Can reactivity to experience sampling be used to decrease alexithymia?

B.Sc. Thesis

Lukas Bern Libbertz
January 2019

Supervisors:
Dr. M.L. Noordzij
Dr. A.M. Sools

Faculty of Behavioural, Management and Social Sciences

University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE.

Abstract

The market for consumer wearable technology is forecasted with a tremendous growth. Among such wearables are products that promise their users better emotional and psychological wellbeing based on physiologic measurements. These consumer wearables are of special interest for psychologic and health research, as well as for treatment. Wearables facilitate the gathering of experience sampling data as participants progress through their daily routine. This way, data is gathered as close to the events researchers are interested in as possible and the effects of recollection biases are limited. Another interesting aspect is that participants are inevitably reactive towards being measured and in research contexts. From prior experience sampling research, it is known that reactivity is present in experience sampling studies, as participants become more aware and reflect more on the topics that they gather data on. Studies in the field of substance abuse already used reactivity effects to make participants aware of their problematic consumption behaviour, achieving positive outcomes of comparable magnitude as other short treatment forms. Therefore, this study aims to contribute to the research in reactivity in experience sampling by testing whether self-assessment and self-measurement effects have the potential to alter emotional self-reflection in participants. Verification of such an effect could help to design future psychology intervention methods that apply experience sampling as treatment. In a one-week emotional reflection experiment using the experience sampled data of 28 wearable users, their scores on the Toronto Alexithymia Scale 20 were compared to scores of a control group of 14 individuals to test for such an effect. It was also tested whether stronger adherence to the experience sampling method had an effect on post-experiment TAS-20 scores. However, the results of this study provide no support for self-assessment effects affecting emotional reflection as measured by the TAS-20. Results of prior studies are discussed, as claims of a treatment like effect due to reactivity could not be replicated. The paper concludes with the recommendation that reactivity can be seen as a helpful contributor to interventions, yet for treatment, the focus should be placed on accompanying exercises.

Table of contents

Introduction

In recent years, more and more technological solutions are created that promise their users deeper insights in mental and bodily processes with claims that such knowledge is advantageous for personal development. Such promises seem to appeal to consumers as the wearable market is forecasted to grow tremendously, from 325 million devices in 2016 to 830 million devices in 2020 (Statista). Examples of such technologies are smart wearable devices such as the Feel Wristband, Spire Stone and Spire Health Tag. Their producers advertise these products with claims that knowledge about emotional and physiological states would contribute to improvements in emotional well-being for their users (Feel, Spire). Yet, as the common saying holds: 'information is not transformation'. However, a possible explanation for such an effect could be that measuring oneself already possesses the potential to change behaviour or cognition. In fact, there are claims that reflecting on personal data may change the perception of personal circumstances that the one measuring them might have (French & Sutton, 2010; Meier, Miller, Lombardi, & Leffingwell, 2017; Rowan et al., 2007; Sutton et al., 2014). This was found to have the potential to change participant behaviour, for example by decreasing risky drinking behaviour or increase the ability of individuals for emotional self-reflection (Meier et al., 2017; Widdershoven et al., 2019). Building on that, the aim of the study was to investigate, whether such an effect can occur in participants that followed a one-week wearable study, as they measure and reflect on their experience sampled data.

Participant reactivity

Inevitably, the simple awareness of participants being research subjects, being measured or being in a research setting affects their performance (Capellan, Wilde, & Zhang, 2017; French & Sutton, 2010; Kypri, Langley, Saunders, & Cashell-Smith, 2007). This effect is called reactivity. The prototypical example of reactivity is the Hawthorne effect. The name is derived from the Hawthorne works in Chicago, where between 1924 to 1933 studies were undertaken on worker productivity. Different levels of lighting and their effect on productivity were assessed, though it was found that turning down the brightness even to "moonlight" levels, still increased productivity (Olson, Hogan, & Santos, 2005). Later on, it was found that workers showed an increase in productivity, partly due to the strong supervision of managing and research personnel that conducted the study and not due to the manipulations in lighting (McCambridge, Witton, & Elbourne, 2014; Olson et al., 2005). This example points out two properties of reactivity that this study wants to focus on. Firstly, that reactivity is a phenomenon capable of changing behaviour in humans and secondly that reactivity can potentially lead to desirable outcomes, such as a productivity increases in the Hawthorne example. To this day, reactivity is an acknowledged phenomenon that is referred to as Hawthorne effect, measurement reactivity or assessment reactivity (Capellan et al., 2017; Olson et al., 2005; Schrimsher & Filtz, 2011). However, the reactivity effect as it is spoken of here refers to individuals being assessed by others. Yet, to test for reactivity in wearable users as they measure themselves and reflect on their measurements, it is necessary to take a look whether self-assessment, self-reporting and the reflection thereupon, has the potential to be reactive.

Reactivity in experience sampling

By self-assessment and self-reporting of data in psychological research concerning daily life it is usually referred to the experience sampling method (ESM). ESM is a sampling method in which individuals are asked to report on various elements of their life, like emotions, activities, moods and symptoms in a structured fashion multiple times a day (Berkel, Ferreira, & Kostakos, 2017; Myin-Germeys et al., 2018). Often these self-reports are given by answering short and identical questionnaires throughout the day, or report on the assessed elements in a diary-like fashion (Berkel et al., 2017; Myin-Germeys et al., 2018) Strengths of this method are that it allows for the caption of data close to the occurrence of the phenomena under investigation as well as that they can be done in real-world contexts as participants go through their daily routines (Barrett & Barrett, 2001; Shiffman, Stone, & Hufford, 2008). These advantages make them a valuable method in gathering real-world data that may otherwise have been distorted by recall biases or simply be forgotten by the participants when they have to report on them retrospectively (Barrett & Barrett, 2001; Ebner-Priemer & Trull, 2009; Myin-Germeys et al., 2018; Shiffman et al., 2008).

With the wide adoption of smartphones and wearables, the ease and potential of carrying out experience sampling studies is increased (Berkel et al., 2017; Morris et al., 2010; Myin-Germeys et al., 2018). Especially, since smart phones are widely adopted nowadays it is less of a cost factor for the researchers to implement their participants smartphones in the study instead of handing out journals or beepers (Morris et al., 2010). Next to that, phones and wearables are equipped with a variety of sensors that can incorporate a wide range of different measurements like stress levels, changes in heart rate, sleep patterns and physical activity (Berkel et al., 2017; Morris et al., 2010; Nelson, Verhagen, & Noordzij, 2016) that provide much information of the context of its wearer.

However, letting participants focus and reflect on the assessed phenomena may change measurement outcomes by itself, due to earlier mentioned reactivity effects. An example of this are the findings of Kypri (as cited in Kypri et al., 2007) reporting on a study where problem drinkers had to monitor their drinking behaviour in a drinking diary. Several individuals of this study were found to report that this diary brought their excessive drinking behaviour to their attention (Kypri et al., 2007 citing: Kypri, 2002). Similar effects were since then reported for self-reports in smoking cessation (Rowan et al., 2007) and self-reports leading to a small increase in adherence to medication (Sutton et al., 2014) as well as improvements on emotion differentiation through experience sampling in individuals suffering from depression (Widdershoven et al., 2019). It is acknowledged that reactivity effects are beneficial in drug and alcohol treatment (Kypri et al., 2007; Meier et al., 2017; Schrimsher & Filtz, 2011) and some even hypothesize that reactivity might serve as a treatment in its own (Schrimsher & Filtz, 2011). However, some report that self-report reactivity overall, is not well researched yet (Meier et al., 2017).

