The Construct Validity and Split-Half Reliability of the Affect Grid in Experience Sampling: Alexithymia and Fluctuations in Core Affect

Bachelor's Thesis

Author	Tim Tiede
Student Number	s1872958
Academic Year	2018/2019
First Supervisor	Dr. Matthijs L. Noordzij
Second Supervisor	Jannis T. Kraiss
Institute	University of Twente
Faculty	Faculty of Behavioral, Management, and Social Sciences (BMS)
Program	Psychology
Specialization	Positive Psychology and Technology (PPT)

Abstract

The affect grid is a single-item measure of the two dimensions of core affect, valence and arousal. Due to its simplicity and time-efficiency it has been frequently used in experience sampling – which is based on repeated, long-term measurement with a single measure in a naturalistic environment. Despite its increasing popularity in this area however, no study so far has validated the psychometric properties of the affect grid in the context of the experience sampling method (ESM). The aim of the current study is to investigate the split-half reliability and construct validity of the affect grid in experience sampling. Split-half reliability was tested for with respect to the within-participant mean scores, standard deviations (SD), and mean squared successive differences (MSSD) of valence and arousal. Construct validity was examined with focus on the fluctuation-characteristics of core affect, represented by the participants' standard deviations and MSSD of valence and arousal, and their relation to alexithymia, measured by two subscales of the TAS-20. The affect grid exhibited adequate split-half reliability in all tested facets. With respect to construct validity, the expected relation between alexithymia and within-participant SD and MSSD was found, however, only in case of arousal. In light of research that provides evidence for a connection of alexithymia to arousal, but not valence, these findings indicate adequate construct validity. Conclusively, this study found the affect grid to exhibit adequate reliability and suggests that it is a valid instrument in experience sampling. Nevertheless, additional validation with respect to construct validity appears necessary.

The Construct Validity and Split-Half Reliability of the Affect Grid in Experience Sampling: Alexithymia and Fluctuations in Core Affect

1. Introduction

Around the year 1885, a young researcher by name of Hermann Ebbinghaus began studying memory dynamics in a single-case design: He memorized throughout two one-year periods around 1300 lists of syllables and examined the rate of forgetting and re-learning over time. The procedure he applied was subjected to several means of standardization such as fixed times of the day, constant rates of repetition, and the ideal of a naturalistic setting (Ebbinghaus, 1885). Because an individual's memory is ever-changing, the investigation of its dynamics required these special conditions to allow firm conclusions. The same applies to the phenomenon of core affect. Core affect represents a person's mood, or feeling and is experienced as dynamic, momentary state. Hence, a single measurement only provides little information about variations in core affect. Here, as in the case of Ebbinghaus, special measurement procedures are of importance that allow the assessment of fluctuations.

The experience sampling method (ESM) is a procedure similar to the way Ebbinghaus approached the dynamics of memory. Both share the characteristics of frequent, repeated sampling of the same individual[s], which requires a measure that can be filled in quickly, without much effort and independent of the participants' location. A measure feasible in this regard is the affect grid. The affect grid is a frequently used measure of core affect that has proven useful to experience sampling research (e.g. Kuppens, Oravecz, and Tuerlinckx (2010); Colomo-Palacios, Casado-Lumbreras, Soto-Acosta, and Garcia-Crespo (2011)). With respect to its psychometric properties, the affect grid has been validated in the assessment of traits, mostly in the context of single use (Killgore, 1998; Russell et al., 1989). However, when it comes to long-term measurement of core affect with the ESM, information on the validity or reliability of the affect grid are nonexistent. This is striking in view of its increasing use in experience sampling.

On these grounds, this study is intended to investigate the split-half reliability of the affect grid and its construct validity in the context of the ESM. Here is to add that, for validity, the focus lies on the long-term fluctuations of core affect.

2. Theoretical Framework

Experience sampling

Experience sampling, also referred to as ecological momentary assessment, time budget sampling, or ambulatory assessment, is a sampling procedure with a long history that became increasingly popular during the affective revolution in the 1980s (Schimmack, 2003). Its main purpose is the investigation of dynamic constructs over a potentially long period of time (Csikszentmihalyi, 1992; Palmier-Claus et al., 2011; Trull & Ebner-Priemer, 2013). Whereas a test in common research is conducted only once or a few times, the ESM differs in that it is based on short-interval, repeated testing of the same measure over a potentially long period of time (Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009).

This can be advantageous in the measurement of dynamic phenomena, as the used test typically can be filled in quickly and with ease. Therefore, after the briefing no further supervision is required so that the measurement can be conducted ecologically valid in a naturalistic environment by means of wearables such as smartphones (Palmier-Claus et al., 2011). These characteristics allow an insightful investigation in the fluctuations of dynamic phenomena such as an individual's core affect (Csikszentmihalyi & Larson, 2014; Ebner-Priemer et al., 2009; Russell, 2003).

Core affect and the Affect Grid

Core affect as introduced by Russell (2003) is a conceptualization of emotion that consists of two distinct dimensions - *valence*, ranging from pleasure to displeasure, and *arousal*, ranging from sleepy to activated. These two dimensions combined represent what is commonly called mood, emotion, feelings or affective state.

A prominent measure of core affect also used in experience sampling is the *affect grid*, introduced by Russell et al. in 1989 and since then used in varying designs (e.g. Ebner-Priemer et al. (2015)). It is a single-item measure consisting of the dimensions compromised in core affect, *valance*, located along the horizontal axis, and *arousal*, along the vertical axis. The present study's affect grid is depicted in Figure 1. The affect grid has been validated in use for single assessment of facial expressions or emotion-related words and exhibited adequate reliability and convergent as well as discriminant validity (Killgore, 1998; Russell et al., 1989). However, as aforementioned, not with regard to its psychometric properties in the context of the ESM. One model that involves long-term dynamics of core affect is the DynAffect.

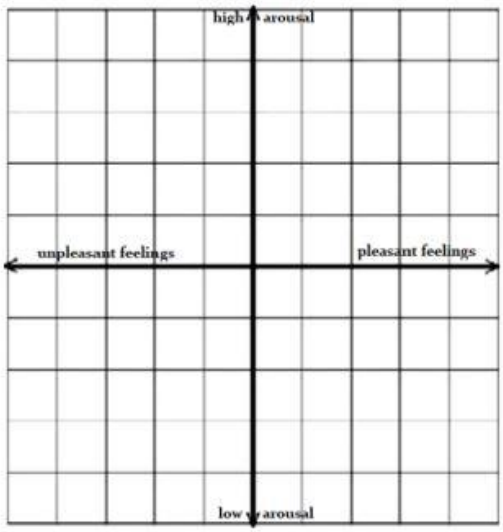

Figure 1. The affect grid administered in the current study. The participants set a mark in the space of the coordinate system to indicate their momentary state of feelings. The crossing point of the two axes indicated a neutral state.

DynAffect

The DynAffect is an explanatory model introduced by Kuppens et al. (2010) that describes the dynamics of valence and arousal based on three propositions: Every individual has a characteristic, and fairly stable 1. *affective home base*, 2. individual *core affect variability*, and 3. *attractor strength* (Ebner-Priemer et al., 2015; Kuppens et al., 2010). The affective home base is a person's affective comfort zone, the area in which the core affect is in average located. (Kuppens et al., 2010). Core affect variability refers to the individuals' typical range of core affect-fluctuation around the affective home base, whereas the regulatory processes that, with individual pace, redirect the core affect towards the home base are called attractor strength. Conceptually, these are represented in an individual's core affect mean score (home base), standard deviation or variance (core affect variability), and pace of change, represented through the mean squared successive difference [MSSD] (attractor strength) (Ebner-Priemer et al., 2015; Kuppens et al., 2010).

Validation of a measure

Commonly in scientific practice, to accept the outcomes of a measure as representative it must be *validated* with regard to its content, the criteria, and the construct, overall referred to as validity.

Validity. *Content validity* refers to the question whether the relevant facets of the construct are represented in the test items, and is commonly investigated by means of a factor analysis (Dooley, 2009). As the affect grid consists of only one item, content validity in this

case is not verifiable (Russell et al., 1989). *Criterion validit*y refers to the correspondence of the measure's outcome with the situation at a specific time- either the present state, called concurrent validity, or a future situation, called predictive validity (Dooley, 2009). Finally, *construct validity* is the extent to which a test measures what it is supposed to measure (Cronbach & Meehl, 1955). The validation of the construct is achieved by correlating a measure's outcome with other constructs that are known to be related (Dooley, 2009). This is also a common way of validity assessment in experience sampling and hence in line with the purpose of the present study (Csikszentmihalyi & Larson, 2014; Palmier-Claus et al., 2011). However, in contrast to the validation of the affect grid in the past, experience sampling creates the opportunity to validate the affect grid with respect to its long term-characteristics, in other words when measuring the *fluctuation* of core affect over time (Palmier-Claus et al., 2011; Russell et al., 1989). A construct known to be related to long term-fluctuations in emotional experience is alexithymia, which therefore will be the means for the investigation of the affect grid's construct validity in the present study.

