University of Twente
Faculty of Behaviour, Management and Society
Department of Positive Psychology and Technology

How to measure how you feel?

Assessing test-retest reliability of the Affect Grid in the Experience Sampling Method and exploring convergent validity with the Positive and Negative Affect Scale

Bachelor Thesis

First Supervisor: Matthijs L. Noordzij

Second Supervisor: Jannis T. Kraiss

By Laura Sophie Stevens

s1878654

Table of Content

Abstract 2

1. Introduction 3
   - 1.1 The Experience Sampling Method 3
   - 1.2 Reliability and validity of ESM questions 3
   - 1.3 Core affect and the dynamics of affect 4
   - 1.4 Measuring core affect 6
   - 1.5 Current study 7

2. Methods 8
   - 2.1 Design 8
   - 2.2 Participants 8
   - 2.3 Materials and measures 9
   - 2.4 Procedure 12
   - 2.5 Data analysis 12

3. Results 14
   - 3.1 Descriptive statistics 14
   - 3.2 Reliability assessment 16
   - 3.3 Validity assessment 17
   - 3.4 Additional analysis with a lower compliance rate 19

4. Discussion 20
   - 4.1 Theoretical reflection 20
   - 4.2 Practical and scientific implications 22
   - 4.3 Limitations and future research 22
   - 4.4 Conclusion 23

References 24

Appendices 28

Abstract

The Experience Sampling Method (ESM) is receiving increasing attention in the field of health and social research and has already been implemented by numerous researchers. This method turned out to be a powerful addition to more conventional survey studies, because it allows the researcher to study psychological constructs in flux, like the experience of feelings or psychiatric symptoms, in the context of everyday life. Surprisingly, little attention has been paid to the psychometric properties of ESM questions. Therefore, the current study aims to contribute to this growing interest in the ESM by assessing test-retest reliability and convergent validity of ESM questions in the framework of measuring core affect. Core affect is made up of the dimensions valence and arousal and is thus far employed within numerous studies for capturing the dynamic nature of affective experiences. Thus, a repeated measures design study in the framework of the ESM with signal-contingent reporting was conducted, in which 19 people were asked to respond to the Affect Grid, six times a day, i.e. every two hours between 10 AM and 8 PM, for one week via a smartphone application. Besides the Affect Grid, which served as the ESM question, participants had to complete the Positive and Negative Affect Scale (PANAS), which was assumed to be a converging survey. Grounded in the assumption of an between-subject stability in the mean of core affect dimensions (i.e. arousal and valence), test-retest reliability was assessed by correlating the split halves of answers of the Affect Grid. Convergent validity was determined by evaluating the associations between averaged state scores produced by the Affect Grid and the trait scores detected by the PANAS. Findings of a bivariate correlation analysis (BCA) supported excellent estimates of test-retest reliability. Scores of a linear mixed model analysis revealed a preliminary trend for agreement between both measures, but observed correlation coefficients of a BCA showed only weak correlations with the convergent survey (PANAS). Overall, the findings suggest that the Affect Grid is a reliable and valid tool for assessing core affect in terms of the affective home base. The study was successful as it was the first attempt for investigating the psychometric properties of ESM questions. Our findings underline the importance for further research with respect to the reliability and validity of ESM questions for filling the current lack of knowledge.

1. Introduction

1.1 The Experience Sampling Method

The Experience Sampling Method (ESM) has turned out to be a useful tool in psychological research studies, because it allows the researcher to study the dynamic nature of human psychological experience as it naturally occurs in the normal environment of the participant. Hence, the researcher can describe patterns of an individual's daily experience and come up with accurate measurements about cognitive and emotional subjective states (Csikszentmihalyi, 2014). Surveys within the ESM set up provide a coarse insight into the fluctuations of an individual, while traditional survey design studies have the ambition to capture more stable aspects of an individual in a detailed manner. This turned out to be especially interesting for gathering data about the quality of life and the experience of feelings, for instance by linking the variability of feelings in the context of daily routine to psychiatric symptoms (Myin-Germeys, Kasanova, Vaessen, Vachon, Kirtley, Viechtbauer, & Reininghaus, 2018). Although there is a converging interest in studying peoples subjective emotions and feelings in everyday life, very little is known yet about the methodological quality of ESM questions (Csikszentmihalyi, 2014).

1.2 Reliability and validity of ESM questions

The assessment of psychometric properties of measures is an essential part of health and social research, because it is the pre-requisite of quantifying human behaviour (Dooley, 2001). Concerning the quality of a survey it has to be questioned if it measures the concept which it is intended to measure (i.e. how valid is the tool) as well as if the responses are constant and consistent over time (i.e. how reliable is the tool). Whereas classical surveys are quite long and not (often) administered on only one occasion, ESM questions tend to be very short, but repeated very often. Because of this, it is questioned, if the short ESM surveys can reflect constructs of interest in a valid and reliable way.

To determine validity of a scientific survey three types of validity are often considered: criterion-, construct- and content validity. To establish convergent validity aggregated results from ESM questions can be compared to a convergent measurement tool. One study by Versluis, Verkuil, Lane, Hagemann, Thayer, and Brosschot, (2018) examined the relation between trait variation and state variation in the context of Emotional Awareness (EA). Based on the latent state-trait theory, they wanted to investigate whether the ESM questions for EA as

indicator for state variation show significant correlations with a 10-item survey about trait EA (Steyer, Schmitt, & Eid, 1999). In the same vein, Littman, Satia, White, Bowen, and Kristal (2006), explored the convergent validity of a two-item survey about stress in the ESM set-up by means of a 4-item survey, a 10-item survey and a 53-item survey. Thus, Littman et al. (2006) wanted to explore whether the 2-item ESM survey can measure the complex construct of stress likewise to the more complex and longer surveys. Using this approach, researchers have been able to estimate whether ESM questions can capture constructs of interest which they are intended to measure.

Reliability refers to the extent to which a measure produces consistent results across situations. This aspect appears to be contradicting with the essence of ESM of studying natural fluctuations in psychological experiences. Hence, perfect reliability in that sense might defeat the purpose of the ESM. But recent research observed that individuals displayed a stable pattern in their fluctuations over time (Versluis et al., 2018; Kuppens, Oravecz, & Tuerlinckx, 2010). Kuppens, Oravecz, and Tuerlinckx (2010) collected data ten times a day for two weeks, whereas Versluis et al. (2018) faced their participants six times per day with ESM questions for only two days. Although, both studies can be differentiated by the complexity of ESM data, they came up with similar results about the stability in average scores. Therefore, several lines of evidence suggest that reliability assessment of ESM questions can be approached in terms of test-retest reliability, by which correlations of scores from repeated measurements can estimate the degree of reliability (e.g. Dooley, 2001; Versluis et al., 2018; Csikszentmihalyi, 2014). A starting point for the current study displays the method employed by Csikszentmihalyi (2014), who examined the individual stability of responses by correlating mean scores and standard deviations of the first half with the mean scores and standard deviations of the second half of the week.

1.3 Core affect and the dynamics of affect

Increasing numbers of research studies currently point to the construct of core affect, instead of referring to concepts like “feelings” or “emotions”, (e.g. Yik, Russell, & Steiger, 2011; Kuppens, Oravecz, & Tuerlinckx, 2011). Core affect encompasses two dimensions, namely valence (pleasure-displeasure) and arousal (arousal-sleepiness) which can be seen in Figure 1 (e.g. Russel & Barrett, 1999; Russel, Weiß and Mendelsohn, 1989). According to a definition provided by Russel (2003, p.148), core affect is “that neurophysiological state consciously accessible as the simplest raw (non-reflective) feelings evident in moods and emotions”. Additionally, Russel (2003, p.148) states that core affect can be defined as “primitive, universal

and simple". These definitions highlight the aspect that everyone is always able to verbalize how one feels in terms of core affect, i.e. how active and how pleased one is (Yik, Russell, & Steiger, 2011).

