

**The Effects of VAS vs. Likert Scales on Emotion Dynamics of Negative Affect in
Experience Sampling Methods**

Lea Staudigel

Faculty of Behavioural, Management, and Social Sciences

Department of Psychology, University of Twente

202000384: BSc Thesis Psychology

First Supervisor: Dr. Jannis T. Kraiss

Second Supervisor: Dr. Thomas R. Vaessen

June 27, 2023

Abstract

Background: The Experience Sampling Method (ESM) captures real-time emotional states and is often used to investigate emotion dynamics, encompassing variability, instability, and inertia. Commonly used response formats such as Likert scales and Visual Analogue Scales (VAS) have been associated with endpoint aversion or the ceiling effect. The influence of these response formats on emotion dynamics remains largely unknown. This study aims to examine the effects of Likert scales and VAS on emotional variability, instability, and inertia of negative affect (NA) within ESM questionnaires.

Method: A randomized controlled trial was conducted using a between-subject design. Participants were recruited using convenience and volunteer sampling, resulting in a final sample size of 89 individuals (63.49% female, 36.51% male, mean age = 29.94; Likert $n = 53$, VAS $n = 36$). Each participant was asked to complete ten semi-randomly scheduled questionnaires per day over the course of one week. Fixed-effect and linear mixed models were utilized to evaluate the effect of response format on emotional variability, instability, and inertia.

Results: Scores were similarly distributed between conditions. No statistically significant effects of response format were found on emotional variability ($b = .033, p = .588$), on emotional instability ($b = .002, p = .984$), and on emotional inertia ($b = -0.062, p = 0.077$).

Conclusion: This study concludes that the choice between Likert scale and VAS does not appear to have a significant effect on reported emotion dynamics. However, the generalizability of these results is limited due to the unrepresentative sample obtained and the low compliance rate.

The Effects of VAS vs. Likert Scales on Emotion Dynamics of Negative Effect in Experience Sampling Methods

Our daily psychological states are, as the name suggests, a constant companion in our life. Naturally, researchers are trying to understand how they change, when they change, and how these states and their expression differ in individuals. What this requires is a reliable way to test and record emotions in individuals without relying on retrospective methods, which are known to come with biases related to recall ability (Myin-Germeys & Kuppens, 2022) or a general inability to accurately assess emotional experiences (Larson & Csikszentmihalyi, 2014). That is where the Experience Sampling Method (ESM) comes into play, a research approach that has been gaining exponential popularity in the past decade (Myin-Germeys & Kuppens, 2022), but still lacks investigation into how certain design choices could affect resulting data. ESM is characterized by asking individuals to give self-reports on their current emotional state or activities throughout the day on several occasions. Because ESM data is collected in real-time and administered in the participant's natural environment, it is considered to have high ecological validity (Myin-Germeys & Kuppens, 2022).

Response Scales: Likert Scale and VAS

In ESM questionnaires, Visual Analogue Scales (VAS) and Likert scales are commonly used as response formats (Eisele et al., 2022). On a VAS, the respondent must indicate their position between the lower and upper points of the scale, often from 0 to 100, using a slider. When using a Likert scale, on the other hand, answer options often range from one to five or from one to seven, and participants must choose between these given response options. Except for the fact that continuous scales have a natural advantage in measuring continuous variables as they potentially lead to more differentiated responses (Eisele et al., 2022), research regarding the possible effect of response scales on data in ESM studies is missing.

In the context of retrospective data-gathering techniques, some studies have found equal reliability and validity as well as similar means and SDs for both Likert scale and VAS (Gries et al., 2018; Kuhlmann et al., 2017; Lukacz et al., 2004).

Other studies did mention potential effects of the scales on response styles. Studer (2012) suggests that using a VAS reduces the endpoint aversion that characterizes the response style with Likert scales. In other words, when using a Likert scale, participants tend to score more in the centre of the scale compared to when using a VAS. This implies that the SD, when using a Likert scale, is smaller compared to the SD when using a VAS. That is

because the scores are more distributed on the entirety of the scale when using a VAS, rather than only towards the centre when using a Likert scale. Reduced endpoint aversion when using a VAS also suggests that responses are more likely to be extreme, indicating an extreme response style (ERS).