Alexithymia

Alexithymia is a personality construct characterized by difficulties to perceive and describe own and others' emotional states (Taylor, Ryan, & Bagby, 1985; Versluis et al., 2018). Moreover, alexithymic individuals have difficulties in attending to and distinguishing own emotions (Lane, Sechrest, Riedel, Shapiro, & Kaszniak, 2000; Lane et al., 1996; Thompson, Dizén, & Berenbaum, 2009). This appears to be especially relevant when reporting own emotions over a longer period of time, as it is to expect that individuals with difficulties in distinguishing emotional states exhibit less variability in their reported emotional experience. Additionally, as emotion recognition should be slower when being inattentive, a slower change in reported emotions is expected.

Reliability. Next to validity, the second factor relevant to the adequacy of a measure is reliability. A reliable measure produces close-to identical results in the same measurement situations (Dooley, 2009). Reliability is commonly separated into parallel test-, interrater-reliability, and internal consistency (Dooley, 2009; Field, 2014).

Test-retest reliability indicates whether a measure produces stable results when being conducted at different points in time, whereas parallel test-reliability looks at the correspondence of the results of one test to those of another, conceptually identical (parallel) test, conducted at the same time (Dooley, 2009). In studies that require categorizing or coding of collected materials by independent raters, interrater-reliability represents the agreement

between the raters in the assessment of the materials (Dooley, 2009). Interrater and test-retest reliability are inapplicable when measuring core affect with the ESM, as here (1). a personal state is assessed, wherefore no rater but the individuals themselves must 'rate', and (2.) changes are expected, so that a second measurement only will most likely show different results. Finally, *internal* consistency refers to the reliability of the test-items *within* the measure, examining whether the results of different parts of the measure are indicating the same outcome (Dooley, 2009). Two indices of internal consistency are 'Lambda 2' and 'Cronbach's alpha', which, however, are inapplicable to single-item measures such as the affect grid, as they require tests with a number of different items (Cronbach, 1951; Dooley, 2009; Russell et al., 1989).

In a basic form, internal consistency tested for by either splitting the results in half and comparing them (split-half reliability), or by comparing the results of all even with all odd-numbered items of a measure (odd-even reliability) (Dooley, 2009). Split-half reliability is the recommended manner to test for reliability in experience sampling as it gives insight into stable patterns throughout the scorings (Csikszentmihalyi & Larson, 2014; Palmier-Claus et al., 2011). The assumption to find stable patterns in core affect might seem counterintuitive considering that the ESM is usually applied to dynamic constructs, however, past research in experience sampling found individuals to exhibit relative stability in their everyday emotional experience (Csikszentmihalyi & Larson, 2014). In line with this reasoning, Kuppens et al. (2010) introduced an explanatory model that specifies these individual characteristics with regard to core affect under the name DynAffect.

The present study

The affect grid is a measure that becomes a increasingly popular instrument for ESM. However, to this point in time, no information exists on its psychometric properties in this context. Building on the DynAffect and the characteristics of alexithymia, this study aims to validate the affect grid under the conditions of the experience sampling method. Split-half reliability will be investigated with regards to the within-person mean score (home base), standard deviation (core affect variability) and MSSD (attractor strength). To assess construct validity, the characteristic variability and MSSD in core affect will be correlated to the participants' results on the TAS-20. The hypotheses are as follows:

H1: The affect grid exhibits adequate split-half reliability with regard to the within participant mean scores

H2: The affect grid exhibits adequate split-half reliability with respect to the within participant standard deviation

H3: The affect grid exhibits adequate split-half reliability concerning the within participant mean squared successive difference (MSSD).

H4: The participants' scores on the subscales of the TAS-20 are negatively correlated with the standard deviation of their ratings on the affect grid.

H5: The participants' scores on the subscales of the TAS-20 are negatively correlated with the mean successive difference of their ratings on the affect grid

3. Method

Design

The present correlational study involved the experience sampling method (ESM) in a one-group repeated measures design to test for the within-participant fluctuation of core affect over a period of one week. All participants filled in 3 questionnaires at the beginning of the study followed by 42 core affect measures (six times a day, over a period of seven days). The data was gathered by means of a mobile application developed by the University of Twente, called TiiM (The incredible intervention Machine). The study was approved in May 2019 by the ethics committee of the University of Twente, with data collection taken place in May and April of the same year.

Participants

For the purpose of this study, 52 participants were recruited by means of convenience sampling. Most participants were befriended students of the researchers, studying at German universities. Due to technical issues with the app, 7 participants (13.47%) had to be excluded from the analyses. Of the remaining 45 participants, 43 were German. Overall, 29 participants (64.40%) were female and 16 (35.60%) male, with age ranging from 18 to 52 (mean age of 22.76; SD= 6.17).

Inclusion criteria. For individuals to become part of the study, requirements included adequate English proficiency with regard to reading and understanding, as well as permanent access to a smartphone with stable internet connection. Finally, the individual had to express

willingness to install the app TiiM and to spend the respective time to repeatedly fill in the measures.

Materials and Measures

TiiM. TiiM (The incredible intervention Machine) is a free application available for IOS and Android that has been developed by the BMS-lab of the University of Twente. This app was the means of administration of the questionnaires and the affect grids and the platform within which the researchers created the study. All measures, their sequence and their timing in the way they would later appear on the smartphone of the participants were set-up through TiiM.

Demographic Data. At the beginning of the study, the participants were asked through the app to provide information on age, nationality, and gender.

The Toronto Alexithymia Scale (TAS). To measure for alexithymia, two of the three subscales of the Toronto Alexithymia Scale (TAS-20) were administered*; difficulties in identifying feelings* (DIF), consisting of 7 items (e.g. *'I am often confused about what emotion I am feeling'*) and the subscale *difficulties in describing feelings* (DDF), consisting of 5 items (e.g. *'It is difficult for me to find the right words for my feelings'*) (Bagby, Parker, & Taylor, 1994; Taylor et al., 1985) [For a complete overview of the items see Appendix B]. In accordance with the original measure, the 12 items were answered on a 5-point Likert scale, ranging from 1 = strongly disagree to 5 = strongly agree. By reason of insufficient reliability the author omitted the third subscale of the TAS-20, *externally oriented thinking* (EOT) in this study (Kooiman, Spinhoven, & W Trijsburg, 2003).

Psychometric properties of the TAS. The TAS-20 is a well validated measure introduced by Bagby et al. (1994), exhibiting high internal consistency (Cronbach's alpha =.81, with values of .75 found for both subscales) and adequate split-half reliability (.77 with $p < .01$) (Bagby et al., 1994; Versluis et al., 2018). Mixed results have been reported with regard to the factor structure, which, even though found to be insignificant by Leising, Grande, and Faber (2009), has frequently been validated (Parker, Taylor, and Bagby (2003); Parker, Michael Bagby, Taylor, Endler, and Schmitz (1993); Soo Seo, Chung, Deog Rim, and Jeong (2009)). These results were confirmed in the present study, with a Cronbach's alpha of

.78 for both subscales combined, .70 for DIF and .81 for DDF. Lambda 2 was satisfactory with .82 for both subscales combined, with .74 for the DIF and .82 for DDF.

The Affect Grid. To measure the participants' core affect, the test that has been used is a slightly adjusted version of the original one-item affect grid by Russell et al. (1989). As visible in Figure 2, the version of the present study indicated the two dimensions, valence and arousal, in the form of axes, depicting a coordinate system. The horizontal axis represented valence (with the indication words 'unpleasant feelings' to the left and 'pleasant feelings to the right') and the vertical axis arousal ('high arousal' at the top and 'low arousal' at the bottom). In this aspect, it is similar to the affect grid used by Kuppens et al. (2010), as they also included two axes and indication words at their tails. However, it is different in that the affect grid by Kuppens et al. (2010) consisted of a 99x99 two-dimensional grid, whereas that of the present study is a 100x100 grid, consisting of 10x10 squares. Here is to add, that the researchers decided against the commonly used arousal-sleepiness terminology as indication words for the arousal-dimension and instead used the terms 'low arousal' and 'high arousal'. This way, so-called lexical priming through the potential negative connotation of sleepiness should be avoided (for a comprehensive overview on lexical priming see Hoey (2012). Throughout the study the participants received the instruction 'please indicate how you are feeling' when a measurement took place and were allowed to set a mark anywhere in the grid. The crossing point of both axes represented a neutral state.