Alexithymia as measurement of emotional self-reflection

When it comes to self-report on emotions, experience sampling has proven itself to be a valuable tool (Hernandez et al., 2016; Kashdan, Barrett, & McKnight, 2015; Morris et al., 2010; Widdershoven et al., 2019). Yet, reporting on personal emotions requires an individual to be aware of them and being able to communicate emotional states, which is not as self-evident as one might think. In fact, being unable to identify and communicate personal emotions is a severe threat to emotional wellbeing and it is what defines the construct of

alexithymia (Bagby et al., 1994). Originally, it was defined by Nemiah, Freyberger and Sifneos (1976, as cited in Taylor 2000). The characteristics of alexithymia are said to be difficulties in identifying and describing personal emotions, difficulties in holding apart bodily sensations stemming from emotional arousal or feelings, reduced imagination and fantasy and an overly externally oriented way of thinking. Yet, alexithymia is not a psychopathology by itself, but a construct containing many cognitive and affective characteristics that were observed in classic psychopathologies and substance abusers (Taylor, 2000), which makes it a transdiagnostic factor. It was found to be comorbid with several psychological disorders such as depression (Parolin et al., 2018; Viganò et al., 2018), anxiety (Viganò et al., 2018), substance abuse (Parolin et al., 2018), and eating disorders (Altamura et al., 2018). A further implication is that an inability to communicate personal emotional distress makes it difficult to receive help (Taylor, 2000) and therefore implicates treatment (Altamura et al., 2018).

Alexithymia was shown to be reducible by mindfulness-based treatments and other treatments that enhance emotional awareness (de Groot & Rodin, 1997; Haase et al., 2015; Norman, Marzano, Coulson, & Oskis, 2018; Santarnecchi et al., 2014). Additionally, a reduction of alexithymia was not only shown to be beneficial for patients suffering from psychopathologies as it was shown to benefit non-clinical patients as well. It was proven that reduced alexithymia helped patients suffering from coronary heart disease that had no former history of psychopathologies (Beresnevaite, 2000). In her study, Beresnevaite successfully decreased the scores on the alexithymia construct in her participants as they followed a 4-month study involving group psychotherapy. The alexithymia construct was measured with the Toronto Alexithymia Scale 20, a twenty item Likert-scale questionnaire indicating a likely presence of alexithymia when scores are higher than 51 (Bagby et al., 1994). After the study, the scores on alexithymia decreased from 70.8 (SD=5.5) to 62.8 (SD=9.8). In yet another study, a decrease in alexithymia helped elite-athletes to greater adaptation to stressful situations (Haase, et al., 2015). It is also proposed to reduce alexithymia as a preventive measure in non-clinical populations (Norman et al., 2018).

However, as alexithymia refers to the ability of identifying and communicating personal emotional states, it hast to be pointed out that the common lay terminology of emotions is not as well suited to distinguish emotions as one might think. Russel (2003) points out, that emotions, in the sense that they are spoken of, are laden with folk theoretical assumptions as if the distinct differences between emotions such as fear or anger can be easily identified. Though, when it comes to scientific measurements there is not much evidence that each individual emotion would possess e.g. a unique physiological pattern that is elicited when the specific emotion comes into action (Feldman Barret, 2006; Cacioppo, et al., 2000). So, if physiologic responses do not provide accurate means to differentiate between emotions as people speak of them, then how can individuals be expected to make sense of their physiology and label it accurately?

Core-affect

A possible solution to the problem described above comes by the proposed core affect model by Russell (2003; 2009). In an earlier version it was referred to, and still is, as the circumplex model of affect (Hernandez et al., 2016; Morris et al., 2010; Russell, 1980). It has to be noted that the circumplex model of affect and the core-affect model are not virtually identical, yet, as the core-affect model is the predecessor of the circumplex model, further distinctions will not be made here (Russell, 1980, 2003, 2009).

Core affect is said to be a universal, simple and primitive concept of emotional state in the sense that it can exist without being labelled, interpreted or needed to be attributed to a cause (Russell, 2003). It is a changeable, neurophysiological state that is always present in an individual and potentially consciously accessible. Core-affect represents the simplest raw emotional state that is present at any moment. Once an individual becomes aware of their core-affect, it can be described as moods and emotions (Russell, 2003, 2009). Therefore, it is lying beneath common emotional terms and concepts such as anger or joy and it manifests into emotional descriptions when attention is directed towards it. Core-affect is said to be located on two dimensions. The horizontal dimension is called valence and it refers to whether an affective state at a given moment would be perceived as pleasurable or displeasuring if it were to enter the consciousness (Russell, 2003, 2009). The vertical dimension, called arousal, describes whether an affective state in a given moment would be perceived as activating or deactivating in terms of mobilization energy if one were to redirect the attention on personal emotional states (Russell, 2003, 2009). The concept holds that at a given moment, the core-affect of an individual is always made up by both dimensions and represented as a mark on the coordinate system (Russell, 2003, 2009). Depending on where the current core-affect is placed on the two dimensions, core-affect can be labelled with everyday emotional terms (Figure 1.).

In the past, wearable and smartphone studies that assessed emotions in a self-report fashion made use of the circumplex model of affect (Hernandez et al., 2016; Morris et al., 2010). Both these studies used a two-dimensional axis model, as proposed by Russel (1980; 2003), and implemented it on participants phones to use it as a self-report, experience sampling questionnaire that participants had to fill in multiple times a day. In the study of Morris et al. (2010) eight participants were equipped with a mobile phone application for mood reporting that also featured prompts for cognitive reappraisal and physical relaxation exercises that they had to attend to in a one-month field study. The study features an extensive report on each participant as they progressed through the study and notes that participants perceived the axis model as useful to 'check in with myself', as sparking curiosity in personal moods as well as increasing the confidence in personal feelings. Participants felt, that it would lead to better self-understanding. Generally, it is noted that the model was seen as easy to understand and apply (Morris et al., 2010).

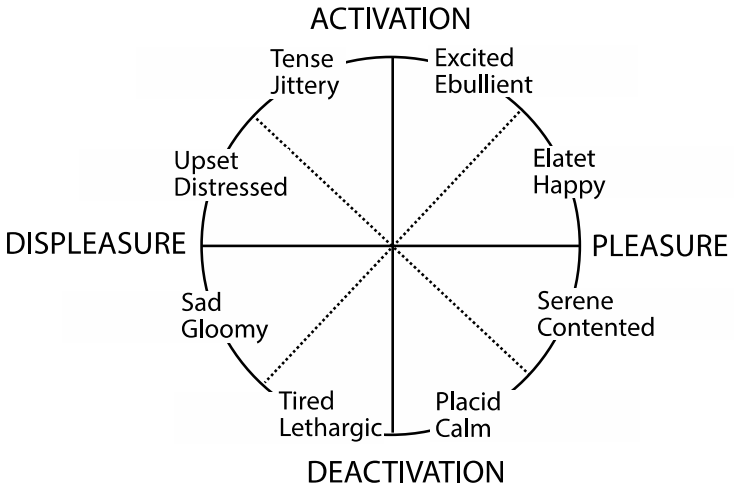

Figure 1. Core-affect model as it was proposed by Russel (2003)

Research question and hypotheses

As reactivity was shown to be almost inevitably present in research with human participants and as experience sampling seems a further facilitator of reactivity, it was assumed that reactivity will also affect emotional self-reflection. Next to that, the construct of alexithymia was pointed out to be a capable construct to measure the ability of an individual to engage with personal emotions and make reason of physiological and emotional states. Due to the comorbidity of alexithymia with numerous psychopathologies and other threats to mental health, a reduction in alexithymia can be considered an improvement of emotional and mental wellbeing. Since other studies already effectively used the core-affect model in wearable and smartphone-based studies, it was used in this one as well. Subsequently, to test the initial research question whether a possible reactivity effect in a one-week reflection experiment on self-reported core-affect can reduce alexithymia scores in wearable-users, the following hypotheses were tested.