Furthermore, because a VAS has been shown to be more sensitive, especially towards the endpoints, it might lead scores to *not* cluster at the endpoints of the scale (Hayes, 2013), meaning more varied scores that have a higher SD, while Likert scale responses would instead cluster at the centre. In contradiction to this stands the fact that Likert scales seem to be more susceptible to the ceiling effect than a VAS (Voutilainen et al., 2016). However, the Likert scale being vulnerable to the ceiling effect would also suggest a lower SD for Likert scales, because participants' scores cluster around the upper end of the response scale. While it therefore still seems debatable whether Likert scale responses are characterized by endpoint aversion or the ceiling effect, it makes no difference for the effect on the SD. The endpoint aversion and the ceiling effect would lead to clustered scores that suggest a lower SD compared to responses when using VAS, as this format has been shown to be less susceptible to both patterns.

Currently, ESM researchers have no information if the choice of response scale affects the study results, ultimately leaving that choice up to personal preference. This could not only create problems with comparability, but the possibility of response formats leading to different response styles could also seriously threaten the reliability and construct validity of ESM studies.

Emotion Dynamics

ESM is often used to study emotion dynamics, which is defined as the study of emotional change that aims to understand the processes behind it as well as what it means for an individual's well-being (Kuppens, 2015). Emotion dynamics are assessed using measures of positive affect (PA), e.g., how interested, proud, and happy one feels, and negative affect (NA), e.g., how distressed, hostile, and ashamed one feels. Emotion dynamics are often divided into three separate concepts: emotional variability, emotional stability, and emotional inertia. Emotional variability is considered as the intensity of the emotional fluctuations over time, someone high on emotional variability experiences more extreme levels of emotions compared to their average emotional level (Houben et al., 2015), meaning more extreme emotional responses (Bernstein et al., 2019). Emotional variability is usually calculated with the within-person standard deviation (SD) (Houben et al., 2015). Emotional instability, sometimes called emotional stability, refers to the degree of fluctuations of someone's

emotions over time, someone low on emotional instability has more consistent and stable emotions in response to an event, whereas someone high on emotional instability usually lives through a lot of emotional fluctuations, regardless of their strength (Houben et al., 2015). Emotional instability is generally calculated as the mean squared successive difference (MSSD) between consecutive emotion scores (Houben et al., 2015). Emotional inertia indicates how far emotional states are resistant to change. For someone high on emotional inertia, an emotion triggered by an event may last longer, even after the event has passed (Houben et al., 2015). This also means that the level of emotional inertia expresses how a specific previous emotion is predictive of the same emotion in the present (Bernstein et al., 2019). Emotional inertia is typically measured with the autocorrelation of a variable over time, that is how well the current value of a variable can be predicted by its past values (Bernstein et al., 2019).

As explained above, the use of a Likert scale could cause a smaller SD compared to a VAS scale. In the context of ESM, this could affect the estimates of emotional variability. With an increased SD when using a VAS scale, the observed emotional variability would be higher than when using a Likert scale.

A more pronounced ERS, supposed to be more likely to occur when using a VAS, could influence the estimates of emotional variability and emotional instability, as well as emotional inertia. An ERS could lead to an increased SD and therefore also to an inflated emotional variability score. Furthermore, if a person consistently rates their emotional state on both extreme ends of a scale, the MSSD between consecutive emotion scores will be higher, indicating higher emotional instability. While these effects seem straightforward, an ERS might have differing effects on estimates of emotional inertia. In general, if a participant continuously alternates between scoring on the lower and upper ends of the scale, the autocorrelation of the variable will be underestimated, leading to lower emotional inertia. If the participant however only scores extremely on one end of the scale, the opposite might occur, leading to an inflated score of emotional inertia.

Finally, the ceiling effect, to which the Likert scale is more susceptible, could also have an effect on emotional variability, emotional instability, and emotional inertia. As a ceiling effect suggests a lower SD, it would also mean that emotional variability is lower when using a Likert scale. A VAS simply allows more varied scores because participants can respond more precisely. For emotional instability this means that a ceiling effect makes it more difficult to detect subtle changes in the emotions of the participants, which could lead to a deflation of the MSSD between consecutive emotion scores, indicating a lower emotional

instability. Lastly, a ceiling effect could be the consistent pattern that causes an inflation of the autocorrelation of the variable and therefore an inflation emotional inertia estimate, as participants cluster around the upper end of the scale.