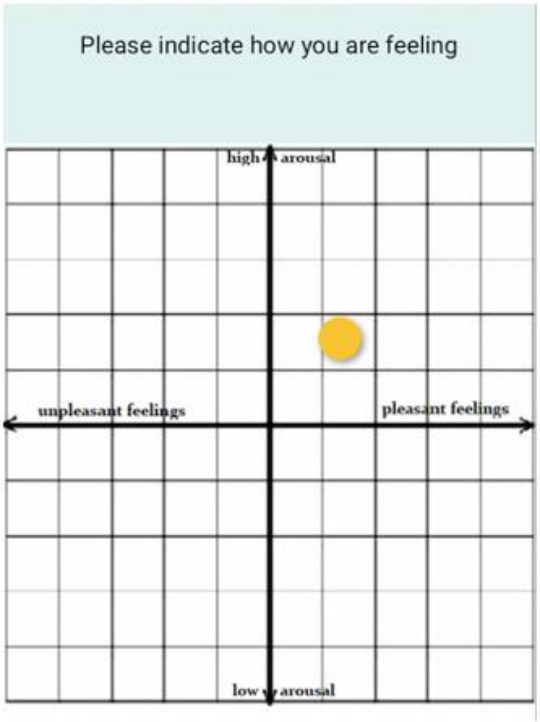

Figure 2. The affect grid of the present study with a score of 27/31 on valence and arousal, respectively.

Psychometric Properties of the affect grid. The affect grid has been found to exhibit adequate psychometric properties; Construct validity has been found to be moderate to high. Both, Killgore (1998) and Russell et al. (1989) found adequate to high convergent validity (.74 to .94 for valence and .63 to .92 for arousal) as well as high construct validity, supporting the orthogonality of arousal and valence (see Russell (2003)). With regard to its reliability, Russell et al. 1989 reported high split-half reliability when being tested for emotion-related terms in multiple studies (.98 for valence and .97 for arousal; n=25). Furthermore, the affect grid has been successfully used to validate affective characteristics of Borderline Personality Disorder (BPD) and individuals in general, as expressed in the DynAffect (Ebner-Priemer et al., 2015; Ebner-Priemer et al., 2009; Kuppens et al., 2010). However, no information exists with regard to the reliability of long-term assessments of personal affective states

Procedure

Preparation. In alignment with recommendations by Conner and Lehman (2012), the researchers ran several pilot tests that they completed themselves after having received ethical approval by the University's ethics committee.

The study consisted of a one-week data collection with the ESM that took place by means of TiiM. In this time period, every participant was instructed to fill in 42 affect grids preceded by three questionnaires. Due to technical issues, one passage had to be re-launched, however, without affecting the data. All students read through an opening statement before the start of the data collection and received detailed instructions via mail after they created an account for the study (for the instructions, see Appendix A*)*. To create an account, the participants had to follow a link provided by the researchers that required to fill in an email address and to choose a custom password. By doing so, the participants gained access to the free app 'TiiM'.

When the researchers had assigned the participants to the study, the measurement started the next day. It began with three questionnaires that preceded the repeated administration of the affect gird. These questionnaires took the participant around 10 minutes to fill in. Afterwards, the first affect grid appeared at 10:00am in the app, succeeded by six more on the respective day, in steps of two hours (at 12, 14, 16, 18, and 20 o'clock). This pattern repeated the following six days. To minimize the number of skipped measurements, the participant received a notification 'please indicate how you are feeling at the moment!' on the mobile screen every time a measurement was scheduled. If not filled in after 30 minutes,

the notification 'Hey! Please take a moment to set a mark', would appear and the test would remain available in the app for another 60 minutes before finally disappearing. After the 42nd affect grid had been filled in on the seventh day of measurement, the participant would not receive any further tests or notifications via the app, and the participation ended. Overall, besides the request to fill in as many tests as possible and to do so as quickly as possible after they were made available, the participants were not restricted in the context in which they would fill in the affect grids.

Data Analysis

In alignment with past research on and with the affect grid, the scores for the two dimensions arousal and valence were analyzed separately (Russell et al., 1989; Russell & Gobet, 2012).

For the final analyses, of the participants who did not experience technical issues, (N=45), 26 participants were excluded by reason of too little data points. The remaining 19 participants had response rates of at least 65% (28 responses), which is in rough alignment with Palmier-Claus et al. (2011), who recommended a minimum of 20 valid responses. To improve the representativeness of the mean values, the author decided for a slightly higher number.

Before the analyses, a normal distribution test was performed. Graphical analyses by means of Q-Q-plots and dot-histograms for both dimensions, arousal and valence, indicated normally distributed sample scores for both dimensions (see Figures 7-9 in Appendix).

Split-half-reliability. To examine the reliability of the affect grid, split-half reliability of the (1) within-participant mean scores, (2) standard deviations, and (3) mean successive differences was calculated by means of a correlational analysis. The obtained data points were split into half at the 21st administered affect grid, irrespective of the number of responses a participant gave. In other words, the datapoints were split with regard to the responses obtained in the first 3,5 days and the second 3,5 days, and not based on the actual number of obtained responses, as this would disregard an equal time frame of the responses given.

Based on previous research in experience sampling, the author rated split half reliability correlations of .50 to .70 as moderate to high and correlations above .70 to .90 as high to very high (Csikszentmihalyi & Larson, 2014).

Validity. With regard to the relation of alexithymia and core affect, a correlational analysis with the mean score on the TAS-20 and core affect variability was performed. In line

with recommendations by Kuppens et al. (2010), core affect variability was represented by the within-participant standard deviation of the scorings per dimensions.

Furthermore, to investigate the relation of attractor strength and alexithymia, the author conducted a correlational analysis with the scores on the TAS-20 and the mean squared successive difference scores (MSSD) per participant.

Results

Descriptive Statistics

The scores of the sample on the two dimensions valence and arousal are graphically illustrated in Figure 3; In view of individual fluctuations in core affect and the changing influence of a naturalistic setting, the outliers con be assumed to not indicate abnormal ratings. Overall, despite similar standard deviations, the sample scored significantly higher in valence (*M*=33,28; *SD*=40,42) than in arousal (*M*=13,91; *SD*=42,63). The within-participant-standard deviation, again with similar standard deviations, was higher for arousal (*M*=37,69; *SD* =10,62) than for valence (*M*=30,65; *SD*=11,11).

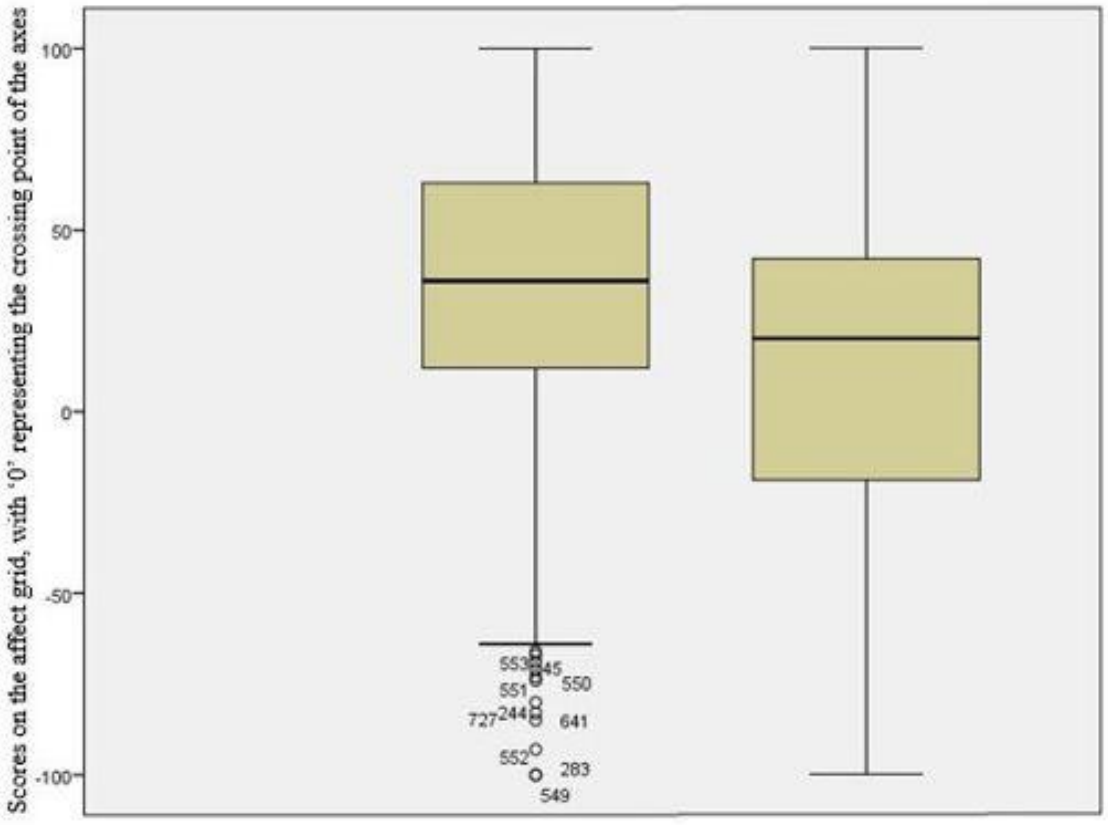

Figure 3. Boxplot of participants' scores on valence (left) and -arousal (right).