Hypothesis 1: Pre-experiment TAS-20 scores are higher than post-experiment scores in the experiment group.

Hypothesis 2: Participants in the experiment group that completed more surveys have a lower post-experiment TAS-20 score than participants who completed fewer surveys.

Hypothesis 3: Pre-experiment TAS-20 scores are not significantly different from post-experiment TAS-20 scores in the control group.

Method

Participants

There were 53 participants in total. The data of 18 participants was used from prior iterations of the study. The 35 participants of this iteration of the study were recruited mainly through convenience sampling and some via SONA Systems of the University of Twente. SONA is an internet platform of the University of Twente in which students of Psychology and Communication studies can apply as participants for research projects to obtain course credits. For this study, the SONA credit reward was 5.25 points. Participants that were recruited outside of SONA did not receive a reward. A requirement for participation in this study was, that the participants owned or at least had constant access to a smart phone. A requirement for participation in the control group of the study was that participants did not use a smart watch or wearable together with mindfulness apps during the duration of the experiment.

The mean age of the participants was 26 years (SD=10), with a minimum age of 19 and a maximum of 70 years. Among the 53 participants there were 31 males making up for 58.5% and 22 females, making up for 41.5%. Most participants were German with 28 participants in total. Then there were 7 Dutch participants and 18 of whom the nationality was unknown. The large number of unknown nationalities is due to the fact that for participants of prior iterations, nationality was not recorded. However, it is likely that they were mainly German or Dutch as well, since the University of Twente is a Dutch university with a high influx of German students. The data of 11 participants was not included in the study. For the experimental group, the data of 5 participants was excluded because it was not present entirely or because they did not complete enough TiiM-surveys. Cut-off criteria for TiiM-surveys was that participants had to fill in at least 20 surveys, though 28 was the goal number. For the control group the data of 6 participants was excluded, because they did not send back the second questionnaire or because they did not fill out the entire questionnaire. All in all, the data of 28 participants was used for the experimental group and the data of 14 participants was used for the control-group.

The study was approved by the Ethics committee of the University of Twente. Before data collection was started, every participant was asked to fill in an informed consent form (see Appendix A).

Materials

Twenty-Item Toronto Alexithymia Scale (TAS-20)

The TAS-20 is a self-report instrument for measuring alexithymia by assessing the ability of identifying and describing personal emotions (Bagby, Parker, & Taylor, 1994). The TAS-20 was published in 1992 and is a revised version of the 26-item Toronto Alexithymia Scale, which was initially published in 1985. The TAS-20 consists of three subscales (Bagby, Taylor, & Parker, 1994). The first subscale consists of 5 items (item numbers: 2, 4, 11, 12, 17) and measures the ability of an individual in describing one's own emotions. The second subscale has 7 items (item numbers: 1, 3 ,6 ,7,9 ,13 ,14) and measures the ability of identifying personal emotions. The third subscale consists of 8 items (item numbers: 5, 8, 10, 15, 16, 18, 19, 20) and it measures an individual's tendency in focussing their thinking externally. The

TAS-20 follows a 5-point Likert scale where a 1 corresponds to completely disagree and a 5 corresponds to completely agree. The scores on items 4, 5, 10, 18 and 19 have to be reversed. The alexithymia score is the sum of scores of all 20 items and the score for each subscale is the sum of scores for the items corresponding to each subscale. Total scores of less than 51 indicate no presence of alexithymia, scores of higher than 51 to 60 indicate a possible alexithymia and scores of 61 and above indicate alexithymia. In the past, the TAS-20 was demonstrated to have a good internal consistency (Cronbach's alpha=.81) and a good test-retest variability (.77, p<.01). The TAS-20 is said to have an adequate concurrent and convergent validity and the general 3 factor structure was shown to be congruent with the alexithymia construct (ACBS). However, Preece et al. (2017, 2018) have some criticism on the TAS-20. They point out that the third subscale has a poor factor loading (<.40) and it should be considered to exclude it. The TAS-20 had been found stable and replicable in both clinical and nonclinical populations (ACBS). The questionnaire form used in this study (see Appendix B) was derived from the former Cognitive Behaviour Therapy and Science Center, USA (cbtscience).

The Incredible Intervention Machine (TIIM)

TIIM is a mobile application that can be downloaded for mobile devices such as smartphones and tablets for both Android as well as iOS. It can be found on appadvice.com or it can be downloaded via Google Play or the Appstore. The TIIM-App was created by the University of Twente's BMS-Lab and was released on the 1.11.2017. The research was conducted using version 1.3, which was released on 21. of September 2018. TIIM is used as a mobile platform to host interventions to give researchers the opportunity to install questionnaires or other experiment and intervention material on the participants phone. This way, participants can be contacted regularly to provide tasks such as filling out questionnaires.

For this study, the participants had to mark a position on a coordinate system ranging from low in energy to high in energy on vertical scale and from unpleasant-pleasant on the vertical scale. The coordinate system was filled in by placing the yellow marker in the coordinate system (Figure 2), that would describe the emotional state of the participant the best. The blue fields surrounding the axes featured terms of emotions that served as examples to help participants in placing the marker. The coordinate system was adopted from the core-affect model of emotion (Kuppens, Oravecz, & Tuerlinckx, 2010). Lastly, a note for iOS devices. At the time this research was conducted, users of iOS devices had problems moving the yellow marker on their screen as the TIIM-App presumably had a flaw when it came to iOS devices. iOS users had to use both hands, one to hold the screen to prevent it from scrolling up and down and one to place the marker. Participants using iOS devices were informed on this issue before the research started, to prevent mistakes.

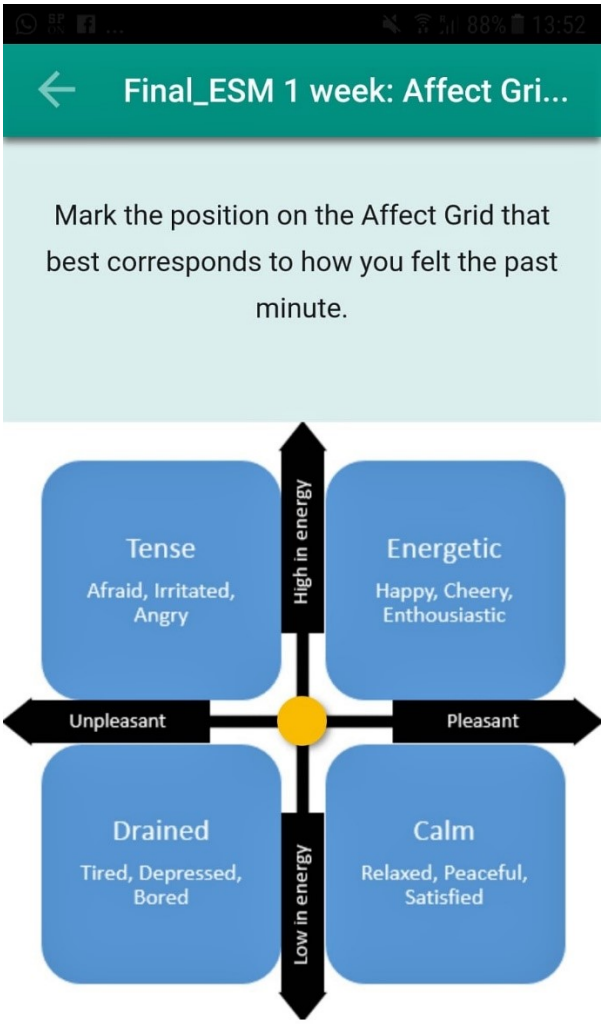

Figure 2. Derived from a screenshot made from the questionnaire as it appeared to participants during the survey. The screenshot shows the adaptation of the core-affect model (Russel, 2003), as it was handled by this study. The terms in the blue fields indicate what emotions would typically fall in this area and served as examples to the participants.