The Current Study

Looking at previous research with retrospective methods, it is apparent that response formats do seem to affect response styles. If this is also the case within ESM, it could potentially lead to an over- or underestimation of emotion dynamics estimates. It is therefore important to understand what effect the choice of response formats has on outcomes obtained from ESM studies.

Seeing as there is no evidence on how response format affects emotional dynamics estimates within an ESM study, the current study aims to investigate this further. The overall research question of this study is: “*What is the effect of using a Likert scale vs. using VAS in ESM questionnaires on findings of emotional variability, emotional instability, and emotional inertia of NA?*”.

Based on the literature about the differing effects of a Likert scale versus a VAS scale in retrospective studies and the subsequent interpretations, the following three hypotheses are proposed:

H1: Emotional variability of NA will be higher when using a VAS compared to when using a Likert scale.

H2: Emotional instability of NA will be higher when using a VAS compared to when using a Likert scale.

H3: Emotional inertia of NA will be different when using a VAS compared to when using a Likert scale.

Methods

Design

In order to test the effect of using a Likert scale vs. using a VAS on measures of emotional variability, emotional instability, and emotional inertia, an ESM study was conducted using a between-subject design with two conditions (Likert scale vs. VAS). This study was approved by the Ethical Committee of the University of Twente (request number: 221244).

Participants

After conducting an *a priori* power analysis for a linear model using G*Power (Faul et al., 2009), setting the effect size to $f = .25$ and $\alpha = .05$, the analysis specified that a sample of 128 participants in total would be necessary to achieve a power of .8, meaning 64 participants per condition. However, Vachon et al. (2019) found a compliance rate of 94% in ESM studies, therefore, the desired sample size was decided to be 136 participants. Participants were recruited through convenience sampling, which was chosen due to it being cost- and time-efficient (Etikan et al., 2016; Stratton, 2021), and through volunteer sampling via the SONA system of the University of Twente. To incentivize participating in this study, students specifically were rewarded with 1.5 SONA credits, and all participants were part of a lottery to win a 40€ Amazon or Bol voucher provided by the researchers. Recruiting took place two weeks before the study started. The inclusion criteria specified all participants be 18 years old or older, own a smartphone, and have English proficiency.

Randomization

Participants were randomized to one of the conditions using stratified randomization by randomizing within each group of participants that were recruited by researcher. The sample was stratified by researcher to account for the similarities that participants might have from being recruited from a researcher's environment. Two other previous teams recruited participants and gathered data that can be used for the current study as well. The first cohort of participants from previous research groups was randomized 50/50. As a new condition was added in the second cohort, it was randomized 30/70 to ensure equal allocation. Because this did not fully work and the current cohort was still randomized 50/50, one condition has slightly more participants compared to the other.

Materials

For the current study, two types of questionnaires were used. The baseline questionnaire inquired about demographic data from the participants and included several measures meant to assess mental well-being. The ESM questionnaires consisted of questions regarding the current emotional state of the participants, their current company, the most recent striking event and how they dealt with it. For the current study, only the current emotional state of the participant was of relevance.

Negative affect was assessed based on the Positive and Negative Affect Schedule (PANAS), which has shown to be a reliable way to measure PA and NA (Watson et al., 1988), and has been widely used to assess affective states in ESM research (Janssens et al., 2020; Thompson et al., 2012; van der Linden et al., 2021). For the current study, only the

measure of NA was relevant. Participants had to answer the four following items: “*How anxious / irritable / down / guilty do you feel right now?*” Depending on the condition, participants answered using a 7-point Likert Scale or a VAS ranging from 0 to 100. Both scales were labelled at the endpoints only with *not at all* and *extremely*.