Figure 4 and 5 show boxplots illustrating the per-participant scoring on both dimensions. For a comprehensive overview of the participants' scores on the subscales and the affect grids see Table 2 and 3 in the Appendix.

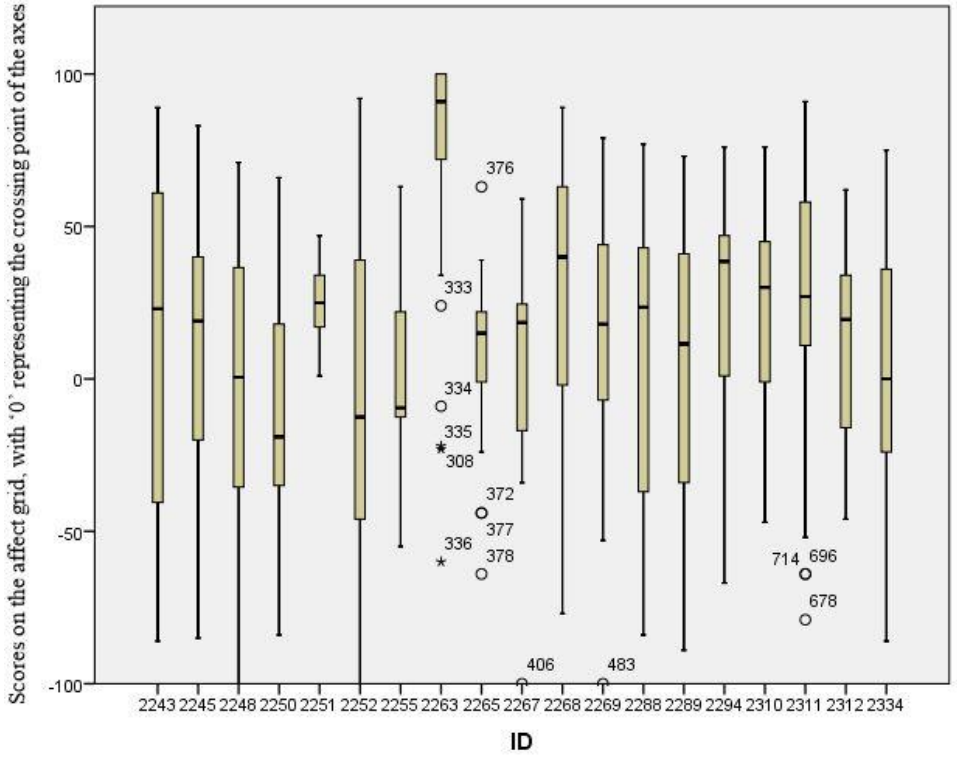

Figure 4. Boxplot per participant of the scores on arousal.

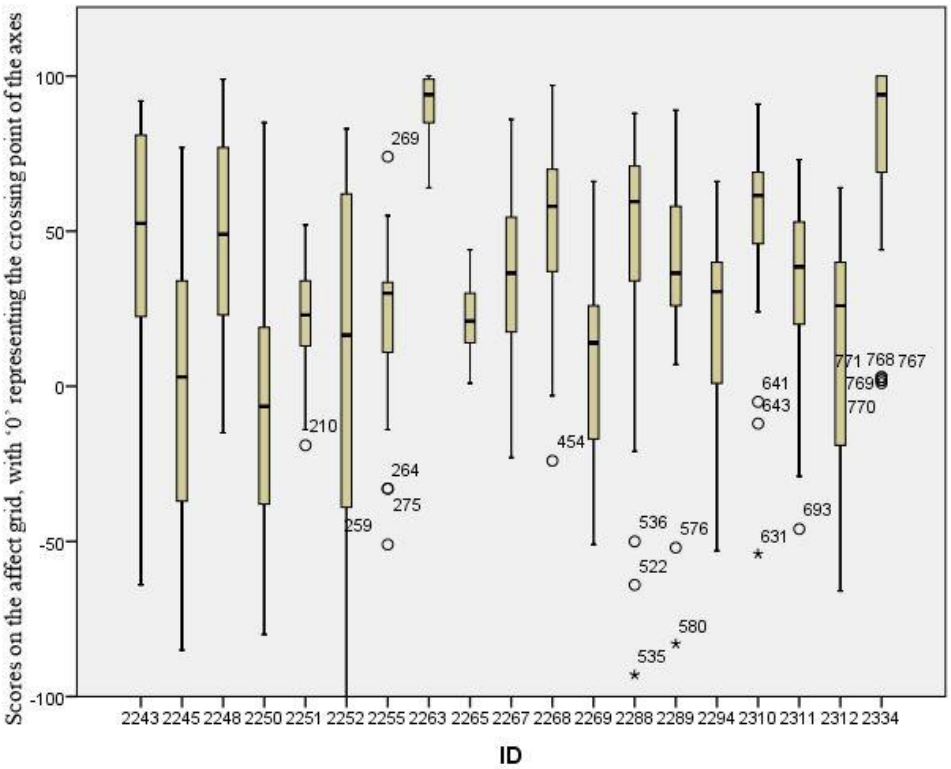

Figure 5. Boxplot per participant of the scores on valence.

Reliability

Overall, all tested split-half correlations were significant. The within-participant mean scores of the first and second half were highly correlated for both valence (r=.803, N=19; p<0.01) and arousal (r=.730, N=19; p<0.01) respectively. Furthermore, the within-participant standard deviations of the halves were moderately correlated in case of valence with a value of .658, (p<0.01), and highly correlated for arousal with .746, (p < .01). Finally, the analysis of the mean successive difference scores of the responses on the first and second 21 administered affect grids revealed a moderate correlation for valence (r=.672, N=19; p<0.01) as well as in case for arousal (r=.686, N=19; p<0.01).

Construct Validity

The results of the correlational analyses of the TAS-20 (sub-)scales with the within-participant standard deviations and mean squared successive differences of arousal and variance and is shown in Table 1. No significant or high negative correlation was found between within-participant standard deviation for valence and the scores on the TAS-20 in any case. However, a moderate negative correlation was observed between within-participant standard deviation of arousal and the subscale difficulties in describing feelings (DDF) (r=-.476, N=19; p<0.05). The same pattern emerged for the within-participant MSSD of valence, that exhibited low and insignificant correlations with all tested scales of the TAS-20. Furthermore, the MSSD of arousal again showed a negative, moderate correlation with the subscale DDF (r=-.475, N=19; p<0.05).

Table 1

Means with standard deviations for valence and arousal and average within-participant standard deviation for the scorings of each dimension.

N	Mean Arousal (SD)	Mean Valence (SD)	Average Scoring-SD Arousal (SD)	Average Scoring-SD Valence (SD)	Average MSD Arousal (SD)	Average MSD Valence (SD)
19	13,91 (42,64)	33,28 (40,42)	37,69 (10,62)	30,64 (11,11)	46,74 (17,16)	36,38 (15,00)

Note. In the present table the mean successive difference scores (MSD), which is the square root of the mean squared successive difference (MSSD), is included to allow a better comparison of the values, as the MSD has the same measurement unit as the coordinates of the affect grid.

Discussion

The purpose of the present study was to validate the affect grid in the context of experience sampling, more precisely with regard to split-half reliability and construct validity. On the question of reliability, the author found moderate to high spit-half reliability in all tested facets, with values in alignment with split-half reliability of other measures in experience sampling (Csikszentmihalyi & Larson, 2014). More precisely, the participants' individual mean score, standard deviation, and mean squared successive difference in core affect all exhibited adequate stability throughout the testing period. With respect to construct validity, the findings demonstrated a negative correlation of variability and pace of change in arousal with the TAS-20 subscale difficulties in describing feelings (DDF). This is in correspondence with the negative relationship of alexithymia and emotional variability that has been hypothesized in the present study and indicated in the past (e.g. Thompson et al. (2009)). However, contrary to the author's expectations, this study did not find a significant negative correlation of variability or pace of change in valence with scores on the TAS-20, independent of subscale. Likely explanations for and implications of these findings for experience sampling and the construct alexithymia will be discussed subsequently.

With values for split-half reliability ranging between .65 and .80, the affect grid exhibits good reliability compared to other measures in experience sampling, typically obtaining correlation coefficients of .55 to .93 (Csikszentmihalyi & Larson, 2014). In view of the increasing amount of experience sampling research involving the affect grid and the ongoing interest in affective states, this finding is of considerable significance (Ebner-Priemer et al., 2009; Kuppens et al., 2010; Russell & Gobet, 2012; Schimmack, 2003). Not only provide these findings a valid ground for conclusions made in past experience sampling studies involving the affect gird, they also open the door for future research on core affect involving the ESM. With increasing interest in implications of emotional experience on psychopathology (Cranford et al. (2006); Csikszentmihalyi (1992); Delespaul (1995); Ebner-Priemer et al. (2009); Kuppens et al. (2010); Palmier-Claus et al. (2011)) the present study found the affect grid being feasible to study everyday long-term experience in core affect. This gives rise to the opportunity to examine and validate the real-time emotional peculiarities of psychological disorders that so far have been reported in retrospective questionnaires.