E4 Wearable, E4 Manager and E4 Connect

The E4 wearable and E4 related programs were no necessity for this subpart of the study, yet they still were vital to the entire study at a whole and mandatory equipment for the experimental group participants. However, they will only be explained briefly here. The E4 wearable, E4 Manager and E4 Connect are distributed by Empatica Inc. in Cambridge, US and Empatica Srl. in Milano, Italy. The E4 is a watch-like wearable featuring a heartrate sensor, a sensor for electrodermal activity, a 3-axis accelerometer to measure movement and a sensor to measure body temperature. The data is saved inside on a flash drive and can be transferred via a micro-USB port to a computer. Data synchronization between E4 wearable and a computer requires the program E4 Manager. The E4 Manager can be downloaded from the Empatica website for free. During this study, version 2.0.1. was used. After the synchronization, the data is stored in separate sessions and uploaded into the E4 Connect cloud. A session is created for each time when the E4 wearable is switched on to start measuring and switched off to end the recording. The cloud service E4 Connect is another product of Empatica Inc. and serves as the online storage for the synchronized data. The

sessions are uploaded automatically, once they are synchronized. E4 Connect is hosted by Amazon Web Services (AWS) that provide safeguards and security from unauthorized access, use and modification of the stored data. The website can show graphical depictions of the recorded data. For a more extensive overview on the E4 wearable, E4 Manager and E4 Connect, Empatica Inc. has a handbook for the E4 wearable and accompanying programs available on their website for free (Empatica).

Design and Procedure

The research featured a longitudinal experience sampling study with a within-subject design for the experimental group, a questionnaire design for control and experimental group and a between group design. For the experimental group, duration of the study was seven days during which the physiologic data was sampled passively and continuously via the E4 wristband and sensors. The psychological data was sampled in self-report fashion by the experimental group participants, which were additionally tasked to fill in a brief TiiM questionnaire every two hours in their waking time. Both groups had to fill in the TAS-20 before and after the 7-day study. For the control-group, this was the only task they had in the study.

Participants of the experimental group received a brief initial description of the experiment either directly by the researchers or via the SONA website. If they were interested to participate, a day and time was set to meet up for further explanations and to start the experiment once they gave written consent. Usually these meetings took place at rented rooms in the University of Twente or at the home of the participants. These initial meetings took between 45-60 minutes. The meetings started with exchanging greetings and the researchers thanking the participants for their interest in the study. Then participant and researcher sat down to a quiet table, making sure that possible distractions were avoided. If the meetings took place in a university room, the researcher made sure in advance that the necessary materials like computer, TAS-20 questionnaires, consent forms, E4 wearable, related programs and materials, were already set up and ready for usage. If the meeting took place at the participants house, the material was presented in the order it was explained to him or her. The explanation started by giving the participant a broader overview over the experiment, its purpose, what the participant could expect of it and what was expected of him or her. Namely, to wear the E4 wearable for 7 days in a row and fill in as many TiiM-surveys, occurring every two hours in a day, as was possible during their waking time. Minimum wearing time was between 10 am and 18 pm. Synchronization, re-charging and handling of the E4 wearable was explained to the participant, though will not be described here as it was not elementary to this subpart of the study.

The TiiM-App was explained to the participant and if the participant used iOS devices, the notification was given on how the participant could avoid the scrolling issue by using both hands. When the participant agreed to take part in the study, the consent form was handed out to sign it. If it was signed, the experiment began with the participant filling in the first TAS-20 and after that, the researcher showed the participant how to download the TiiM-App from the AppStore or Google Play. An account was created for the participant for the TiiM-App which was done by the researcher sending the participant an invitational link to the study. It was recommended to the participant to note the password for the account somewhere safe. The researcher logged-in into the admin site of the TiiM-App and set time and date on which the first TiiM-survey should appear on the participants device after consenting on time and date that suits the participant best. This first TiiM-survey marked the

beginning of the 7-day study. After that, the participant received a detailed information sheet on how to handle the E4 and TiiM-App and the link to a video if a repetition of the information was required. Ideally, researcher and participant met again after the 7 days to hand over the E4 wearable and to download all recorded sessions as well as to debrief the participant. During these debriefings, the participant had to fill in the second TAS-20, was asked how the experiment was perceived and thanked for the participation. However, if problems occurred during the study, meetings between researchers and participants were set up to solve occurring problems or to replace defective E4 wearables. After the debriefing was concluded, the researchers gave the participants data a user number for anonymisation and uploaded the data and both TAS-20 into a shared cloud of the researcher team for further processing.

Participants in the control-group received an introduction as well, though E4 and TiiM-App were not mentioned to them and it was limited to information on the study and the TAS-20. Control-group participants were usually asked personally or via messenger programs like Facebook Messenger or WhatsApp whether they were interested in participating in the study. However, it was disclosed from them that they would form the control group or that the goal of the study was to analyse a possible effect of emotional self-reflection. This was done to prevent unwanted encouraging to inform themselves on the matter. When they were interested, consent form and the first TAS-20 was sent to them. The consent forms and questionnaires were either printed and given to the participants to fill out, or they were sent to the participants via messengers or mail for them to print and fill out. Another possibility for participants was to use the PDF marking and signature function and then send back questionnaire and consent form to the researcher. When seven days had passed, the researcher messaged the participants again to fill in the second TAS-20. After receiving the second TAS-20, participants were thanked for their participation and it was offered to meet for a debriefing or provide further explanations over the study and their role as control-group participants.

Data Analysis

As this particular study was a subpart of a larger study it did not use all of the data that was collected. This study focused on a possible reactivity effect that might lead to a reduction in alexithymia scores in a one-week emotional reflection experiment in wearable users. Therefore, the measurements relevant to this study were the pre- and post-experiment TAS-20 scores of all participants and the frequencies of completed TiiM-surveys for participants of the experimental group. In total, the data of 42 participants were used.

The data analysis was done with SPPS. SPSS is a program developed and distributed by the International Business Machines Corporation (IBM) and is designed to handle statistical calculations. The SPSS license used was provided by the University of Twente and for this study, the SPSS version 25 was used. For the demographic variable age mean, standard deviation, minimum and maximal values were calculated. The variables gender and nationality were calculated in total values and percentages.

For first hypothesis, pre-experiment TAS-20 scores are higher than post-experiment TAS-20 scores in the experimental group, a t-test was done. The t-test was done to compare the pre- and post-experiment TAS-20 scores of the 28 participants in the experimental group.

For the second hypothesis, participants who completed more TiiM-surveys have lower post-experiment TAS-20 scores than participants who completed fewer surveys, a

regression analysis was done. With the regression analysis, the effect of the number of completed surveys on the post-experiment TAS-20 score was assessed for the 28 participants in the experimental group. The number of completed TiiM-surveys was obtained by computing the frequency of either self-reported arousal or valence, as this frequency equals the number of questionnaires a participant filled in.