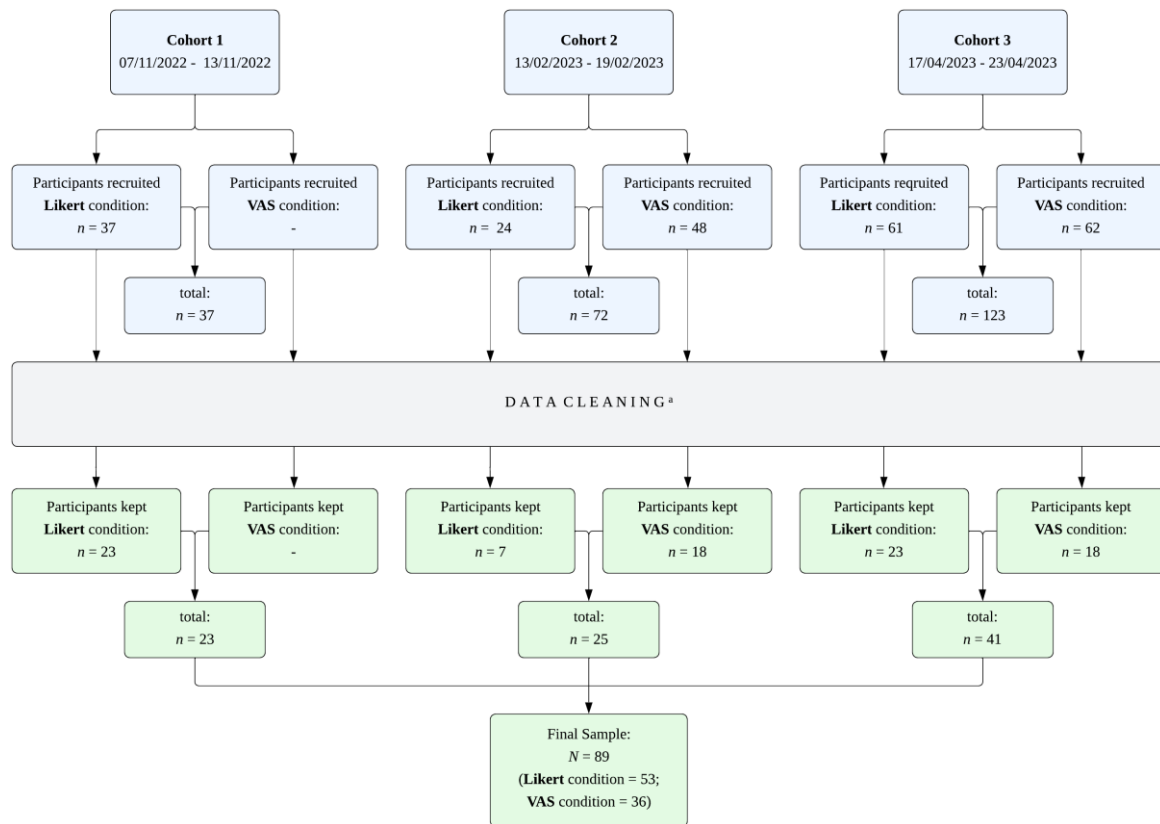
The composite NA score was used as an overall measure of NA by averaging the NA items for each questionnaire, which is a common mode for assessing affective states (Janssens et al., 2020; van der Linden et al., 2021). The Likert condition had a split-half reliability of $r = .893$ for the NA score. The VAS condition had a split-half reliability of $r = .895$ for the NA score.

Procedure

After recruitment, participants were sent an e-mail with information and instructions about the study and access information for the app Ethica (Ethica Data Services Inc., 2023; Appendix A & B). In Ethica, participants had to agree to the informed consent form (Appendix C). Data collection of the three cohorts took place between November 2022 and April 2023 (Figure 1). An ESM-typical period of seven days was chosen to get a complete insight into the participants' week and gather data on weekdays as well as the weekend (Dejonckheere & Erbas, 2022). The first questionnaire that participants received was the baseline questionnaire, which took around 20 minutes to fill out. Afterwards, regardless of whether the baseline questionnaire had been completed, the ESM questionnaires were sent out. Participants received 10 questionnaires a day for 7 days, which is a common assessment frequency for ESM studies (Dejonckheere & Erbas, 2022). These questionnaires were randomized within ten equally sized time intervals so that participants received one questionnaire at a random moment in time, within each of the intervals (Table 1). This semi-random schedule was used because it minimizes participants' reactivity to the questionnaires, as they are triggered unexpectedly (Myin-Germeys et al., 2018). Each of the ESM questionnaires took 1 to 2 minutes to answer and expired after 15 minutes.