Furthermore, in consequence of the adequate reliability found, the present study also provides support for the long-term stability of the three dimensions incorporated in the DynAffect. In the present study the affective home base was represented in the within-participant mean score, core affect variability in the within-participant standard deviation, and the attractor strength in the within-participant MSSD. As aforementioned, for valence and arousal all tested parameters turned out to be considerably stable throughout the testing period. Conclusively, these findings support the belief of relative stability of the DynAffect dimensions.

Finally, the reported negative relation between variability in arousal and alexithymia is consistent with the conceptualization of alexithymia, which includes having difficulties in distinguishing feelings (Bagby et al., 1994; Boden & Berenbaum, 2011). Furthermore, the impaired ability in recognizing (new) emotions, as reported by Lane et al. (2000); Lane et al. (1996) is reflected in the negative correlation of alexithymia with the pace of change in reported arousal. On the one hand, these findings are in line with the hypotheses and provide support for the construct validity of the affect grid in the present study. On the other hand, the expected relationship between alexithymia and core affect was only found with respect to arousal, not to valence. However, considering that valence and arousal are distinct dimensions, this is not necessarily an indicator for a lack in construct validity (Russell, 2003).

One possible explanation is that arousal rather than valence is connected to emotional variability in alexithymia. Indeed, neurological research indicates a distinct role of valence and arousal in alexithymia (Heinzel et al., 2010). Furthermore, in support of the present results, alexithymic individuals were found to exhibit hyperarousal in resting states, which implies a shorter distance to a state of high arousal compared to individuals who are low in arousal in the first place (Luminet, Rime, Bagby, & Taylor, 2004; Stone & Nielson, 2001). Accordingly, individuals in a constant state of hyperarousal experience less variability in arousal compared to those who are not. Moreover, in line with the present finding of a weaker attractor strength in arousal in relation to alexithymia, in a recent study Panayiotou and Constantinou (2017) reported alexithymic individuals to exhibit slower habituative responses to arousing stimuli. In other words, a slower adjustment of the individuals' arousal to external stimuli was found. Finally, in line with the present finding of an insignificant relation of valence and alexithymia, an extensive study by Peasley-Miklus, Panayiotou, and Vrana (2016) that investigated the role of valence and arousal in alexithymia reported physiological measures of arousal but not valence to be related to outcomes in the TAS-20. This speaks in

favor of the representativeness of the present findings. Taken together, these studies provide support for the reported results and the construct validity of the affect grid.

Limitations

With regard to limitations of the present study, there are two factors that could have had an influence on the data. The first concerns the sequence of the affect grid's administration; the study was programmed to administer the affect grid every two hours, and, if not filled in immediately, to remain available for 90 minutes. However, few participants reported that the previous affect grid remained available if it had not been filled in, even after the subsequent measurement was appearing. To the author's knowledge this was the case only for a few participants, however, it appears likely that retrospective answering of multiple responses could have reduced the variability in scoring, as the responses were given at the same time.

Another confounding factor that could have affected the scorings are external events. Of course, this is the case for any experience sampling research, however, in view of the rather 'short' sampling period of one week, an external event could have affected the participant over a considerable part of the testing period. In retrospection the author would therefore recommend including a question item such as 'During the testing period, did an event take place that strongly affected you in your mood in any way? If yes, 1. On which day? 2. Around which time? 3. How long did you feel affected by it? 4. How did the event make you feel? (more pleasant; less pleasant; more aroused; less aroused). Of course, these items cannot be thought to reliably explain detected irregularities in the responses, however, they could allow some insight into reasons for potential irregularities in the data.

On a more general level, also with regard to the conclusions about validity and reliability, caution is advised. Due to the little research conducted on the role of valence and arousal in alexithymia the question of construct validity of the affect grid based on the reported results can be answered on a preliminary basis only. Future correlational ESM-research involving constructs with better validated relations to core affect (e.g. the PANAS by Watson, Clark, and Tellegen (1988)) would allow further conclusions with regard to the affect grid's construct validity.

Implications and suggestions for future research.

The findings of the present study have several implications for research and practice. With regard to implications for future research, the notion of a different relation of arousal and valence to alexithymia seems to be supported in the present study. However, this difference

has received little attention from varying fields of research and lacks a comprehensive picture. A meta-study gathering the findings with respect to a distinct role of valence and arousal to alexithymia could be informative. Alternatively, based on the outcome of the present study the affect grid seems to be a suitable instrument to investigate this distinction further and cross-validate the outcome, e.g. by simultaneous use of a skin conductance measure to investigate arousal.

With respect to practical implications, this study could prove useful especially to the clinical practice. As Derks, Westerhof, Bohlmeijer, (2017) pointed out, one focus of psychotherapeutic treatments is to increase the emotional awareness of patients. The affect grid as long-term measure could contribute to the observation of treatment effects on emotional awareness by investigating enduring changes in core effect variability. Also, concluded in the same study, alexithymia and, closely connected, emotional awareness have been found to be related to borderline personality pathology (BPP). However, these findings have been based on mono-self reports (Derks et al., 2017). With the affect grid a tool is available that allows to investigate these past findings by means of experience sampling. In this regard, as a rather ventured suggestion, it furthermore seems possible that using the affect grid over a longer period of time could by itself contribute to an increase in emotional awareness, as the patient is habituated to attend to the own emotions more frequently. Effects of a decrease in scoring variance over time have been reported by Csikszentmihalyi and Larson, (2014) but could be due to a multitude of reasons. On these grounds, by investigating changes in emotional variability when using the affect grid over a long period of time, in combination with a measure of emotional awareness, not only treatment effects could be observed, but also changes in variability not stemming from the actual treatment.

Conclusion

Taken together, the present study found the affect grid to exhibit adequate split-half reliability in all tested facets. With respect to the question of construct validity, correlations in the expected directions were found, however, only in one of the two facets, namely arousal. In light of research that provides evidence for a connection of alexithymia to arousal, but not valence, these findings speak in favor of adequate construct validity of the affect grid in the present study. Nevertheless, additional validation appears necessary to confirm this conclusion.

References

Bagby, R. M., Parker, J. D. A., & 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. doi:https://doi.org/10.1016/0022-3999(94)90005-1

Boden, M. T., & Berenbaum, H. (2011). What you are feeling and why: Two distinct types of emotional clarity. *Personality and Individual Differences, 51*(5), 652-656. doi:https://doi.org/10.1016/j.paid.2011.06.009

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., & Garcia-Crespo, A. (2011). Using the affect grid to measure emotions in software requirements engineering. *Journal of Universal Computer Science, 17*(9), 1281-1298. doi:10.3217/jucs-017-09-1281

Conner, T., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89-107). New York: New York: Guilford Press.

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology, 32*(7), 917-929. doi:10.1177/0146167206287721

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/BF02310555

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Csikszentmihalyi, M. (1992). *The experience of psychopathology: Investigating mental disorders in their natural settings* (M. W. d. Vries Ed.). Cambridge: Cambridge University Press.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The collected works of Mihaly Csikszentmihalyi* (pp. 35-54). Dordrecht: Springer Netherlands.

Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life : the experience sampling method.* (Doctoral Thesis), Maastricht: University Press, 1995.

Derks, Y. P. M. J., Westerhof, G. J., & Bohlmeijer, E. T. (2017). A meta-analysis on the association between emotional awareness and borderline personality pathology. *Journal of personality disorders*, *30*(3), 362-384. https://doi.org/10.1521/pedi_2016_30_257

Dooley, D. (2009). *Social Research Methods* (4th ed.). New Jersey: Pearson Education Limited.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Dover.

Ebner-Priemer, U., Houben, M., Santangelo, P., Kleindienst, N., Tuerlinckx, F., Oravecz, Z., . . . Kuppens, P. (2015). Unraveling affective dysregulation in borderline personality disorder: A theoretical model and empirical evidence. *J Abnorm Psychol, 124*. doi:10.1037/abn0000021

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology, 118*(1), 195-202. doi:10.1037/a0014868

Field, A. (2014). *Discovering statistics using IBM SPSS statistics* (4 ed.). London: SAGE Publications Ltd.

Heinzel, A., Schäfer, R., Müller, H.-W., Schieffer, A., Ingenhag, A., Northoff, G., . . . Hautzel, H. (2010). Differential modulation of valence and arousal in high-alexithymic and low-alexithymic individuals. *NeuroReport, 21*, 998-1002. doi:10.1097/WNR.0b013e32833f38e0

Hoey, M. (2012). Lexical priming. In *A New Theory of Words and Language*. London: Routledge.