In addition, it is important to define the DynAffect account, which has its roots in the dynamical systems theory (e.g. Vallacher, & Nowak, 1994; Nowak, & Lewenstein, 1994; Camras, & Witherington, 2005). The DynAffect account encompasses three influential factors for core affect, namely (1) the affective home base, (2) the affective variability and (3) the attractor strength, which are visualized in Figure 2. With respect to the present study, the affective home base will be of great importance. This term refers to the baseline functioning of individuals (Kuppens, Oravecz, & Tuerlinckx, 2010; Russel, 2003). Aforementioned, previous research by Versluis et al. (2018) and Kuppens, Oravecz, & Tuerlinckx (2010) discovered that even though people tend to fluctuate in their affective experiences across situations, they display highly stable affective home bases over time.

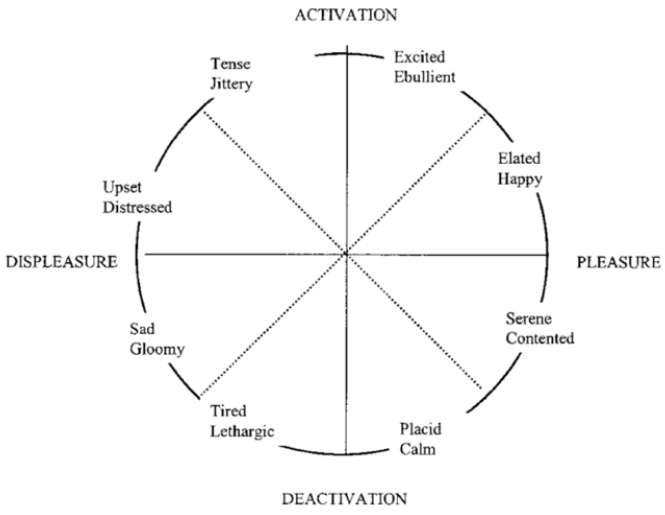

Figure 1. Visualization of the dimensions of core affect by Russel (2003). Valence displays the horizontal dimension, which ranges from displeasure (e.g. misery) to a neutral point to pleasure (e.g. euphoria). Arousal is represented as the vertical axis which ranges from deactivation (e.g sleepiness) to activation (e.g. excitement).

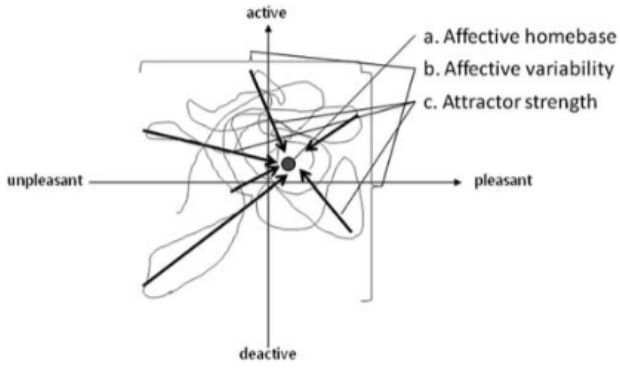

Figure 2. Graphical representation of the DynAffect account of core affect that drives affective change, consisting of (a.) the affective home base, (b.) affective variability and (c.) attractor strength. This model was created by Kuppens, Oravecz, and Tuerlinckx (2010).

1.4 Measuring core affect

Having defined what is meant by core affect, it is now necessary to discuss what measures are already established for measuring this construct. Based on previous observations, it can be outlined that arousal and valence, can be assumed as continuous dimensions, because they entail bipolar opposites and can be characterized as being orthogonal to each other (e.g. Yik, Russell, & Steiger, 2011; Kuppens, Oravecz, & Tuerlinckx, 2010). Common measurement tools for capturing core affect are the Affect Grid and the Positive and Negative Affect Scale (PANAS) (e.g. Crawford & Henry, 2004; Watson & Clark, 1994; Watson, Clark and Tellegen, 1988).

The Affect Grid is a single-item scale, which can be considered as a simple and undisguised measure of core affect (Russell & Gobet, 2010; Russel, Weiß, & Mendelsohn, 1989). The grid is a build-up of two axes, where the X-axis is representative for the continuum of valence (displeasure -pleasure) and the Y-axis displays the continuum of arousal (sleepiness-activation). By setting a mark within the Affect Grid, two answers are generated simultaneously which sum up to the core affect.

Furthermore, the PANAS is a prominent self-measurement tool for measuring the core affect utilizing 20 items that are concerned with the positive and negative affect of an individual (Watson, Clark, & Tellegen, 1988). Negative Affect (NA) measures the extent of experiencing subjective stress and unpleasable feelings and Positive Affect (PA) investigates the extent to which one feels enthusiast and alert (Crawford & Henry, 2004; Watson, Clark, & Tellegen, 1988). Moreover, PA can be placed by high arousal and pleasant feelings at the one end to sleepiness and unpleasant feelings at the other end. NA connects to the range between high arousal and sleepiness on the one end to low levels of arousal and pleasant feelings at the other

end (Feldmann, 1995; Watson, Wiese, Vaidya and Tellegen, 1999). Next, the PANAS is presumed to be sensitive not only for measuring fluctuations in daily core affect, but also by predicting long-term, trait-like stability of core affect.

1.5 Current study

Although an uprising interest in the ESM is noticeable, far too little attention has been paid to the assessment of the psychometric properties of ESM questions. Based on the fact that ESM surveys are quite short and repeated over time, it is questionable if they can capture constructs of interest like traditional surveys. That is why this paper addresses the indicated lack of knowledge by making the first attempt to assess the reliability and validity of core affect measures in the ESM set up. The current study will focus on state and trait measurements for evaluating the quality of the ESM regarding the measurement of core affect. The affect grid was found to be a sensitive tool for measuring core affect on a state-level in a very short amount of time, which is why it is appropriate for its implementation as an ESM question. Furthermore, the PANAS will be implemented for measuring the individual affective home base of each participant on a trait level.

Therefore, the goal of the current study is two-fold: (1) to investigate the reliability in terms of the stability of responses (i.e. over time) of the Affect Grid in the context of the ESM set-up, and (2) to assess the convergent validity of the Affect Grid by means of the PANAS. With regard to the stability of responses, it is hypothesized that average responses significantly correlate. Several studies already observed that people display highly stable patterns in longitudinal data with regard to affective experiences and found significant correlations between split halves of scores (Versluis et al. 2018; Kuppens, Oravecz, & Tuerlinckx, 2010). Next to that, for assessing the content validity of the Affect Grid, it is predicted that the trait measurements of the PANAS about the individual affective home base significantly correlate with the average responses of the state measures to arousal and valence. This hypothesis is grounded in previous research findings, which observed significant correlates between short ESM surveys and convergent measures (Versluis et al., 2018; Littman et al., 2006). More specifically, based on the definition provided by Feldmann (1995) it is hypothesized that there is a positive and significant association between PA and arousal, NA and arousal, as well as between PA and valence. Finally, a significant and negative relationship is suggested between NA and valence.

2. Methods

2.1 Design

The current study can be defined as a correlational study with a repeated measures design within the ESM set up. The data gathering itself was carried out through a smartphone application called "The Incredible Intervention Machine" (TIIM). Next to the Affect Grid, which served as the ESM measurement tool, conventional surveys were employed, namely the Positive and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988) the Neuroticism-Scale (Francis, Lewis, & Ziebertz, 2006) and the Toronto Alexithymia Scale-20 (TAS-20) (Bagby, Parker, & Taylor, 1994) (only the PANAS is taken into consideration within the current study; the other measures were beyond the scope of the current study).

Signal-contingent, fixed-times reporting was carried out, which means that participants received a notification in case new questions are available to answer (Berkel, Ferreira, & Kostakos, 2017). In addition, the question of responding to the Affect Grid was scheduled on regular repeating times, namely six times a day between 10AM and 8PM every two hours for seven consecutive days. Therefore, participants were asked to respond 42 times to the Affect Grid. In contrast to the Affect Grid, the survey data was gathered once at the beginning. By the implementation of both the PANAS and the Affect Grid, a combination of trait sampling, as well as dynamic state sampling was established. While the PANAS is assumed to investigate the affective home base on a trait level, the Affect Grid investigates the constant flux in core affect, therefore, the state scores have to be aggregated to come up with the affective home base in the end.