Data Analysis

All analyses were performed using R Statistical Software (v4.3.0; R Core Team, 2023). Participants who completed less than 1/3 of the questionnaires were removed, meaning only those participants that answered at least 23.33 questionnaires were used for further analysis. This cut-off point was chosen as a low response rate suggests that participants only completed a questionnaire when it was convenient, making their answers unrepresentative (Viechtbauer, 2022b).

Figure 1*Participants per Cohort*

Note. ^a Removing participants that completed less than 23.33 questionnaires.

Table 1*Schedule of the Questionnaires*

Day	Questionnaire	Triggered	Expire
1	Baseline Questionnaire	Once at the beginning of the study	No
Every day (7 days)	ESM Questionnaire	7:30 – 9:00 9:00 – 10:30 12:00 – 13:30 13:30 – 15:00 15:00 – 16:30 16:30 – 18:00 18:00 – 19:30 19:30 – 21:00 21:00 – 22:30	Yes, after 15 minutes

Descriptive statistics were obtained, and the composite NA scores were standardized by subtracting the mean from each score and dividing it by the SD. Following this, the statistics for emotional variability (within-person standard deviation [SD] of NA scores) and emotional instability (mean squared successive difference [MSSD] between consecutive NA scores) were computed using the standardized NA scores. To later assess the autocorrelation of NA scores over time, a lagged NA variable ($t-1$) was created.

To test the first two hypotheses, two fixed effect models were run with condition as the independent variable and the within-person SD or the MSSD for consecutive NA scores as the dependent variable, respectively. Although ESM studies result in repeated measures data, the within-person SD and the MSSD are time-invariant, hence a fixed effect model was appropriate. To avoid unit-of-analysis errors, the data used to run the models only included a single row per participant in which their unique within-person SD and MSSD were stored.

To test the final hypothesis, a Linear Mixed Model (LMM) was run, using the package “nlme” (Pinheiro & Bates, 2000; Pinheiro et al., 2023). Repeated measures are usually organized on two levels: level 1 represents the total observations, while level 2 represents the number of participants. LMMs are the common mode of analysis for ESM data, as they consider the repeated observations of each participant and their dependency (Viechtbauer, 2022a).

For the LMM, the lagged NA variable and condition were set as the independent variable, including an interaction between the two. The NA standardized score was set as the dependent variable, with participants as a random effect, using a random intercept model. Moreover, the variance-covariance structure was specified as autoregressive (1), to model the regression of each observation to the subsequent one (Jongerling et al., 2015). An AR(1) structure assumes that observations closer together in time are more correlated than those further away in time (Kincaid, 2005). With these parameters, the extent of how the current NA state is predicted by its past state is assessed (Jongerling et al., 2015). For all analyses, a p-value of $<.05$ was considered significant.

Results

In total, 232 participants took part in the study. During data analysis, 143 participants were excluded because they completed less than 33.33% of the questionnaires. Therefore, the final sample consisted of 89 participants. Here, 53 participants took part in the Likert condition and 36 participants took part in the VAS condition. In total, 56.56% of participants

were removed from the Likert condition due to insufficient response frequency, while 67.27% of participants were removed from the VAS condition for the same reason. Figure 1 describes how many participants were obtained and removed in what cohort.

The sample from which demographic data was obtained consisted of 63 participants, with 40 females (63.49%) and 23 males (36.51%), aged between 20 and 67 years (Mean = 29.94, SD = 13.79). In Table 2, descriptive statistics of the sample and of the NA composite scores for the two separate conditions are presented.

The NA standardized composite scores for both conditions show a visually similar distribution upon initial observation (see Figure 2). Both distributions are skewed to the right, as the data groups towards the lower endpoint of the scale. The relative frequencies of values of the NA standardized score, assigned to categories, for both conditions separately, can be found in Table 3.