Killgore, W. D. S. (1998). The affect grid: A moderately valid, nonspecific measure of pleasure and arousal. *Psychological Reports, 83*(2), 639-642. doi:10.2466/pr0.1998.83.2.639

Kooiman, C., Spinhoven, P., & W Trijsburg, R. (2003). The assessment of alexithymia: a critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research, 53*, 1083-1090.

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*, 1042-1060. doi:10.1037/a0020962

Lane, R., Sechrest, L., Riedel, R., Shapiro, D., & Kaszniak, A. (2000). Pervasive emotion recognition deficit common to alexithymia and the repressive coping style. *Psychosomatic Medicine, 62*, 492-501. doi:10.1097/00006842-200007000-00007

Lane, R. D., Lee, S., Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic Medicine*, *58*(3), 203-210.

Leising, D., Grande, T., & Faber, R. (2009). The Toronto Alexithymia Scale (TAS-20): A measure of general psychological distress. *Journal of Research in Personality, 43*(4), 707-710. doi:https://doi.org/10.1016/j.jrp.2009.03.009

Luminet, O., Rime, B., Bagby, R. M., & Taylor, G. J. (2004). A multimodal investigation of emotional responding in alexithymia. *Cognition & Emotion, 18*(6), 741-766. doi:10.1080/02699930341000275

Palmier-Claus, J., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P., . . . Dunn, G. (2011). Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta psychiatrica Scandinavica, 123*, 12-20. doi:10.1111/j.1600-0447.2010.01596.x

Panayiotou, G., & Constantinou, E. (2017). Emotion dysregulation in alexithymia: Startle reactivity to fearful affective imagery and its relation to heart rate variability. *Psychophysiology*, *54*(9), 1323-1334. doi:10.1111/psyp.12887

Parker, J., Taylor, G., & Bagby, R. (2003). The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research*, *55*, 269-275. doi:10.1016/S0022-3999(02)00578-0

Parker, J. D. A., Michael Bagby, R., Taylor, G. J., Endler, N. S., & Schmitz, P. (1993). Factorial validity of the 20-item toronto alexithymia scale. *European Journal of Personality*, *7*(4), 221-232. doi:10.1002/per.2410070403

Peasley-Miklus, C. E., Panayiotou, G., & Vrana, S. R. (2016). Alexithymia predicts arousal-based processing deficits and discordance between emotion response systems during emotional imagery. *Emotion, 16*(2), 164-174. doi:10.1037/emo0000086

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172. doi:10.1037//0033-295X.110.1.145

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*.

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. doi:10.2466/03.28.PMS.114.1.125-136

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies, 4*(1), 79-106. doi:10.1023/a:1023661322862

Soo Seo, S., Chung, U.-S., Deog Rim, H., & Jeong, S. (2009). Reliability and validity of the 20-Item Toronto Alexithymia Scale in korean adolescents. *Psychiatry investigation, 6*, 173-179. doi:10.4306/pi.2009.6.3.173

Stone, L. A., & Nielson, K. A. (2001). Intact physiological response to arousal with impaired emotional recognition in alexithymia. *Psychotherapy and Psychosomatics, 70*(2), 92-102. doi:10.1159/000056232

Taylor, G. J., Ryan, D., & Bagby, R. M. (1985). Toward the development of a new self-report alexithymia scale. *Psychotherapy and Psychosomatics, 44*(4), 191-199. doi:10.1159/000287912

Thompson, R. J., Dizén, M., & Berenbaum, H. (2009). The unique relations between emotional awareness and facets of affective instability. *Journal of Research in Personality, 43*(5), 875-879. doi:https://doi.org/10.1016/j.jrp.2009.07.006

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual review of clinical psychology, 9*, 151-176. doi:10.1146/annurev-clinpsy-050212-185510

Versluis, A., Verkuil, B., Lane, R. D., Hagemann, D., Thayer, J. F., & Brosschot, J. F. (2018). Ecological momentary assessment of emotional awareness: Preliminary evaluation of psychometric properties. *Current Psychology*. doi:10.1007/s12144-018-0074-6

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

Appendix

A. Instructions sent to the participants by mail after being assigned to the study

Hey [account name]!

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,
[researchers]
PS: In case of any open questions, issues, or critique please do not hesitate to contact

[Names and corresponding mail addresses of the researchers]

B. TAS-20 with the two subscales administered

I. Difficulties Identifying Feelings

1. I am often confused about what emotion 1 am feeling.
2. I have physical sensations that even doctors don't understand.
3. When I am upset I don't know if I am sad, frightened, or angry.
4. I am often puzzled by sensations in my body.
5. I have feelings that 1 can't quite identify.
6. I don't know what's going on inside me.
7. I often don't know why I am angry.

II. Difficulties Describing Feelings

1. It is difficult for me to find the right words for my feelings.
2. I am able to describe my feelings easily.
3. I find it hard to describe how I feel about people.
4. People tell me to describe my feelings more.
5. It is difficult for me to reveal my innermost feelings, even to close friends.

C. Additional graphics and Tables

Table 2

Gender, age, nationality, number of given responses, mean scores and standard deviations of the participants' scorings on arousal and valence.

ID	DIF	DDF	Total Score	Mean Score Arousal	Arousal SD	Mean Score Valence	Valence SD	MSD Arousal	MSD Valence
2243	27	6	33	12,64	54,62	44,14	44,74	81,05	62,44
2245	16	15	31	7,69	39,96	1,69	38,75	49,83	48,98
2248	22	16	38	6,35	43,21	53,65	28,74	56,76	30,87
2250	28	20	48	-8,33	31,48	-4,57	38,62	41,97	48,95
2251	24	23	47	26,07	10,08	24,43	15,43	10,81	12,84
2252	18	18	36	-8,08	49,68	8,03	54,68	57,83	60,20
2255	20	14	34	5,39	32,03	19,69	27,84	34,69	37,32
2263	17	11	28	76,95	39,89	92,93	8,54	29,49	8,52
2265	13	18	31	10,15	23,72	21,28	10,98	27,42	12,72
2267	16	17	33	5,25	32,10	34,54	30,50	49,65	37,40
2268	21	17	38	22,97	48,04	50,79	30,14	60,86	40,53
2269	20	14	34	15,61	41,62	14,58	31,75	62,75	42,54
2288	16	10	26	6,08	47,25	45,03	42,91	68,97	54,60
2289	23	21	44	3,79	44,54	36,64	32,06	47,49	38,47
2294	26	23	49	22,79	37,43	22,71	27,52	48,15	33,34
2310	17	16	33	23,02	29,58	54,10	26,75	29,97	27,37
2311	13	18	31	26,93	42,15	33,76	27,30	54,32	31,71
2312	19	14	33	13,02	26,20	12,64	32,77	25,55	32,84
2334	11	7	18	3,98	42,48	77,46	32,30	50,41	29,65

Note. Gender (1=female, 2=male), nationality (1= German, 2= other)

Table 3

Means, standard deviations (SD), and mean successive difference scores (MSD) with scores in the two used subscales of the TAS-20 and the respective total score.

ID	Gender	Age	Nationality	N(Scores)	Mean Score Arousal	Arousal Scoring-SD	Mean Score Valence	Valence Scoring-SD
2243	1	21	1	42	12,64	54,62	44,14	44,74
2245	1	21	1	39	7,69	39,96	1,69	38,75
2248	1	21	1	40	6,35	43,21	53,65	28,74
2250	2	21	1	42	-8,33	31,48	-4,57	38,62
2251	1	21	1	42	26,07	10,08	24,43	15,43
2252	1	21	1	36	-8,08	49,68	8,03	54,68
2255	2	24	1	36	5,39	32,03	19,69	27,84
2263	1	22	1	41	76,95	39,89	92,93	8,54
2265	1	19	1	40	10,15	23,72	21,28	10,98
2267	2	22	1	28	5,25	32,10	34,54	30,50
2268	1	23	1	29	22,97	48,04	50,79	30,14
2269	2	22	1	36	15,61	41,62	14,58	31,75
2288	2	21	1	38	6,08	47,25	45,03	42,91
2289	2	20	1	42	3,79	44,54	36,64	32,06
2294	1	21	1	42	22,79	37,43	22,71	27,52
2310	2	24	1	42	23,02	29,58	54,10	26,75
2311	2	23	2	42	26,93	42,15	33,76	27,30
2312	1	19	1	42	13,02	26,20	12,64	32,77
2334	1	24	1	41	3,98	42,48	77,46	32,30

Note. In the present table the mean successive difference scores (MSD), which is the square root of the mean squared successive difference (MSSD), is included to allow a better comparison of the values, as the MSD has the same measurement unit as the coordinates of the affect grid.