For the third hypothesis, participants in the control-group show no significant difference between pre- and post-experiment TAS-20 scores, a t-test was done. The t-test was done to compare the pre- and post-experiment TAS-20 scores for participants in the control-group.

Results

Descriptive

For the 42 participants the TAS-20 scores pre-experiment and post-experiment were calculated together with minimum and maximum scores, the mean and standard deviation. The mean scores, minimum and maximum scores as well as the standard deviation of the two tests can be found in Table 1. A more extensive overview over the TAS-20 scores per participant can be found in the appendix (see Appendix C).

Table 1
Descriptives of the TAS-20 Questionnaire Pre- and Post-Experiment

	N	Minimum	Maximum	Mean	SD
TAS-20 pre	42	26	65	43.1	8
TAS-20 post	42	26	71	43.3	8.4

Inferential

For the first hypothesis a paired sample t-test was done to test whether pre-experiment TAS-20 scores were significantly higher than post-experiment TAS-20 scores in the experimental group. For this calculation, the data of the 28 experimental-group participants was used. There was no significant difference found between the pre-experiment TAS-20 scores (M=41.86, SD=7.53) and the post-experiment TAS-20 scores (M=41.71, SD=7.69); t(27)=.187, p<.85.

The second hypothesis was tested with a single linear regression analysis to test if participants that completed more surveys had lower post-experiment TAS-20 scores than participants who completed fewer surveys. Again, the data of the 28 experimental-group participants was used. The regression analysis yielded no significant result with (F(1,26)=.52, p<.477), with an R^2 of .02. The post-experiment TAS-20 score decreased by .13 points for each additional survey that was completed (Figure 3).

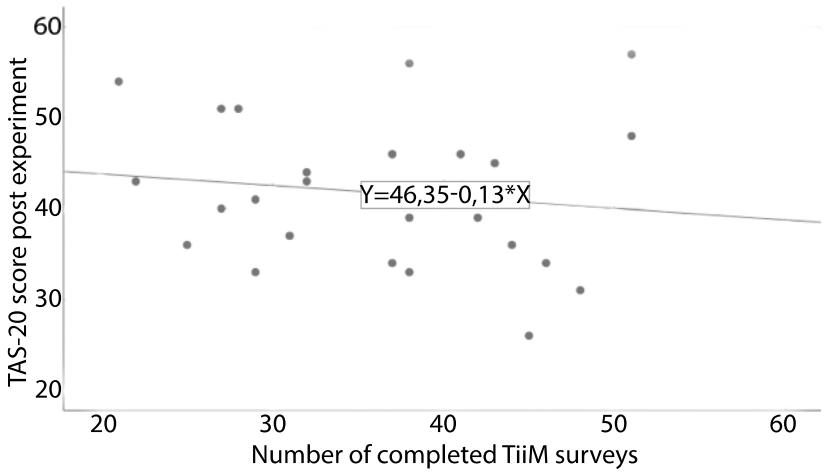

Figure 3. Post-experiment TAS-20 score by number of completed TiiM-surveys.

For the third hypothesis another paired sample t-test was done to test whether the scores on the pre- and post-experiment TAS-20 in the control group are not significantly different from each other. For this calculation, the data of the 14 control group participants was used. There was no significant difference found between the pre-experiment TAS-20 scores (M=45.71, SD= 8.516) and the post experiment TAS-20 scores (M=46.57, SD=9.24); t(13)=-.798, p<.44.

Discussion

Conclusion

The aim of this study was to explore whether reflecting on personal emotions, as represented by core-affect, can lead to a reduction in alexithymia scores in a one-week emotional reflection experiment based on core-affect in wearable users. However, statistically significant support for self-reflection on core-affect having an effect on alexithymia scores was not found. The assumption that such an effect exists was derived from research into reactivity towards being measured and that multiple studies attributed a treatment-like effect of reactivity on their participants (Capellan et al., 2017; Kypri et al., 2007; Morris et al., 2010; Schrimsher & Filtz, 2011). Additional substantiation for the research was derived from the reactivity facilitating nature of experience-sampling as it has the potential of bringing the assessed phenomena to the notice of the measurer (Morris et al., 2010; Kypri et al., 2007; Schrimsher & Filtz, 2011). Though this study did not apply treatment exercises or treatment recommendations of any kind, other studies pointed out that experience sampling gave their participants the feeling of understanding their emotions better (Morris et al., 2010). Reflections on core-affect, were successfully used in prior mobile and wearable studies (Hernandez et al., 2016; Morris et al., 2010). As alexithymia refers to the ability to make reason of personal feelings and bodily sensations and due to its high comorbidity with other psychopathologies, it seemed to be an appropriate construct to assess in terms of emotional-wellbeing. Based on this theorisation, it was hypothesized that the experimental group, that had to measure their core-affect during the one-week experience sampling study, would show a reduction in alexithymia scores due to the reactive nature of measuring. It was also hypothesized, that participants who filled in more surveys, would show a stronger effect than others, as filling in more surveys was understood as better adherence to the study and self-reporting. Next to that, it was hypothesized, that participants of the control group, that only filled in the TAS-20, would show no such effect.

Yet support for these assumptions was not found in this study. The first two hypotheses were not substantiated by the results of this study, though the third was. The t-test for the first hypothesis indicates no significant difference between the pre- and post-experiment TAS-20 score in the experiment group. In fact, the difference between mean pre- and post TAS-20 scores after one week was only 0.15 points, which is too small to assume an effect rather than measurement fluctuations. In comparison, after the 7-week training course in the study of Haase et al. (2015), TAS-20 mean scores dropped in the first subscale by 1 point, for 4.28 points in the second subscale, though increased by 1 point in the third. Additionally, in the study of Beresnevaite (2000), the mean TAS-20 scores decreased from 70.8 (SD=5.5) to 62.8 (SD=9.8) in a 4-month study of group psychotherapy in a sample of patients suffering from coronary heart disease. Yet, this study achieved no significant changes in TAS-20 scores that can be considered comparable to those studies. However, as the TAS-20 scores in the pre-experiment assessment were high in the study of Beresnevaite, a portion of the reduction of TAS-20 scores might be attributable to a regression towards the mean. Nonetheless, the first hypothesis is refuted as there is no statistical substantiation to the assumption that there was an effect on the participants by reflecting on core-affect for a week was found. The regression analysis for the second hypothesis was not significant as well and thereby not in favour of the hypothesis that more completed surveys would lead to a lower post-experiment TAS-20. However, the tendency was in the direction of the hypotheses, as for each completed survey the post-experiment TAS-20 score decreased by

.13 points yet taking a look at the figure 3 makes it clear that this tendency is hard to visualize. Therefore, it cannot be assumed that completing more surveys had a decreasing effect on post-experiment TAS-20 score or would lead to improvements in emotional reflection. For the third hypothesis, there was no significant difference found between pre-experiment and post-experiment TAS-20 scores in the control group. In fact, the mean post-experiment TAS-20 score was with 0.86 points slightly higher than the mean pre-experiment TAS-20 score. Therefore, the hypothesis is supported that there would be no significant difference in TAS-20 scores between the pre- and post-experiment test. However, since there was no significant result in the first two hypotheses either, the third hypothesis does not contribute to the evaluation of a possible reactivity effect other than there was no such reactivity effect.