H1: The Effect of Response Format on Emotional Variability

There was no significant effect found of condition on the measure of emotional variability, $F(1, 87) = 0.296, p = .588$ (Table 4).

H2: The Effect of Response Format on Emotional Instability

Results showed that the condition did not have a significant effect on the measure of emotional instability, $F(1, 87) = 0.0004, p = .984$ (Table 4).

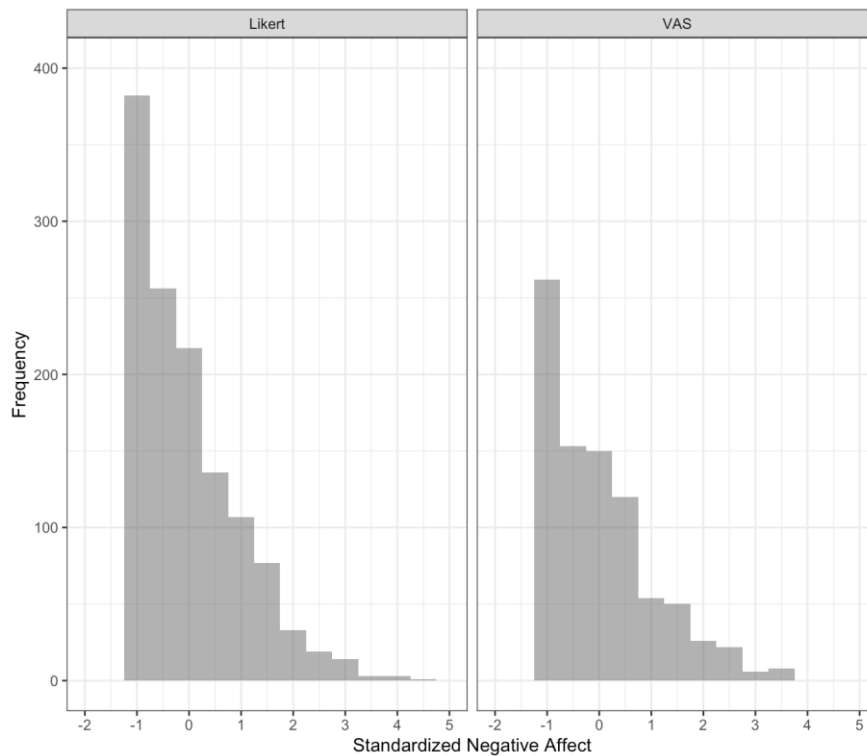
Table 2*Descriptive Statistics of the Baseline Questionnaire Sample and the NA Composite Scores*

Variable	Likert ($n_{Baseline} = 28$)		VAS ($n_{Baseline} = 35$)	
Mean Age (<i>SD</i>)	29.9 (13.7)		30 (14)	
	<i>n</i>	%	<i>n</i>	%
Gender				
Female	17	60.71	23	65.71
Male	11	39.29	12	34.29
Nationality				
German	27	96.43	34	97.14
Dutch	1	3.57	1	2.86
Occupation				
Working	12	42.86	10	28.57
Student	8	28.57	12	34.29
Studying and Working	8	28.57	11	31.43
Self-employed	0	0	1	2.86
Other	0	0	1	2.86
Education (highest degree completed)				
High School	9	32.14	16	45.71
Bachelor	13	46.43	11	31.43
Master	5	17.86	8	22.86
Other	1	3.57	0	0
NA Composite				
	Likert ($n_{ESM} = 53$)		VAS ($n_{ESM} = 36$)	
Mean (<i>SD</i>)	1.95 (1.03)		19.8 (17)	
Skewness	1.37		1	
Kurtosis	1.88		0.58	

Note. $N_{Baseline} = 63$. Level 2 $N_{ESM} = 89$. Lower $N_{Baseline}$ than N_{ESM} due to the baseline questionnaire not being triggered for those in the Likert condition during the first cohort.

Figure 2

Histogram of the NA Standardized Composite Scores for the Likert and VAS Condition



Note. The Likert condition has level 1 $n = 2014$, the VAS condition has level 1 $n = 1371$. This accounts for the large visual difference between the two histograms.