2.2 Participants

Participants were gathered through convenience sampling, as well as personal contact. Therefore, people were recruited by asking if they are willing to invest time and effort in engaging in the current study. In total, N=52 respondents volunteered their time for the study, where 34 (65.4%) were females and 18 (34.6%) were males. The age of the participants ranged between 18 and 52, with an average of 22.6 (SD=5.8). Also, the majority of the participants were of German nationality, while only two had other nationalities.

The threshold for the compliance rate was set by 65% (responding to at least 28 Affect Grids) (Chen, Cordier, & Brown, 2015; Palmier-Claus et al., 2011; Berkel, Ferreira, & Kostakos, 2017). By taking the established threshold in consideration, 26 participants had to be excluded from the analysis. Besides, 7 participants were not able to proceed with the study due

to technical issues with the smartphone application. Therefore, there was a sum of 33 (63.5%) excluded participants in the end, which leaves a sample of N=19.

2.3 Materials and measures

For the current study, only two of the implemented measurement tools are of great interest for the state sampling and trait sampling of core affect, namely the PANAS and the Affect Grid, which were both implemented in the TIIM smartphone application.

The Incredible Intervention Machine (TIIM)

TIIM is a smartphone application, which is available for IOS and Android and was designed by professionals of the BMS Lab, which is part of the University of Twente. This smartphone application enables researchers to conduct a study within the context of the ESM. A set of questions is sent to participants on predefined dates and times. Whenever new questions are available, participants are notified by the application with a push notification, so that they are reminded to answer the questions in a specific time frame until the question expires. Researchers can customize the application in terms of their research purpose, for instance setting time slots or deciding about the type of answer (e.g. open answer or multiple choice).

To log into the application, the participants had to register via a starting page in an internet browser through a link received by the researcher. The link, which could be opened in any internet browser on any device, referred the people to a so-called starting page, that displayed the informed consent (Appendix B), demographical questions about name, gender and age, and a registration page, on which they had to sign in with their e-mail address. After they have signed in, the researcher had to assign each respondent to the study as a participant, which automatically activated the sending of an e-mail which contained all the relevant information about the procedure of the study, theoretical background of the Affect Grid, instructions about downloading the application and contact details of the researcher in case of questions (see also Appendix C). For the full participation, they had to download the smartphone application from the app store. By assigning each participant to a specific intervention, the questions became available on the custom time slots, and the push notifications were activated.

The Positive and Negative Affect Scale

The PANAS was previously introduced as a measurement tool, which enables not only the measuring of state-level core affect but also for trait-level core affect. Depending on the

research purpose of the researcher the instructions can be adjusted concerning the items, e.g. “To what extent have you felt this way today?” (today) or “To what extent do you feel this way right now?” (moment). For this study, the survey was employed to measure the individual affective home base, thus, long-term instructions were adapted, e.g. “To what extent do you generally feel distressed?” (see also Appendix A). For the PANAS items, a Likert scale was used, which ranged from (1) “not at all” to (5) “very much”. The interface of the application for the PANAS items is represented in Figure 3.

Research by Crawford and Henry (2004) revealed strong internal consistency for the PA-scale (r = .88 - .09) and the NA-scale (r = .84- .87), thus, it can be assumed that the PANAS possesses adequate reliability. Next, to that, they proved that the PANAS measures two distinct factors that are at least moderately negatively correlated. Although existing literature presents PA and NA, not as completely orthogonal constructs, results of previous studies show that they are nearly perfectly orthogonal to each other due to a VARIMAX rotation of factor analyses which increases the overall discriminant validity. Also, Watson (1988) revealed that the PANAS displays high convergent validity with other measures of valanced affects.

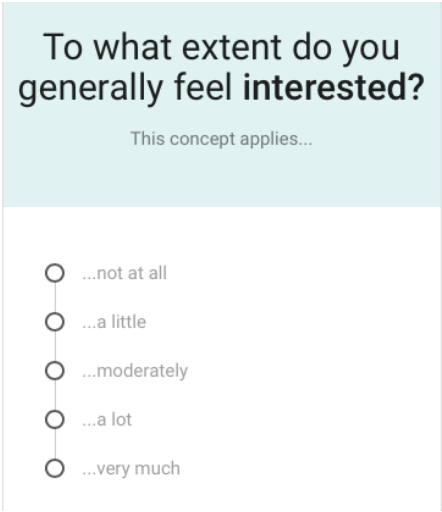

Figure 3. The interface of the TIIM application in the form with which the PANAS was administered to the participants.

The Affect Grid

Likewise to Russel, Weiß and Mendelsohn (1989), a 10 x 10 two-dimensional grid was employed in the application to visualize the axes. Different to Kuppens, Oravecz and Tuerlinckx (2010) and Russel, Weiß and Mendelsohn (1989), the affective labels (Stress,

Depression, Relaxation, Excitement) within the grid were avoided to reduce potential biases based on the assumption of lexical priming (Hoey, 2012). For instance, the label of "depression" appeared to be leading, because it evokes the connection to a mental disease. It is assumed that some participants cannot or do not want to relate their feelings to something labelled as depressive. From a technical point of view, the smartphone application itself entailed a fixed space within the interface which enables scoring. Therefore, the Affect Grid had to completely fit within that given space to make scoring in the complete visualized range of - 5 to 5 possible, which is why the labels of the axes were also integrated within the grid instead of being placed outside the grid.

The Affect Grid was presented with the statement ahead "Please indicate how you feel". Using the touchscreen of the smartphone each participant was able to set a dot within the grid. As it is represented in Figure 4 the yellow dot was a visualization for the participants to set their mark. The participants were able to move the dot within the scoring field during their decision making. The middle point of the visualized dot was the mark for the resulting score. Similar to Kuppens, Oravecz, and Tuerlinckx (2010), the scoring was visualized as a 10 x 10 grid, but for coming up with more detailed data, the axes ranged between -100 and 100 by translating the set dot into numerical values.

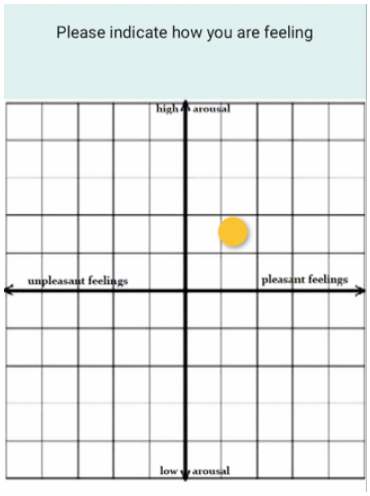

Figure 4. Affect Grid implemented in the current study within the smartphone application TIIM. The respondent is asked to indicate how one is currently feeling by placing the yellow dot to a suitable place in the grid. The X-axis is representative of feelings ranging from unpleasant to pleasant and the Y-axis is displaying the range from low arousal to high arousal.

2.4 Procedure

The ethics committee of the University of Twente approved the study in March 2019. After the official approval, the data collection was started and endured throughout April. The smartphone application was checked by employing a pilot study, which was filled in by the researcher beforehand, to assess the working of the time slots as well as the overall interface and usability for the user concerning the Affect Grid (Mehl, & Conner, 2012).

In the beginning, the participants were asked to fill out the three questionnaires, which were already mentioned, namely the PANAS, the Neuroticism Scale and the TAS-20. Thus, 50 items had to be answered in total, which took the participants around 15 minutes. The three conventional surveys represented the first part of the data collection, by which only the data of the PANAS had to be considered for the analysis of the current paper. Next to that, the second part of the data collection consisted of the Affect Grid, which was displayed for the whole period of seven days, six times a day within the time frame of 10 AM until 8 PM every two hours (10 AM, 12 AM, 2 PM, 4 PM, 6 PM, 8 PM). Thus, the participants were asked to respond on 42 occasions. In case a new Affect Grid for scoring is available, the participant was notified via a push notification on the smartphone. After the notification, the participant had 90 minutes to respond until the Affect Grid expired. The scoring of the Affect Grid took no more than one minute for each participant. In sum, the participants were asked to invest approximately one hour for the current study.