Dot-Plot with Normal Curve of the Samples' Scores on the Dimensions Valence and Arousal

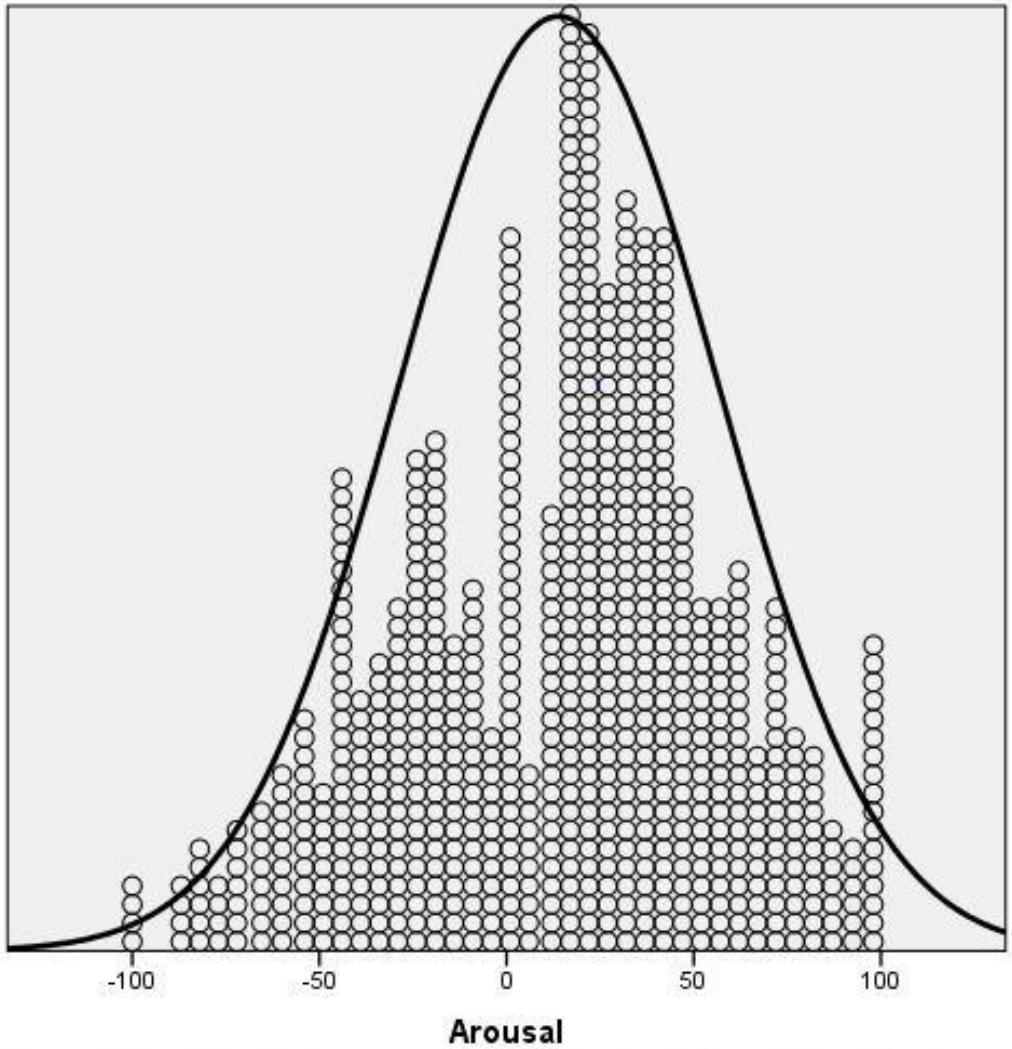

Figure 6. Normal distribution curve with simple dot plot of the participants' scorings on arousal.

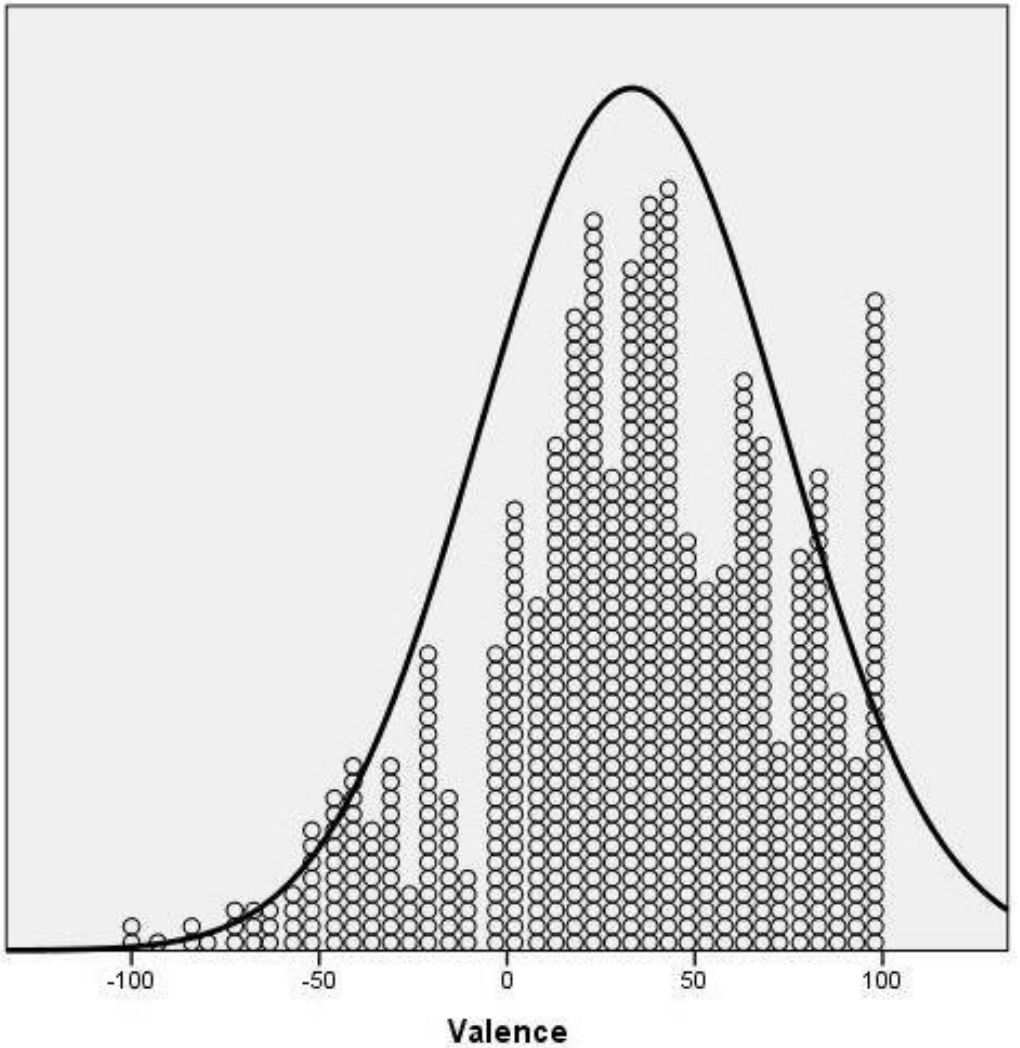

Figure 7. Normal distribution curve with simple dot plot of the participants' scorings on valence.

