☐ ☐ ☐ ☐ ☐	(slider)
Do you have any comments regarding either of the mood capturing methods?			

Application-related questions						
What was your experience using the image within the application?	(negative)	☐ ☐ ☐ ☐ ☐	(positive)
What was your experience using the slider within the application?	(negative)	☐ ☐ ☐ ☐ ☐	(positive)
How easy to use did you find working with the image within the application?	(difficult)	☐ ☐ ☐ ☐ ☐	(easy)
How easy to use did you find working with the slider within the application?	(difficult)	☐ ☐ ☐ ☐ ☐	(easy)
How responsive was the image assessment within the application?	(unresponsive)	☐ ☐ ☐ ☐ ☐	(responsive)
How responsive was the slider assessment within the application?	(unresponsive)	☐ ☐ ☐ ☐ ☐	(responsive)
How intuitive do you consider working with the image within the application is?	(not intuitive)	☐ ☐ ☐ ☐ ☐	(very intuitive)
How intuitive do you consider working with the slider within the application is?	(not intuitive)	☐ ☐ ☐ ☐ ☐	(very intuitive)
Do you think you would use the image for mood assessment?	(unlikely)	☐ ☐ ☐ ☐ ☐	(likely)
Do you think you would use the slider for mood assessment?	(unlikely)	☐ ☐ ☐ ☐ ☐	(likely)
Do you think you can use the image approach within the application for assessing mood without a clarification from a technical person?	(no)	☐			☐	(yes)
Do you think you can use the slider approach within the application for assessing mood without a clarification from a technical person?	(no)	☐			☐	(yes)
Is there some part of the functionality, which didn't work for you?			
Did you encounter any technical difficulties, while working with the application?			
Do you have any comments regarding the application?			