2.5 Data analysis

The gathered data was exported as a CSV file from the TIIM dashboard. After making some adjustments to the file, e.g. splitting the answers of the Affect Grid into two distinct answers to separate the subscale of arousal and valence, the file was imported into SPSS, version 24 (Statistical Program for Social Sciences). Beforehand, the data of the PANAS as well as the data of the Affect Grids were merged into one file, which allows the calculation of correlations.

Descriptive statistics. To get an overall impression of the group about their affective home bases and their tendency to fluctuate in core affect, descriptive statistics in form of means and standard deviations were calculated. Next, boxplots were created to illustrate the fluctuations for each participant concerning valence and arousal.

Bivariate correlation analysis. To assess the stability of responses, a Pearson correlation was calculated. An alpha level of .05 was selected to decide about the significance of the results. Based on the fact that the dataset missed values from the not completed Affect Grids, the mean responses to the Affect Grids were split into two different forms. By splitting the answers not only based on the individual middle of each participant but by also considering the splitting based on even and odd numbers of the Affect Grid, two correlation coefficients for each construct were computed to enrich the conclusion about the overall consistency between answers. With regard to the analysis, a Pearson coefficient of > .9 (> -.9) is assumed as very strong, a result which falls in the range of .5 (-.5) to .9 (-.9) is considered as strong, the range between .3 (-.3) and .5 (-.5) displays moderate correlations and results which are < .3 (< -.3) are weak (Cohen, 1992).

Linear mixed model analysis. To investigate the validity of the Affect Grid, a linear mixed model analysis was conducted in terms of a repeated covariance type of an autoregressive model. Therefore, the ESM scores for both arousal and valence set as dependent variables and the scores of the PANAS are fixed as covariates. The goal of this analysis is to examine the association between the trait scores (PANAS) with the longitudinal ESM scores (Affect Grid). Furthermore, marginal means for persons (i.e. average for a person across all time points) and for the Affect Grids (i.e. average across all persons at a time point) were estimated by selecting the ESM Scores of valence and arousal as respective dependent variables by also setting participants and time point as fixed factors.

3. Results

3.1 Descriptive statistics

Affect Grid. In sum 740 Affect Grids were scored concerning the final sample of N = 19, which adds up to an average response rate of 92,73%. This response rate can be rated as higher than the set compliance rate of 65% as well as the total response rate of the sample of participants who answered at least one Affect Grid (N = 45) which is 62,5% with 1.182 answered Affect Grids in sum. Overall, the mean score of the final sample tended to be higher for valence (M = 33.31) than for arousal (M = 14.39). Nevertheless, the scores indicate a rather neutral affective home base with the tendency to experience "excitement". This trend can also be seen by taking a closer look at the individual level (see also Figure 5 & 6), in which the highest mean was found to be 91.77 for valence, while the mean of arousal only peaked up to 78. Besides, the tendency to have lower scores in arousal was also seen in the lowest mean score by having -5.13 for valence and -10.56 for arousal. Therefore, the stated ranges of mean scores, indicates that participants tend to differ concerning their affective home bases.

Turning to overall variation in terms of standard deviations, valence (SD = 40.31) and arousal (SD = 42.7) were found to reflect similar values. This tendency of resemblance can be also detected on the individual level by comparing the range in variation of valence (8.98 – 52.53) with arousal (10.36 – 56.65) (see also Appendix D)

PANAS. In sum, participants scored weak for PA (M = 2.34, SD = 0.6) and moderately for NA (M = 3.26, SD = 0.72). By taking a more closer look at the individual level, it is striking, that individual differences can be detected. The range of means for PA (1.2 – 3.3) and NA (2.0 – 5.0) shows that more variation can be found between individuals for NA than to PA. The outcomes for every individual can be found in Appendix E.

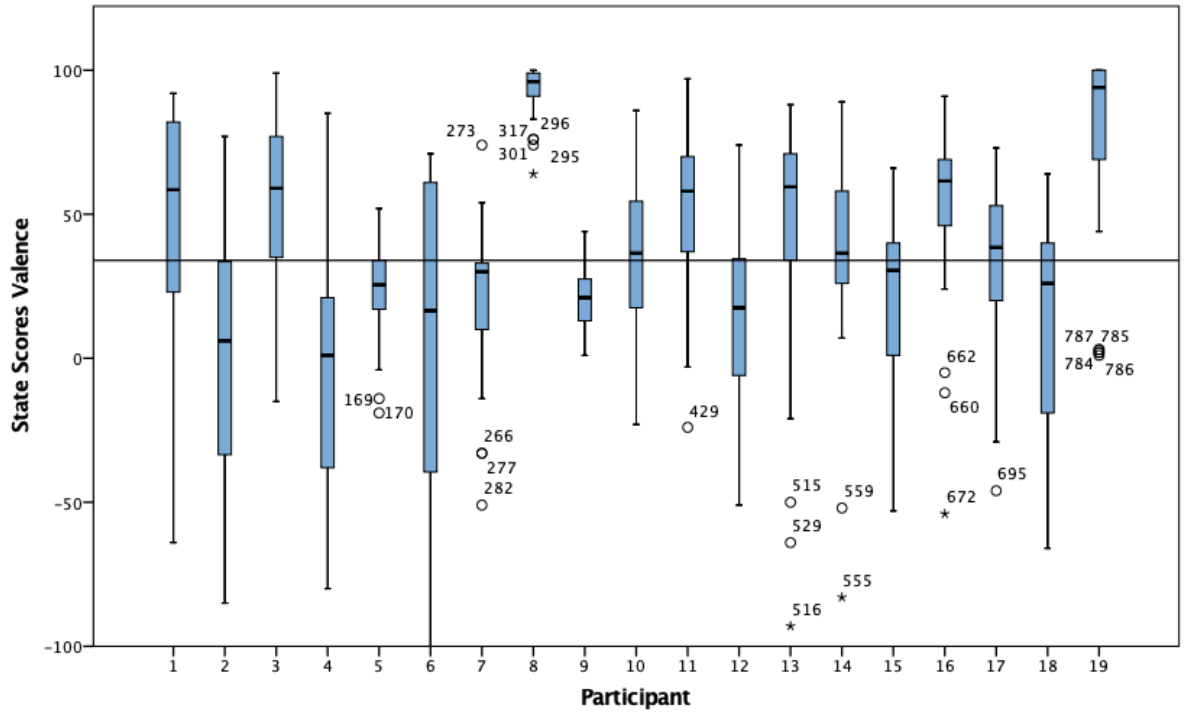

Figure 5. Boxplot illustrating the variation in experiencing valence for each participant based on the scores gathered from the Affect Grid, by also displaying a guiding line set at the group mean (M = 33.31). It is noteworthy that participants 8 and 19 display outlying mean scores.