Theoretical reflections and implication

In the light of these findings, there is no support for a reactivity effect that has treatment- or intervention-like effects on the participants when it comes to emotional self-reflection. However, this is not necessarily a contrasting finding to the concept of reactivity. McCambridge et al. (2014) derived at the conclusion that a single reactivity, or Hawthorne effect, might not exist per se. In their analysis of the Hawthorne effect, they pointed out, that there are too many different effects, mechanisms, potential outcomes and conditions in which reactivity might be present to simply point towards one single Hawthorne reactivity effect that potentially explains it all. So, reactivity might still have been present in the research, yet not in the form that was assessed for. As the concept of reactivity can be elusive and as it is not clearly foreseeable how participants will react to the assessment, it can neither be fully confirmed or fully rejected on the grounds of this study. However, and as the initial Hawthorne studies have shown, there are mechanisms at play, that point towards such an effect. Therefore, it is shared in to the demand of McCambridge et al. (2014), that further research is needed in the field of reactivity.

Yet concerning reactivity in experience sampling, prior studies explicitly pointed out that participants became reactive to the phenomena that were assessed. Kypri (2002, cited by Kypri et al., 2007) pointed out that self-report on alcohol consumption brought problematic drinking behaviour to the attention of the participants. Rowan et al. (2007) pointed out that self-reports on smoking lead to a reactivity in patients, though after the study was concluded. Sutton et al. (2014) found a slight increase in medication adherence due to reactivity and Widdershoven et al. (2019) found that engaging in experience sampling lead to improvements in emotional differentiation. Yet, this study did not obtain significant results to support the involvement of reactivity effetcs. However, a contrast of their studies to this one is, that all of them had durations of several weeks. Even in light of the prolonged duration of these studies, their findings were not always significant, though leaning towards the hypotheses or only significant in certain aspects of the study. This indicates that even after adhering to several weeks of a study, the effects on participants due to assessment and self-report had mainly tendencies towards the hypotheses and no clear and strong support for a reactivity effect. Considering now, that the mean scores on the TAS-20 were lower by .13 points post-experiment and the regression also yielded a, though not significant, decrease in post-experiment scores based on number of completed surveys, the measurement effect on participants might have been more meaningful if the duration of the study had been comparable. So, in theory, the measured might increase if this study would have been prolonged and further substantiation could have been provided to the claim that

self-report measurements affect participants in the measured construct. Nonetheless, verification of this theory requires a repetition of the experiment spanning over several weeks. The conclusion for now and based on the outcomes of this study should be, that an effect of self-reporting on participants was not supported.

This conclusion however gives little to no support towards the claim that self-assessment in itself has the potential to act as an intervention or treatment. Schrimsher and Filtz (2011) find assessment to be a valuable treatment-like method to decrease hazardous drinking that excels in feasibility and might be ideal for individuals that have been placed on a treatment waiting-list to bridge the time. While this proposal has its merits concerning the gathering of self-report data on drinking behaviour during the waiting time that can be used in later treatment, the hopes should not be too high that self-report alone will change much before the real treatment begins. While Schrimsher and Filtz propose that self-reporting would increase self-efficacy and might motivate for changing behaviour it has to be pointed out that this is not necessarily the case. It is inconclusive that everyone following assessment will derive at the conclusion that the behaviour, in this case concerning drinking, must change. After all, it is a self-reflective process that the participant has to undergo to derive at such a conclusion that might be facilitated by seeing personal recordings of drinking behaviour, though it does not have to be a clear causal path as individuals go about their personal reflective process. Based on the findings of this study, the argument is shared that self-reporting may hold beneficial value to accompany treatments. Especially, due to the reduction of recall biases and the reflective processes that may be facilitated or initiated in participants. However, it should be refrained from inflating the benefit of experience sampling as a treatment-like method until further support is gathered.

Lastly, a note on the application of the Toronto Alexithymia Scale 20. While the TAS-20 is a test that is often used in the context of psychopathologies (Altamura et al., 2018; Parolin et al., 2018; Viganò et al., 2018) there are also examples of it being used in the context of non-psychopathologic populations. Examples are professional athletes (Haase et al., 2015) and patients suffering from coronary heart disease, where participants with a history of psychopathology were excluded (Beresnevaite, 2000). It is pointed out that the TAS-20 could be a valuable tool for preventive measures in non-clinical populations (Norman et al., 2018). However, in this study, there were no significant TAS-20 changes observed. At the one hand, this can be because there was no effect on emotional self-reflection and alexithymia, as pointed out above. On the other hand, it cannot be excluded that the TAS-20 might have difficulties in assessing differences in populations that already have a healthy score on alexithymia. In this study, the mean pre-experiment TAS-20 score was 43.1, though the indication of alexithymia begins with 51, which was reached by only 6 participants of 42 in this study (Appendix X). As a comparison, in the study of Beresnevaite (2000), the mean TAS-20 score pre-study was 70.8 indicating the presence of clinical alexithymia. Therefore, it is proposed that as participants were already moderately healthy considering alexithymia in this study and thus, there was not much room for improvement on alexithymia. Next to that, it is also pointed out that the study of Haase et al. (2014) employed a 7-week intensive course in mindfulness training featuring several 3-hour long training days, which dwarfs the self-reflection task on core-affect that had to be completed at least 4 times a day and took approximately 10 seconds. Also, the study of Beresnevaite (2000) featured weekly group therapy sessions over a period of 4 months. Based on that, it is assumed that the TAS-20 either lacks discriminative value in populations that are already healthy or that meaningful change in alexithymia requires more intensive treatments than a one-week self-reflection task.

Strong points and limitations of the study

A clear strong point of the study is its setup. While other studies required their participants to receive extensive trainings, frequent meetings with the researchers or other activities, this study was far less obtrusive as participants simply had to wear the E4 and answer questionnaires on their phone every 2 hours. This way, the study was able to capture participant responses continuously as they carried on with their real-world routines instead of being confined to a laboratory context. Furthermore, researchers pointed out that the effect of assessment on treatment outcomes is not well researched yet (Meier et al., 2017). Thus, this research adds to the pointed-out gap as insights can be derived from this study on how reactivity may impact or may not impact self-report assessments. Another strong point is the number of participants that this study was conducted with, 28 in the experimental group and 14 in the control-group. This number of participants was higher than in other studies as the study of Morris et al. (2010) had 8 participants and Hernandez et al. (2015) had 15. Yet another strong point is that participants were not specifically imprinted on the theme of mindfulness or emotional reflection as the study on a whole sought to investigate a connection between core-affect and physiologic responses. This way, participants were less inclined to explore emotional reflection techniques and applying them which could have implicated the results. Yet another strong point is that the study relied on the TAS-20 as a measurement of alexithymia, which is still a widely applied testing method with adequate psychometric properties. However, it has to be pointed out, that the TAS-20 is a mainly clinical test, though it was also applied successfully with non-clinical participants (Haase et al., 2015).

Yet this study has limitations as well. Preece et al. (2017,2018) acknowledged the adequacy of the psychometric properties of the TAS-20 though, pointed out that the subscale ‘externally oriented thinking’ of the TAS-20 has insufficient internal consistency and should be ignored. Preece et al. (2018) proposed another testing model of alexithymia, the attention-appraisal model of alexithymia, which is said to be supported by statistics as well as being able to answer questions that the TAS-20 could not. While this sounds promising for getting more accurate results, Preece et al. lack the decades of verification and validation of the TAS-20. Additionally, the TAS-20 is a rather short questionnaire which makes it easier to administer, though it has reduced accuracy than more extensive questionnaires. Ultimately, the TAS-20 is no perfect tool, but an adequate one. Another limitation is the duration of the experiment. One week of continuous measurement is fairly long compared to lab studies, yet for investigating the effects of emotional reflection, one week is short and other studies ran for several weeks and months (Haase et al., 2015; Morris et al., 2010). Some participants reported that they were frustrated with the functioning of the TIIM-App on iOS devices, despite being informed about the occurrence of problems beforehand. This might have reduced their willingness to fill in as much surveys as would have been possible. Yet another limitation is, that the control-group was of no comparable size to the experimental group as the data of 6 of the initial 20 control-group participants had to be excluded. Reflection on core-affect could have have been another limitation to this study. While core-affect can be understood as a simpler model of emotions, less biased by folk-psychological (Russel, 2003), it might have been counter-intuitive to participants that are not familiar with the research and rational behind it. Some participants reported that when filling out the surveys, they often stayed near the middle of the axis model, because they had no conception of what the extremes would mean for them. They also reported that within the pre-set timeframe of two

hours and sometimes even within the last two minutes their emotions changed frequently which made it difficult for them to pinpoint their feelings on the axis model. This was especially stressed as being confusing by a participant who had three children and reported a lot of emotional fluctuation during the day.