Table 3

Relative Frequencies of Observations of the NA Composite Scores for the Likert and VAS Condition

Category	Likert ($n = 2014$)		VAS ($n = 1371$)	
	Count	%	Count	%
1	722	38.33	617	45
2	799	39.67	386	28.15
3	235	11.67	203	14.81
4	167	8.29	117	8.53
5	28	1.39	36	2.63
6	9	0.45	12	0.88
7	4	0.2	0	0

Note. Level 1 $N = 3385$. For ranges of values for the categories see Appendix D.

Table 4

Model Results of Condition on Emotional Variability, Emotional Instability, and Emotional Inertia

Effect	Std. Estimate	SE	df	95% CI		p
				LL	UL	
Emotional Variability ^a						
Condition ^b	.033	.062	87	- 0.09	.158	.588
Emotional Instability ^a						
Condition ^b	.002	.12	87	- 0.236	.241	.984
Emotional Inertia ^{c d}						
Condition ^b	-0.019	.069	87	- 0.156	.117	.78
NA(t-1)	.632	.027	2012	.58	.685	< .001
Condition ^{a *} NA(t-1)	-0.062	.035	2012	- 0.062	.007	.077

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

^a $N = 89$.

^b 0 = VAS Condition, 1 = Likert Condition.

^c Dependent variable: Negative Affect

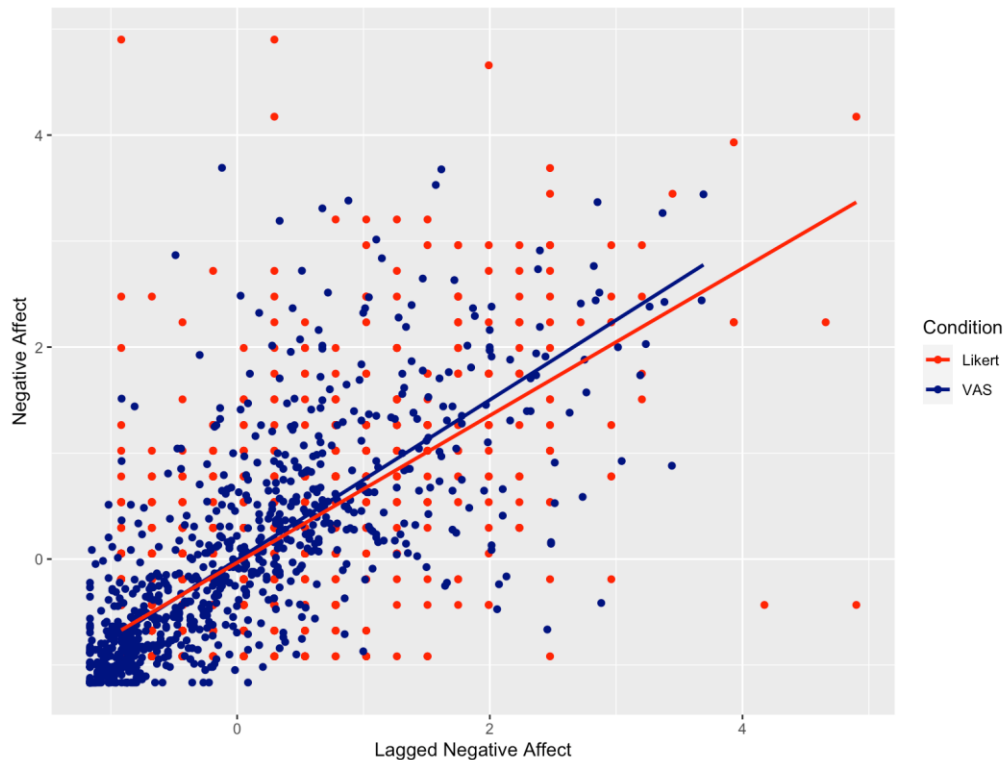
^d Level 1 $N = 2103$, Level 2 $N = 89$.

H3: The Effect of Response Format on Emotional Inertia

The interaction between the lagged NA variable and condition was not significant, $b = - 0.062$, $p = 0.077$ (Table 4). This suggests that the effect of the lagged NA variable on NA scores does not significantly differ across conditions. Figure 3 shows the regression lines of the effect of the lagged NA variable on NA, of condition separately.