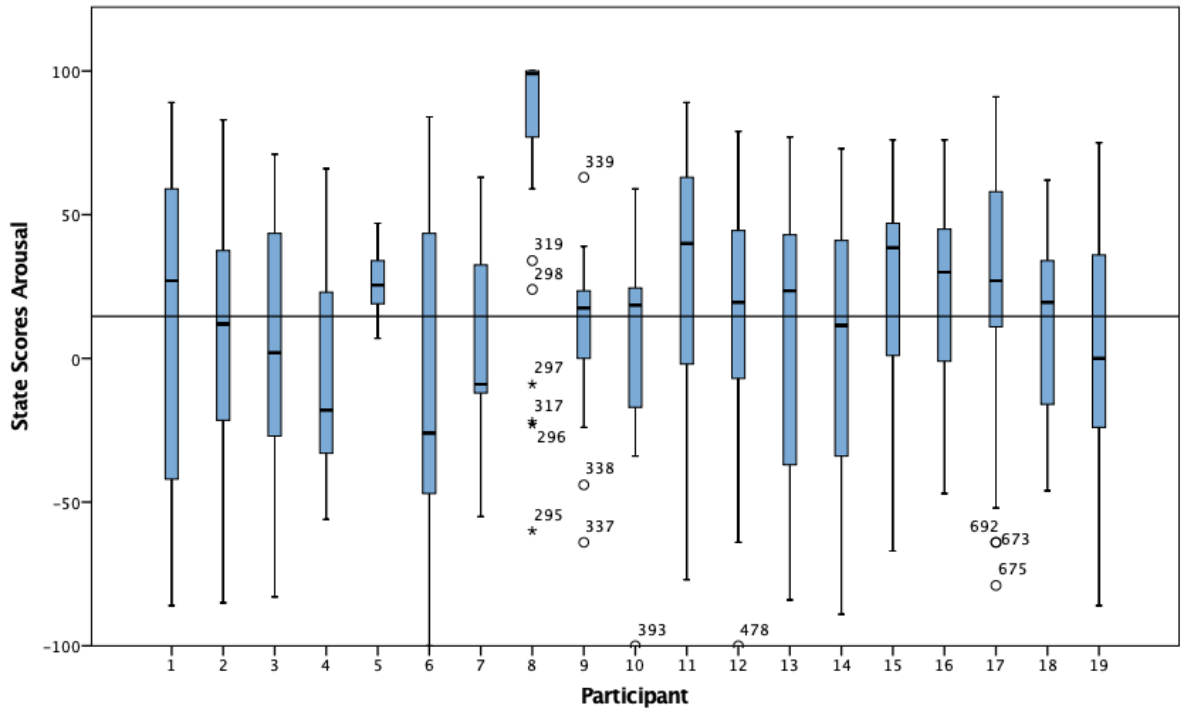

Figure 6. Boxplot illustrating the variation in experiencing arousal for each participant based on the scores gathered from the Affect Grid, by also displaying a guiding line set at the group mean (M = 14.39). Likewise to valence, participant 8 displays outlying scores.

3.2 Reliability assessment

Variability. Based on a mixed model analysis it is detectable that there is no significant linear effect between state arousal and time point (p = .083) as well as between state valence and time point (p = .715).

Stability of responses. Before calculating the correlation coefficients using a Pearson correlation analysis, scatterplots were created (Figures 7, 8 and 9). Based on the scatterplots, it is visible that especially the construct of valence tends to display a linear relationship, but also the scatterplots for arousal indicates a linear association between scores. Moreover, there were significant, positive correlations found between the scores of valence based on a Pearson correlation analysis (first and second half: r = .81; even and odd numbers: r = .92). Likewise the associations between the answers to Arousal, which turned out to be significant and positive (first and second half: r = .604; even and odd numbers: r = .749).

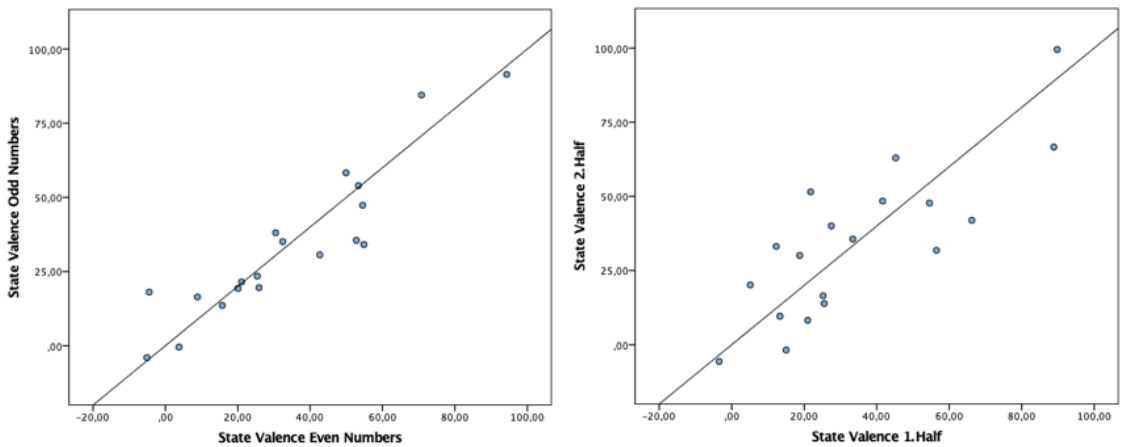

Figure 7. Scatterplots illustrating the strength of the linear relationship between the split halves (split in the middle, split between even and odd numbers) of answers to the Affect Grid for valence. A linear relationship can be assumed based on this visualization.

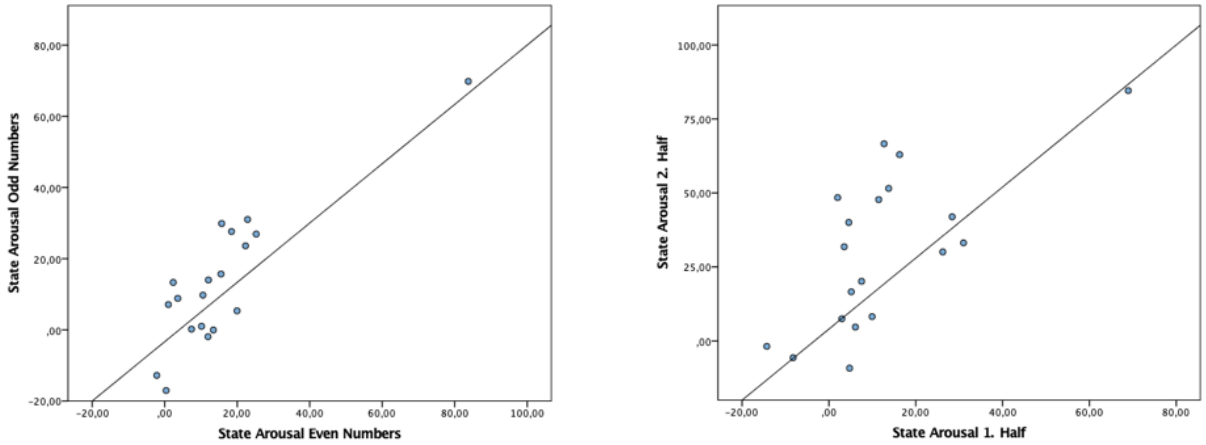

Figure 8. Scatterplots illustrating the relationships between the split halves (split in the middle, split between even and odd numbers) of answers to the Affect Grid for arousal. The scatterplots indicate one outlying participant. A linear relationship can be assumed based on these scatterplots.

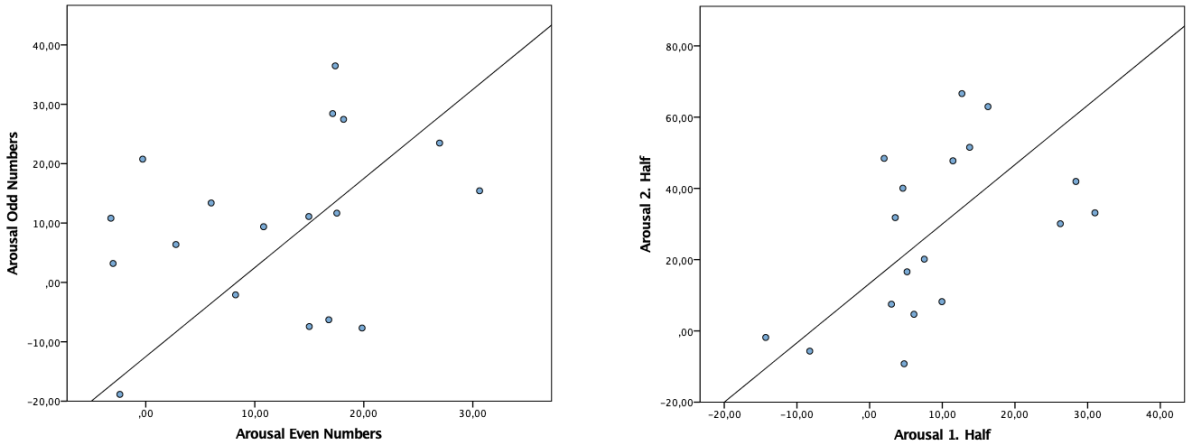

Figure 9. Scatterplot illustrating the relationship between the split halves of answers to the Affect Grid excluding participant 8, who can be outlined as outlier within this sample. It is noticeable that the range of the axes reduced to a maximum value of 30 (1. half) and 40 (2. half) after rejecting the diverging scores of participant 8.