Suggestions for further research and recommendations and final statement

Based on the findings of this study, claims that reactivity to self-assessment would have treatment-like outcomes were not substantiated. Therefore, it is pointed out, that despite the experience sampling aspect may be a helpful contributor to administering interventions, this effect should not be inflated. Instead, focus and research should be placed on applications and programs that provide accompanying exercises and how they can profit from reactivity effects. Most of the studies that reported a positive effect on alexithymia and emotional reflection featured such exercise programs (Haase et al., 2015; Morris et al., 2010; Norman et al., 2018). Therefore, for intervention design, experience-sampling should be considered to be a helpful tool, but not the sole method to achieve positive intervention outcomes. After all, a reactivity effect on emotional-reflection was not supported and building interventions relying strongly on such an effect cannot be recommended based on this studies outcome. Next to that, building interventions around the possibility that some participants may arrive at certain desirable conclusions or reflections alone, cannot be recommended. It is the purpose of health and psychotherapeutic personnel to help their patients re-think their problematic behaviour or situations and alter it based on that. Simply presenting them with information and hoping that some of it stays with them is no ideal treatment when human health is on the line. This is why it is strongly advocated to better rely on useful exercises and treatments and if reactivity may add something useful, to see it as positive side-effect of the treatment, rather than consider it the treatment in itself. Finally, it is hoped that this research contributed to the study of reactivity effects and how they can be applied in the arising methodology using wearable technology as well as clarifying on the potential magnitude of such reactivity effects in treatments.

References

Association for Contextual Behavioral Science (ACBS), nd. Retrieved from: https://contextualscience.org/TAS_Measure

Altamura, M., Porcelli, P., Fairfield, B., Malerba, S., Carnevale, R., Balzotti, A., Bellomo, A. (2018). Alexithymia Predicts Attrition and Outcome in Weight-Loss Obesity Treatment. *Frontiers in Psychology*, *9*, 2432. https://doi.org/10.3389/fpsyg.2018.02432 (Original work published)

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on psychological science*, *1*(1), 28-58.

Barrett, L., & Barrett, D. J. (2001). An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review*, *19*(2), 175–185. https://doi.org/10.1177/089443930101900204 (Original work published)

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of psychosomatic research*, *38*(1), 23-32.

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *Journal of psychosomatic research*, *38*(1), 33-40.

Bagby, R. M., Tayler, G. J., & Parker, J. D., nd

Beresnevaite, M. (2000). Exploring the Benefits of Group Psychotherapy in Reducing Alexithymia in Coronary Heart Disease Patients: A Preliminary Study. *Psychotherapy and Psychosomatics*, *69*(3), 117–122. https://doi.org/10.1159/000012378 (Original work published)

Berkel, N., Ferreira, D., & Kostakos, V. (2017). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys (CSUR)*, *50*(6), 93. https://doi.org/10.1145/3123988 (Original work published)

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of emotions*, *2*, 173-191.

Capellan, J., Wilde, M. H., & Zhang, F. (2017). Measurement Reactivity in a Randomized Clinical Trial Using Self-Reported Data. *Journal of Nursing Scholarship*, *49*(1), 111–119. https://doi.org/10.1111/jnu.12266 (Original work published)

Cognitive Behaviour Therapy & Science Center, nd. TAS-20, retrieved from: https://cbtscience.files.wordpress.com/2014/02/tas-20.pdf

de Groot, J., & Rodin, G. (1997). Coming Alive: The Psychotherapeutic Treatment of Patients with Eating Disorders. *The Canadian Journal of Psychiatry*, *43*(4), 359–366.

https://doi.org/10.1177/070674379804300403 (Original work published)

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*(4), 463. https://doi.org/10.1037/a0017075 (Original work published)

Feel, nd. Retrieved from: https://www.myfeel.co

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology*, *15*(3), 453–468. https://doi.org/10.1348/135910710x492341 (Original work published)

Haase, L., May, A. C., Falahpour, M., Isakovic, S., Simmons, A. N., Hickman, S. D., … Paulus, M. P. (2015). A pilot study investigating changes in neural processing after mindfulness training in elite athletes. *Frontiers in Behavioral Neuroscience*, *9*, 229. https://doi.org/10.3389/fnbeh.2015.00229 (Original work published)

Hernandez, J., ff, D., Infante, C., Maes, P., Quigley, K., & Picard, R. (2016). Wearable ESM: differences in the experience sampling method across wearable devices, 195–205. https://doi.org/10.1145/2935334.2935340 (Original work published)

International Business Machines Cooperation (IBM), nd. Retrieved from: https://www.ibm.com/nl-nl/?lnk=m

Kashdan, T. B., Barrett, L., & McKnight, P. E. (2015). Unpacking Emotion Differentiation. *Current Directions in Psychological Science*, *24*(1), 10–16. https://doi.org/10.1177/0963721414550708 (Original work published)

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042. https://doi.org/10.1037/a0020962 (Original work published)

Kypri, K., Langley, J. D., Saunders, J. B., & Cashell-Smith, M. L. (2007). Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking. *Addiction*, *102*(1), 62–70. https://doi.org/10.1111/j.1360-0443.2006.01632.x (Original work published)

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015 (Original work published)

Meier, E., Miller, M., Lombardi, N., & Leffingwell, T. (2017). Assessment reactivity: A randomized controlled trial of alcohol-specific measures on alcohol-related behaviors. *Addictive Behaviors*, *67*, 44–48. https://doi.org/10.1016/j.addbeh.2016.11.025 (Original work published)

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., & Deleeuw, W. (2010). Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness. *Journal of Medical Internet Research*, *12*(2), e10. https://doi.org/10.2196/jmir.1371 (Original work published)

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513 (Original work published)

Nelson, E. C., Verhagen, T., & Noordzij, M. L. (2016). Health empowerment through activity trackers: An empirical smart wristband study. *Computers in Human Behavior*, *62*, 364–374. https://doi.org/10.1016/j.chb.2016.03.065 (Original work published)

Norman, H., Marzano, L., Coulson, M., & Oskis, A. (2018). Effects of mindfulness-based interventions on alexithymia: a systematic review. *Evidence-Based Mental Health*, ebmental-2018-300029. https://doi.org/10.1136/ebmental-2018-300029 (Original work published)

Olson, R., Hogan, L., & Santos, L. (2005). Illuminating the History of Psychology: Tips for Teaching Students about the Hawthorne Studies. *Psychology Learning & Teaching*, *5*(2), 110–118. https://doi.org/10.2304/plat.2005.5.2.110 (Original work published)

Parolin, M., Miscioscia, M., Carli, P., Cristofalo, P., Gatta, M., & Simonelli, A. (2018). Alexithymia in Young Adults With Substance Use Disorders: Critical Issues About Specificity and Treatment Predictivity. *Frontiers in Psychology*, *9*, 645. https://doi.org/10.3389/fpsyg.2018.00645 (Original work published)

Preece, D., Becerra, R., Allan, A., Robinson, K., & Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences*, *119*, 341-352.