Normal Q-Q Plots for Samples' Scores on the Dimensions Valence and Arousal

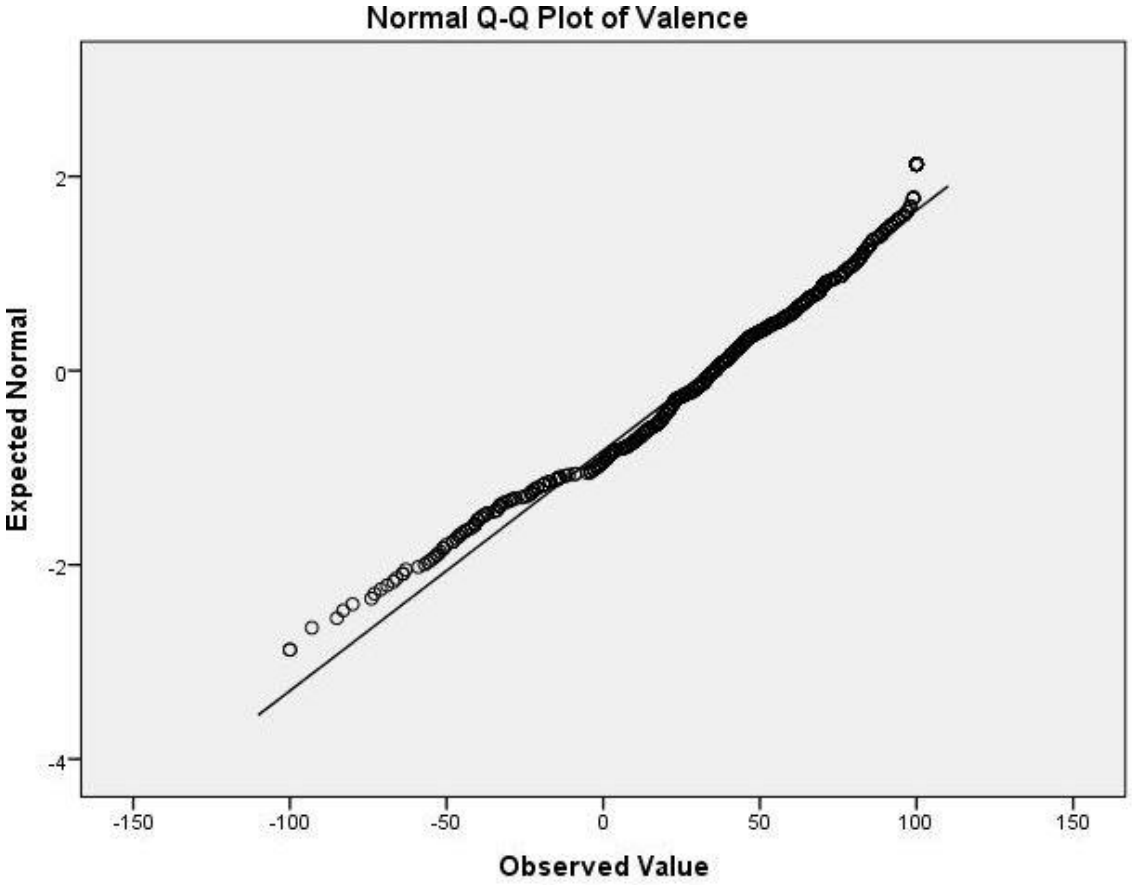

Figure 8. Normal Q-Q plot of the valence-scorings.

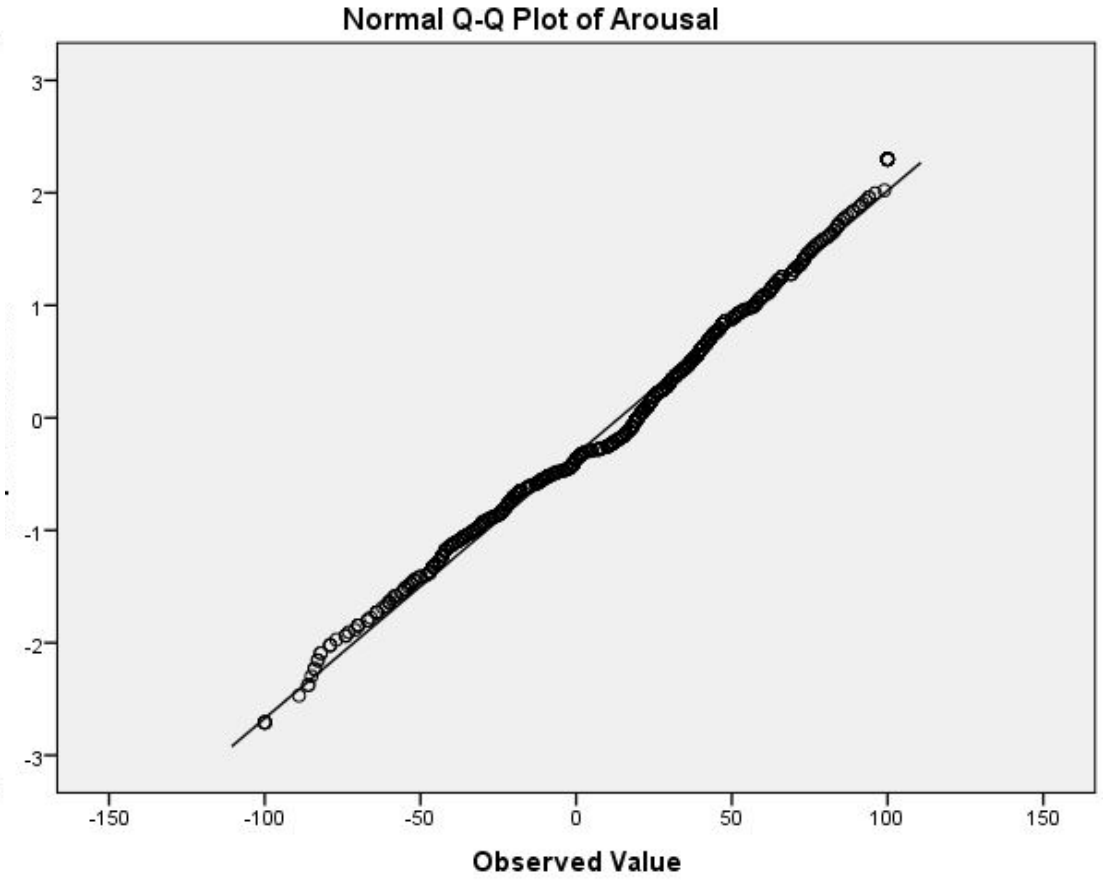

Figure 9. Normal Q-Q plot of the arousal-scorings.

Participant means of Valence and Arousal Scorings

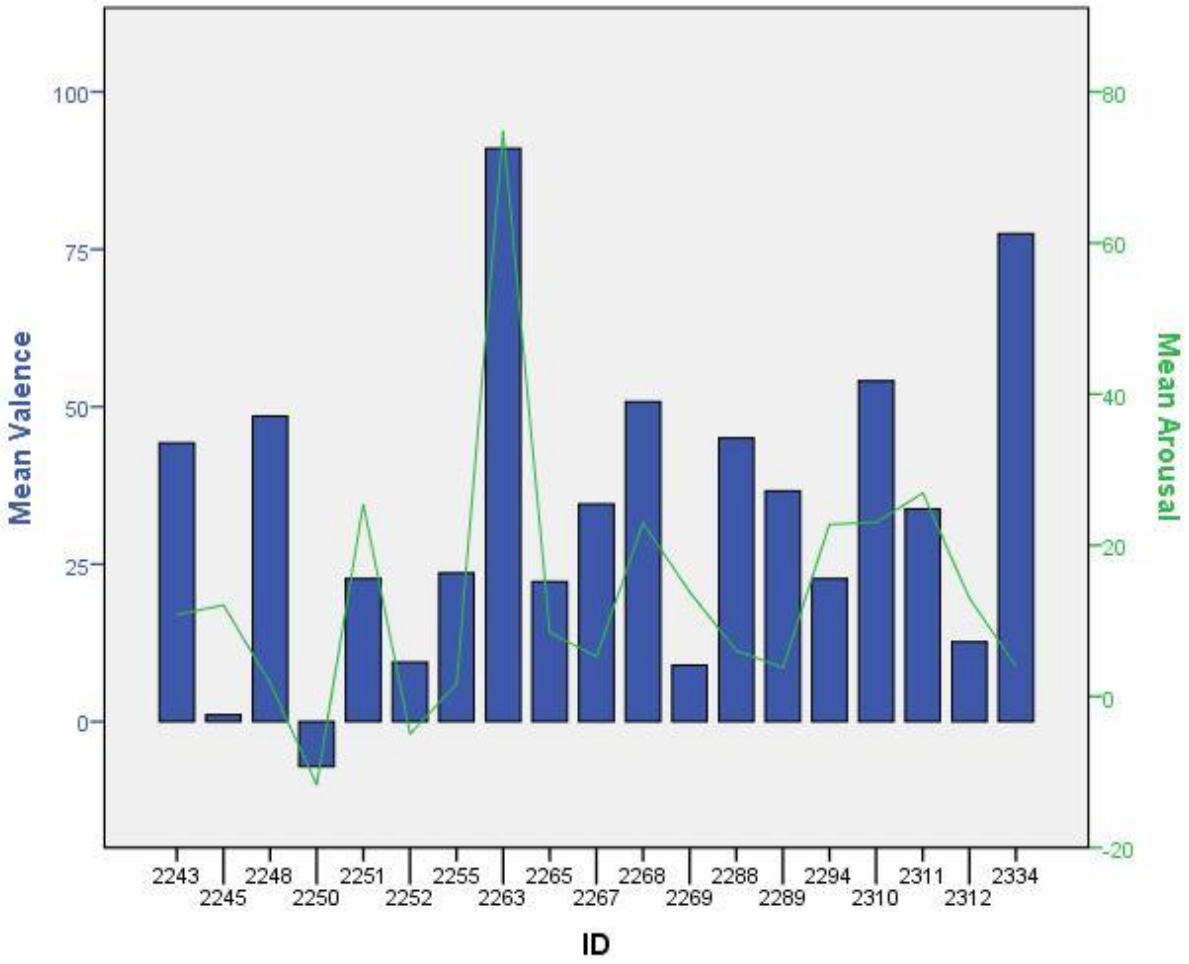

Figure 10. Relation of mean-valence and mean-arousal.

ACKNOWLEDGEMENTS

The author thanks Lucia Rabago Mayer for her support in creating the intervention in TiiM.

The author thanks Juilan Noormalal for his support in handling the obtained data.