3.3 Validity assessment

By estimating associations of the trait measurements with the state measurements in the framework of linear mixed modelling by including the trait scores as fixed covariate it was striking that PA is a significant covariate for state arousal ($p = .034$) and valence ($p = .001$) and NA is a significant covariate for valence ($p = .011$), but not for arousal ($p = .186$). Furthermore,

a Spearman correlation analysis was done by using the computed estimated means for participants and time points. A moderate and positive association was found between the state scores of valence and the trait score for PA (r = .33, p = .01) and a weak, positive relation between the state scores of arousal and the trait score for PA (r = .132, p = .01). Furthermore, a weak and negative association was found between state valence and trait NA (p = -.13, r = .01) and between state arousal and trait NA scores (p = -.08, p = .04). The significance for the associations is also visible in the provided Figures (Figures 10 and 11), where it noticeable that the scores for the state constructs and the trait constructs are related.

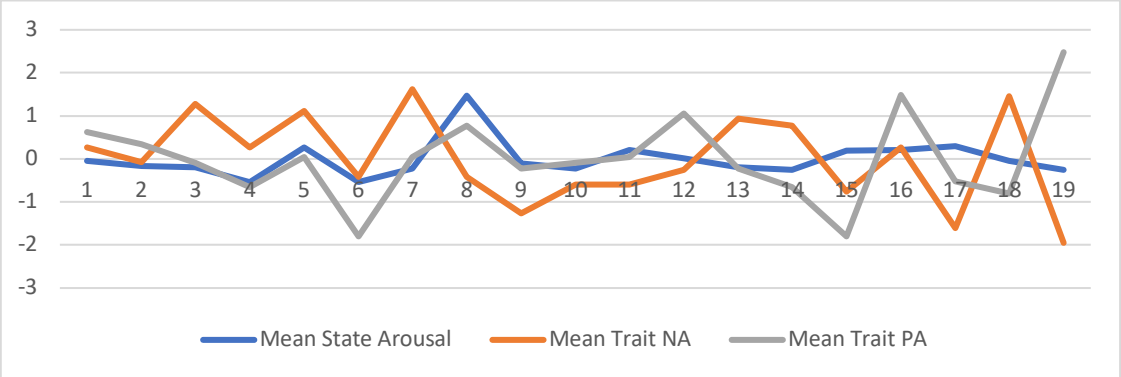

Figure 10. Line Graph illustrating the association between the mean scores of state arousal and the mean scores for trait NA and trait PA.

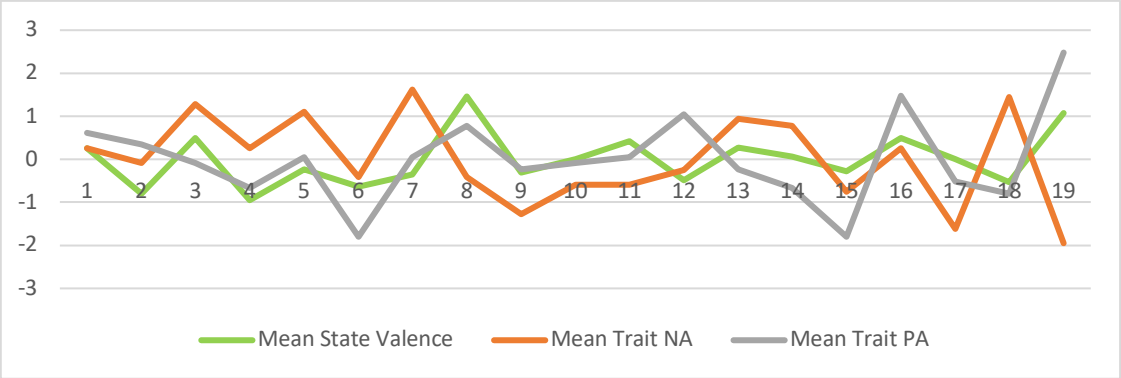

Figure 11. Line Graph illustrating the association between the mean scores of state valence and the mean scores for trait NA and trait PA.

3.4 Additional analysis with a lower compliance rate

Based on the fact that the compliance rate of 65% resulted in a high exclusion rate for the overall data of the participants, an additional analysis with a moderate compliance rate was conducted. The compliance rate was set to 40% which resulted in a threshold of 17 answered Affect Grids. This created a new sample of 29 participants who responded in sum to 962 Affect Grids (Valence: M = 17.86, SD = 43.47; Arousal: M = 36.19, SD = 40.67). The results of the sample with a lower compliance rate tend to be in line with the previous results without striking diverging scores, although more participants and thus, more data were included. The respective results of the bivariate correlation analysis can be found in Appendix F.

4. Discussion

4.1 Theoretical reflection

A considerable amount of literature has been published which have already made use of the ESM. However, psychometric properties of ESM questions remained unclear. This lack of knowledge raises an issue, because the assessment of psychometric properties can be outlined as the most fundamental aspect within health and social research (Dooley, 2011). Therefore, the present study was designed to create a first starting point to evaluate the reliability and validity of ESM questions.

Overall, the findings of this study provide preliminary support for the reliability and validity of the Affect Grid in the context of the ESM. Findings suggest good test-retest estimations, based on moderate to strong associations between aggregated scores. This coincides with previous studies investigating the stability of longitudinal data in the context of subjective experiences (e.g. Csikszentmihalyi, 2014; Kuppens, Oravecz and Tuerlinckx, 2010). Therefore, the stated hypotheses could not be rejected and the findings support the conclusion that the Affect Grid can capture affective baselines of people over time. Furthermore, validity values, including correlation coefficients of states scores of the Affect Grid and trait scores of the PANAS, showed weak associations. This finding is not in line with earlier findings reported by Hektner, Schmidt, and Csikszentmihalyi (2007), who proposed strong convergent validity for ESM measures and comparable one-time measurements in the framework of individual characteristics. In addition, it was surprising that the association between NA and arousal turned out to be weak and negative. This finding is contrary to the theoretical assumption by Feldmann (1995) which has suggested that higher levels of arousal are related with higher levels of NA. This inconsistency might be due to a matter of rotation, because PA and NA are assumed to be a 45° rotation of the arousal-valence space. Already Russel, Weiss, and Mendelsohn (1989) observed this inconsistency within their study. Nevertheless, these findings, while preliminary, indicated a trend of an agreement between the results of the Affect Grid and the PANAS.

It is interesting to note that all forms of split answers for both arousal and valence in this study displayed significant, (moderate to) strong, positive relationships. This also accords with earlier observations, although different approaches were taken in the ESM thus far by Freeman, Larson, and Csikszentmihalyi (1986), Kuppens, Oravecz and Tuerlinckx (2010), Diener and Larsen (1984) or Versluis et al. (2018). The first study by Freeman et al. (1986), which was concerned with changes in affective states, can be defined as more complex, due to more

frequent reporting of results as well as the inclusion of a follow-up week of testing after two years. Diener and Larson (1984) also find high stability and consistency estimates over a period of six weeks with two moments of reporting per day. This is different to the procedure of Versluis et al. (2018), who employed a rather short period of two days of collecting data with six times of signal-reporting in the context of emotional variability. Especially, findings by Kuppens, Oravecz and Tuerlinckx (2010) have to be outlined, because they also employed the Affect Grid as means of an ESM question. Different to the current study, they reported results ten times a day for two weeks within a first session and 50 times per day for four consecutive days during a second session. To date various time scales as well as different frequencies of reporting have led to similar observations. Hence, it could conceivable be hypothesized that the Affect Grid can detect stable patterns in affective experiences as it can also be concluded on basis of the current results.