Preece, D., Becerra, R., Robinson, K., & Dandy, J. (2017). Assessing alexithymia: Psychometric properties and factorial invariance of the 20-Item Toronto alexithymia scale in nonclinical and psychiatric samples. *Journal of Psychopathology and Behavioral Assessment*, 1-12.

Rowan, P. J., Cofta-Woerpel, L., Mazas, C. A., Vidrine, J., Reitzel, L. R., Cinciripini, P. M., & Wetter, D. W. (2007). Evaluating reactivity to ecological momentary assessment during smoking cessation. *Experimental and Clinical Psychopharmacology*, *15*(4), 382. https://doi.org/10.1037/1064-1297.15.4.382 (Original work published)

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161. https://doi.org/10.1037/h0077714 (Original work published)

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145. https://doi.org/10.1037/0033-295x.110.1.145 (Original work published)

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion*, *23*(7), 1259–1283. https://doi.org/10.1080/02699930902809375 (Original work

published)

Santarnecchi, E., D'Arista, S., Egiziano, E., Gardi, C., Petrosino, R., Vatti, G., … Rossi, A. (2014). Interaction between Neuroanatomical and Psychological Changes after Mindfulness-Based Training. *PLoS ONE*, *9*(10), e108359. https://doi.org/10.1371/journal.pone.0108359 (Original work published)

Schrimsher, G. W., & Filtz, K. (2011). Assessment Reactivity: Can Assessment of Alcohol Use During Research be an Active Treatment? *Alcoholism Treatment Quarterly*, *29*(2), 108–115. https://doi.org/10.1080/07347324.2011.557983 (Original work published)

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415 (Original work published)

Spire, nd. Retrieved from: https://spire.io/pages/stone

Sutton, S., Kinmonth, A.-L., Hardeman, W., Hughes, D., Boase, S., Prevost, T. A., … Farmer, A. (2014). Does Electronic Monitoring Influence Adherence to Medication? Randomized Controlled Trial of Measurement Reactivity. *Annals of Behavioral Medicine*, *48*(3), 293–299. https://doi.org/10.1007/s12160-014-9595-x (Original work published)

Statista, nd. ‚Statistics & Facts on Wearable Technology', retrieved from: https://www.statista.com/topics/1556/wearable-technology/

Taylor, G. J. (2000). Recent Developments in Alexithymia Theory and Research. *The Canadian Journal of Psychiatry*, *45*(2), 134–142. https://doi.org/10.1177/070674370004500203 (Original work published)

Viganò, C. A., Beltrami, M. M., Bosi, M. F., Zanello, R., Valtorta, M., & Maconi, G. (2018). Alexithymia and Psychopathology in Patients Suffering From Inflammatory Bowel Disease: Arising Differences and Correlations to Tailoring Therapeutic Strategies. *Frontiers in Psychiatry*, *9*, 324. https://doi.org/10.3389/fpsyt.2018.00324 (Original work published)

Widdershoven, R., Wichers, M., Kuppens, P., Hartmann, J. A., Menne-Lothmann, C., Simons, C., & Bastiaansen, J. A. (2019). Effect of self-monitoring through experience sampling on emotion differentiation in depression. *Journal of Affective Disorders*, *244*. https://doi.org/10.1016/j.jad.2018.10.092 (Original work published)

Appendices

Appendix A

Informed consent form

26.10.2018 University of Twente, The Netherlands

Informed Consent

Informed Consent for the research **"Does your smartwatch know how you feel?"**

I hereby declare that I have been informed in a manner which is clear to me about the nature and method of the research as described in the introduction of the research. My questions have been answered to my satisfaction. I agree by my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my given data is to be used in scientific publications or made public in any other manner, then the data will be made completely anonymous. My personal data will not be disclosed to third parties without my consent. If I request further information or help for the research, now or in the future, I may contact the researcher:

Lukas Libbertz: l.b.libbertz@student.utwente.nl tel.: +4917622515097

If you have any complaints about this research, please direct them to the researchers or the secretary of the Ethics Committee of the Faculty of Behavioural Sciences at the University of Twente, Drs. L. Kamphuis-Blikman P.O. Box 217, 7500 AE Enschede (NL), telephone: +31 (0)53 489 3399; email: l.j.m.blikman@utwente.nl).

By signing this informed consent, I agree to participate in this study as well as give the permission to the researchers to use the gained data. I have been informed, that I can withdraw from the experiment at any time.

Date: _____ **Signature**: _____

Appendix B

Toronto Alexithymia Scale 20

Please answer the following questions, using the scale provided:

(1) Completely disagree
(2) Disagree
(3) Neutral
(4) Agree
(5) Completely agree

1.	I am often confused about what emotion I am feeling.	1 – 2 – 3 – 4 – 5
2.	It is difficult for me to find the right words for my feelings.	1 – 2 – 3 – 4 – 5
3.	I have physical sensations that even doctors don't understand.	1 – 2 – 3 – 4 – 5
4.	I am able to describe my feelings easily.	1 – 2 – 3 – 4 – 5
5.	I prefer to analyze problems rather than just describe them.	1 – 2 – 3 – 4 – 5
6.	When I am upset, I don't know if I am sad, frightened, or angry.	1 – 2 – 3 – 4 – 5
7.	I am often puzzled by sensations in my body.	1 – 2 – 3 – 4 – 5
8.	I prefer to just let things happen rather than to understand why they turned out that way.	1 – 2 – 3 – 4 – 5
9.	I have feelings that I can't quite identify.	1 – 2 – 3 – 4 – 5
10.	Being in touch with emotions is essential.	1 – 2 – 3 – 4 – 5
11.	I find it hard to describe how I feel about people.	1 – 2 – 3 – 4 – 5
12.	People tell me to describe my feelings more.	1 – 2 – 3 – 4 – 5
13.	I don't know what's going on inside me.	1 – 2 – 3 – 4 – 5
14.	I often don't know why I am angry.	1 – 2 – 3 – 4 – 5
15.	I prefer talking to people about their daily activities rather than their feelings.	1 – 2 – 3 – 4 – 5
16.	I prefer to watch "light" entertainment shows rather than psychological dramas.	1 – 2 – 3 – 4 – 5
17.	It is difficult for me to reveal my innermost feelings, even to close friends.	1 – 2 – 3 – 4 – 5
18.	I can feel close to someone, even in moments of silence.	1 – 2 – 3 – 4 – 5
19.	I find examination of my feelings useful in solving personal problems.	1 – 2 – 3 – 4 – 5
20.	Looking for hidden meanings in movies or plays distracts from their enjoyment.	1 – 2 – 3 – 4 – 5

TAS

Appendix C

TAS-20 Scores Pre- and Post-Experiment per Participant ID

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
TAS-pre	33	48	43	48	36	47	61	39	38	34	49	37	37	48	50	42	39	46	36	47	53	44	48	26	35	36	38	34	65	53	49	53	44	48	40	35	34	43	49	48	45	34
TAS-post	33	48	46	51	33	45	57	39	31	36	39	34	36	44	56	42	40	43	37	46	54	51	43	26	34	43	41	41	71	49	44	59	43	48	43	43	36	41	51	47	41	36

Note. Participant ID's from 1 to 17 were participants from the first iteration of the study, ID's from 18 to 28 were from this iteration of the study and participant ID's from 29 to 42 were participants from the control group.