By taking a closer look on the mean scores of the sample about the Affect Grid and the PANAS it was striking that the average responses for both could be rated as quite similar. This observation already indicates that the Affect Grid captured similar baselines as the PANAS, which suggests a first trend of convergence between the measures. Moreover, the correlation coefficients between aggregated state scores and trait scores indicated that both measures tend to reflect coinciding results. In sum, it is possible, that the Affect Grid in the ESM set up measures what it is intended to measure in accordance with traditional surveys like the PANAS. Such an interpretation is supported by earlier findings by Russel, Weiss, and Mendelsohn (1989), who already detected good estimates for convergent validity between the PANAS and the Affect Grid. These findings are also in line with more recent studies, which were described earlier by Versluis et al. (2018) as well as Littman et al. (2006), who explored the convergent validity between short ESM surveys and more complex one-time surveys. Both lines of research concluded that the ESM surveys equally captured the constructs of interest to the traditional surveys. It is therefore likely that short ESM questions like the one-item Affect Grid can measure constructs of interests just like multiple-item surveys.

Surprisingly, this study did not find significant differences between analysis on basis of different compliance rates, although the sample with a lower compliance rate included numerous missing values. The outcomes of the sample with a compliance rate of 40% indicated similar findings than the sample with a higher compliance rate of 65%. With respect to aggregated scores, this observation may support the hypothesis that missing values does not play an influential role for the predictive power of state scores produced in the ESM set up.

4.2 Practical and scientific implications

The theoretical context of the Affect Grid is grounded in the DynAffect account which was originally understood as a model for the dynamics of basic affect. However, it became also increasingly prominent in the abnormal range of variation (Ebner-Priemer et al., 2015). Not only bipolar personality disorders, but nearly every disorder is connected with dynamic fluctuations in core affect (Myin-Germeys et al., 2018). The insights gained from this study provided a building block for future research and gave a first indication that the Affect Grid is a reliable and valid tool and appears to be promising for measuring affective home bases for core affect. By reliably detecting the affective baseline of an individual, first suggestions about the affective experiences can be drawn. According to Ebner-Priemer et al. (2015) it can be assumed that more negative home bases give rise to the experience of more negative than positive emotions, which can be due to regulatory processes or personal characteristics. Therefore, the Affect Grid could represent a supplementary aid, if not the only one, in mental health research to control for mean scores of core affect.

Nevertheless, there is still room for further progress in determining the quality of the Affect Grid as an ESM question. While this study provided preliminary evidence detecting the average baseline of an individual, future research can also consider the variability in core affect. Recent research claimed that greater variability in emotions is associated with lower psychological well-being (Myin-Germeys et al., 2018; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007). In addition, from great interest is also the time needed for recovery, i.e. when the core affect is pulled back to the affective home base. Reliable and valid data about the variability in affective experiences can be helpful not only with identifying inadequate emotion regulation theories, but it can also support conclusions about future courses of symptoms (Myin-Germeys et al., 2018).

4.3 Limitations and future research

The current study poses several limitations that have to be considered. First of all, the scope of this study was limited to information about aggregated scores over time and the correlations between average scores on a trait and a state level. There is, therefore, a definite need to further investigate the quality of ESM questions from other angles, because several questions still remain to be answered. While this study focused on the affective home base, future studies should also take the other parts of the DynAffect account into consideration, namely the affective variation as well as the attractor strength.

Secondly, the correlation coefficients between the aggregated state scores (Affect Grid) and the trait scores (PANAS) turned out to be weak. Thus, the conclusion about convergent validity has to be drawn with caution. An explanation for this could be the possibility that the PANAS did not create a comparable accurate picture of the affective home base than the Affect Grid (Hektner, Schmidt, & Csikszentmihalyi, 2007). To develop a full picture of convergent validity, additional studies will be needed that implement a comparable survey which is more accurate in the beginning, like the extended form of the PANAS (PANAS-X), which includes 60 items instead of 20 items. Watson and Clark (1984) suggested, that the PANAS-X can draw a more accurate picture of the positive and negative affect of participants, compared to the PANAS.

Thirdly, the current study implemented an Affect Grid which displayed the dimension of arousal ranging between "high arousal" and "low arousal". According to various definitions, arousal describes the dimension with high arousal at the one end to lethargy or sleepiness at the other end (Russel, Weiss, & Mendelsohn, 1989). Therefore, the current study misses the range between low arousal and lethargic. This would be a fruitful area for further work.

4.4 Conclusion

The present study has been one of the first attempts to thoroughly examine the reliability and validity of ESM questions in the field of human feelings. Based on the current findings it can be suggested that the Affect Grid entails good test-retest reliability and convergent validity with the PANAS in the context of the ESM. Furthermore, this study provided a first starting point for the influence of missing values in the ESM by considering different compliance rates. Finally, these findings can be used as a building block for further research to fill the current knowledge gap about the psychometric properties of ESM surveys.

References

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.

Camras, L. A., & Witherington, D. C. (2005). Dynamical systems approaches to emotional development. *Developmental Review*, *25*(3-4), 328-350.

Chen, Y. W., Cordier, R., & Brown, N. (2015). A preliminary study on the reliability and validity of using experience sampling method in children with autism spectrum disorders. *Developmental Neurorehabilitation*, *18*(6), 383-389.

Cohen, L. (1992). Power Primer. *Psychological Bulletin,* 112(1), 155-159.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(3), 245–265. https://doi.org/10.1348/0144665031752934

Csikszentmihalyi, M. (2014). *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. https://doi.org/10.1007/978-94-017-9088-8

Diener, E., & Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of personality and social psychology*, *47*(4), 871.

Dooley, D. (2001). *Social Research Methods* (4th ed.). New Jersey, USA: Prentice-Hall.

Ebner-Priemer, U. W., Houben, M., Santangelo, P., Kleindienst, N., Tuerlinckx, F., Oravecz Z., ... & Kuppens, P. (2015). Unravelling affective dysregulation in borderline personality disorder: A theoretical model and empirical evidence. *Journal of abnormal psychology*, *124*(1), 186.

Eysenck, H. J. (1968). Eysenck personality inventory manual. *San Diego: Educational and Industrial Testing Service*

Francis, L. J., Lewis, C. A., & Ziebertz, H. (2006). The short-form revised Eysenck personality Questionnaire (EPQ-S): A German edition.

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, *69*(1), 153.

Freeman, M., Csikszentmihalyi, M., & Larson, R. (1986). Adolescence and its recollection: Toward an interpretive model of development. *Merrill-Palmer Quarterly (1982-)*, 167-185.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). Psychometrics of ESM data. *Experience sampling method: Measuring the quality of everyday life*, 103-121.

Hoey, M. (2012). *Lexical priming: A new theory of words and language*. Routledge.

Jaccard, J. J. (1974). Predicting social behavior from personality traits. *Journal of Research in Personality*, *7*(4), 358-367.

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.

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*(2), 262.

Larsen, R. J. (2000). Toward a Science of Mood Regulation. *Psychological Inquiry*, *11*(3), 129–141. https://doi.org/10.1207/s15327965pli1103_01

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In *Flow and the foundations of positive psychology* (pp. 21-34). Springer, Dordrecht.

Littman, A. J., White, E., Satia, J. A., Bowen, D. J., & Kristal, A. R. (2006). Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology*, *17*(4), 398 403. https://doi.org/10.1097/01.ede.0000219721.89552.51

Mehl, M. R., & Conner, T. S. (2012). *Handbook of Research Methods for Studying Daily Life*. New York: The Guilford Press. Retrieved from http://ezproxy2.utwente.nl/login?url=http://search.ebscohost.com/login.aspx?direct=true &db=nlebk&AN=408484&site=ehost-live

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.

Nowak, A., & Lewenstein, M. (1994). *Dynamical systems: A tool for social psychology?*. Academic Press.

Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A. E. G., ... & Dunn, G. (2011). Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica*, *123*(1), 12 20.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Russell, Y. I., & Gobet, F. (2012). Sinuosity and the affect grid: a method for adjusting repeated mood scores. *Perceptual and motor skills*, *114*(1), 125-136.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal.. *Journal of Personality and Social Psychology*, *57*(3), 493–502. https://doi.org/10.1037/0022-3514.57.3.493

Scherer, K. R. (1982). Emotion as a process: Function, origin and regulation. *Social Science Information*, *21*(4-5), 555–570. https://doi.org/10.1177/053901882021004004

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389-408.

Vallacher, R. R., & Nowak, A. E. (1994). *Dynamical systems in social psychology*. Academic Press.

Watson, D: 1988, 'The vicissitudes of mood measurement: Effects of vary- ing descriptors, time frames, and response formats on measures of positive and negative affect', Journal of Personality and Social Psychology 55(1), pp. 128–141.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales.. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form.* Ames: The University of Iowa.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of personality and social psychology*, *76*(5), 820.

Appendices

Appendix A

The Positive and Negative Affect Scale (PANAS)

To what extend do you generally feel…	(1) not at all	(2) a little	(3) moderately	(4) a lot	(5) very much
1. interested?					
2. distressed?					
3. excited?					
4. upset?					
5. strong?					
6. guilty?					
7. scared?					
8. hostile?					
9. enthusiastic?					
10. proud?					
11. irritable?					
12. alert?					
13. ashamed?					
14. inspired?					
15. nervous?					
16. determined?					
17. attentive?					
18. jittery?					
19. active?					
20. afraid?					

Appendix B

Informed Consent

Dear participant!

Welcome to this one-week experience sampling study. With this study you will experience 'Experience sampling'. A relatively new and popular methodology in psychological research in which participants answer a limited amount of questions a number of times per day during their normal everyday life. In the current case, this will only be one question six times per day over a period of seven days.

You are herewith being invited to participate in our study *How to measure how you feel.* This study is being done by us, *Laura Stevens, Julia Müller, and Tim Tiede* from the Faculty of Behavioural, Management and Social Sciences at the University of Twente.

The purpose of this research study is to evaluate the validity of the Affect Grid and will take you approximately 5 minutes per day over a period of seven days to complete. Before this period begins, you will quickly fill in three questionnaires. The data obtained will be used for statistical analyses of the psychometric properties of the Affect Grid and constructs related to ratings in feelings.

Your participation in this study is entirely voluntary and you can withdraw at any time. Furthermore, you are free to omit any question at any time.

We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by storing the data obtained safely and as anonimously as possible.

Once you created an account, please download the app TiiM - the incredible intervention machine and log in with the chosen credentials. The study will begin at 10:00 am the next day. **You will receive an e-mail to the address of your account with all relevant information. Please read it carefully before you start the study.**

In case of any open questions, issues, or critique please do not hesitate to contact

Julia Müller (j.muller-1@student.utwente.nl)
Tim Tiede (t.tiede@student.utwente.nl)
or
Laura Stevens (l.s.stevens@student.utwente.nl)

Thank you for your interest and willingness to participate

Sincerely,
Julia Müller, Tim Tiede, and Laura Stevens.

Appendix C

E-Mail

This E-Mail was sent to all participants before the start of the study:

"Thank you for taking part in our study.

In the following, we will provide a short outline on the set-up and the theoretical background of this study. In general, this study serves the investigation of changes in core affect over time.

Core affect represents what is commonly called mood or feelings, and consists of two dimensions- valence and arousal. **Valence** represents how **pleasant** you feel at the moment and ranges from **unpleasant** to **pleasant**. **Arousal** represents how **activated** you feel at the moment and ranges from **sleepy** to **activated**.

These two dimensions are represented in the **Affect Grid**, the test you will fill in repeatedly in the course of this study. The Affect Grid is a coordinate system with *valence represented on the x-axis* (horizontal) and *arousal on the y-axis* (vertical). You can refer to the point where both axes cross as *'neutral state of feelings'*. By setting a mark somewhere in this coordinate system you indicate how you feel- both pleasant/unpleasant and sleepy/activated in one mark!

(this means, that if you feel very active and pleasant you set a mark far to the right upper corner, if you feel pleasant but rather sleepy/inactive you set it far to the right but in the lower corner)

(please read the information above carefully- understanding this is necessary to successfully fill in the Affect Grid!)(you can also make a screenshot of this description in case you want to read it at a later point again)

Your task is *to set a mark in the coordinate system every two hours between 10:00 and 20:00 o'clock* (six times per day) for 7 days. You do not have to keep track of the time yourself, the app will send you a notification every two hours within this time frame to fill in the Affect Grid. *(therefore please allow the app to send you notifications!)*

If you are not able to fill in the grid right away, you can do so a few minutes later, but please try to do so in time. We hope that you can answer as many times as possible!

To start the study, please download the app TiiM - The incredible intervention machine and log in with the credentials you chose. You will start by answering three questionnaires and then the study will begin- We hope you are as excited about this study as we are!

Best,
Julia Müller, Laura Stevens, and Tim Tiede.
PS: In case of any open questions, issues, or critique please do not hesitate to contact
Julia Müller (j.muller-1@student.utwente.nl), Laura Stevens (l.s.stevens@student.utwente.nl)
or Tim Tiede (t.tiede@student.utwente.nl)

Appendix D
Descriptive Statistics of the Affect Grid

Descriptive Statistics of the longitudinal scores of Valence and Arousal for each participant.

ID		N	Minimum	Maximum	Mean	Std. Deviation
2243	Valence	46	-64	92	42.78	44.24
	Arousal	46	-86	89	9.85	54.65
2245	Valence	42	-85	77	.40	37.91
	Arousal	42	-85	83	9.69	40.34
2248	Valence	46	-15	99	52.37	31.73
	Arousal	46	-100	71	.61	44.93
2250	Valence	48	-80	85	-5.13	38.04
	Arousal	48	-84	66	-10.96	33.10
2251	Valence	48	-19	52	22.96	15.32
	Arousal	48	1	47	25.31	10.36
2252	Valence	42	-100	83	12.57	52.53
	Arousal	42	-100	92	.10	51.08
2255	Valence	42	-51	74	21.48	26.90
	Arousal	42	-55	63	4.57	30.75
2263	Valence	47	64	100	91.77	8.98
	Arousal	47	-60	100	78.00	37.51
2265	Valence	46	1	44	21.17	10.73
	Arousal	46	-64	63	9.07	23.97
2267	Valence	28	-23	86	34.54	30.50
	Arousal	28	-100	59	5.25	32.10
2268	Valence	29	-24	97	50.79	30.14
	Arousal	29	-77	89	22.97	48.05
2269	Valence	42	-51	74	11.33	31.22
	Arousal	42	-100	79	14.60	39.94
2288	Valence	38	-93	88	45.03	42.91
	Arousal	38	-84	77	6.08	47.26
2289	Valence	42	-83	89	36.64	32.06
	Arousal	42	-89	73	3.79	44.54
2294	Valence	42	-53	66	22.71	27.52
	Arousal	42	-67	76	22.79	37.43
2310	Valence	42	-54	91	54.10	26.75
	Arousal	42	-47	76	23.02	29.58
2311	Valence	42	-46	73	33.76	27.30
	Arousal	42	-79	91	26.93	42.15
2312	Valence	42	-66	64	12.64	32.77
	Arousal	42	-46	62	13.02	26.20
2334	Valence	41	1	100	77.46	32.30
	Arousal	41	-86	75	3.98	42.48

Appendix E
Table including the mean score for each Dimension of PA and NA for each participant.

Descriptive Statistics of the final Sample (N = 19) for the PANAS.

ID	Mean Positive Affect*	Mean Negative Affect**
2243	2.5	3.7
2245	2.3	3.7
2248	3.1	3.2
2250	2.5	2.8
2251	3.0	3.3
2252	2.1	2.0
2255	3.3	3.3
2263	2.1	3.8
2265	1.6	3.1
2267	2.0	3.2
2268	2.0	3.3
2269	2.2	4.0
2288	2.9	3.1
2289	2.8	2.8
2294	1.9	2.0
2310	2.5	4.3
2311	1.4	2.9
2312	3.2	2.7
2334	1.2	5.0

*Mean of Items 1, 3, 5, 7, 9, 11, 13, 15, 17, 19
** Mean of Items 2, 4, 6, 8, 10, 12, 14, 16, 18, 20

Appendix F
Results of an additional analysis with the full dataset

Results of a Spearman Correlation Analysis for the Associations between State and Trait Scores based on a sample with a lower Compliance Rate.

	N	Spearman Correlation	Sig. (2-tailed)
State Valence and Trait PA	29	.47	.01
State Valence and Trait NA	29	-.27	.16
State Arousal and Trait PA	29	.24	.22
State Arousal and Trait NA	29	-.09	.63