Figure 3

Regression Lines of the Effect of NA Lag on NA Coloured by Condition



Discussion

This study examined the effect of response format on emotional variability, emotional instability, and emotional inertia within an ESM study, specifically the difference between using a Likert scale versus using a VAS. Overall, results indicated a non-significant effect of the response format on all three concepts of emotion dynamics.

Main Findings

Emotional variability, emotional instability, and emotional inertia scores showed no significant difference depending on the response format. All three hypotheses that (1) emotional variability, (2) emotional instability, and (3) emotional inertia of NA will be higher when using a VAS compared to when using a Likert scale, must therefore be rejected. These findings are contradicting to the reviewed literature and the consequent assumptions based on that literature.

The results of this study suggest that the choice of response format in ESM studies does not have an effect on an individual's reported emotional fluctuations, neither intensity nor frequency, nor in how far an individual's emotions remain static over time. This is in line with the findings of Gries et al. (2018), Kuhlmann et al. (2017), and Lukacz et al. (2004) who

found no differences in distribution between responses using Likert scales or VAS in retrospective studies. In addition to this, it seems that neither condition suffered from endpoint aversion. Rather, both conditions showed a positive skewness in their distributions and more scores at the lower end of the scale. The incongruity between the literature on which the hypotheses were based and the study findings could be accounted for by several factors.

Firstly, emotion dynamics might generally not be strongly influenced by the choice of response format. Emotion dynamics are complex constructs influenced by various factors, including situational context, individual differences, and internal processes (Kuppens & Verduyn, 2017). It is possible that these factors outweigh the potential impact of response formats, leading to non-significant effects in this study. Similarly, Kuhlmann et al. (2017) pointed out that the discrepancy between their findings of little difference between VAS and Likert and other studies' findings of existing differences might be because of the different assessed constructs. The same could apply here, since the current study focused solely on NA, whereas other studies investigated happiness (Studer, 2012) or patient satisfaction (Voutilainen et al., 2016). Additionally, von Klipstein et al. (2023) found in their ESM study that distributions of NA scores are characterized by a floor effect for non-depressed individuals. This would, for one, explain the distribution found in the current study, but could also account generally for the non-significant findings, since the floor effect could mask any differences in VAS and Likert. Therefore, exploring the effects of response format on other emotional constructs beyond NA, such as PA, would provide a more comprehensive understanding of the influence of response format on emotion dynamics. Examining the individual NA items could also be of interest, as the composite score could possibly conceal subtle differences.

Secondly, individual differences also play a role when it comes to response patterns that are found in the data. Arce-Ferrer (2006) for example demonstrated that individuals from Mediterranean and Latin American cultures exhibit higher levels of ERS compared to individuals from other cultural backgrounds. As the sample used in the current study was relatively homogenous in that it only consisted of German and Dutch participants, effects that might be visible in specific cultures potentially got lost in the current study. This could also account for the discrepancies between the findings of previous literature and this study, considering that other studies may have used culturally different or more diverse samples. Incorporating these factors into the investigation could be an avenue for future research. Examining how situational characteristics or individual and cultural differences interact with

response formats and therefore maybe moderate the influence of response formats on emotion dynamics could reveal nuanced effects and help optimize the design and implementation of ESM studies.

Thirdly, the ecological validity of ESM might mitigate the potential influence of response formats on emotion dynamics. ESM allows for immediate reporting of everyday life experiences within participants' natural environments, minimizing the reliance on retrospective recall and therefore potential memory biases (Myin-Germeys & Kuppens, 2022). The hypotheses of this research, however, were solely based on findings of retrospective studies. This would suggest that the immediate and contextually grounded nature of ESM data collection could make ESM studies more robust and provide a more accurate reflection of individuals' emotions, irrespective of the response format used, as compared to retrospective studies. This could explain the contradicting findings of the current study and of retrospective studies.

Still, the question remains open, what other criteria determine the suitability of the response format within ESM studies, if not the effect on emotion dynamics. For instance, what could also influence the design choice of an ESM study is the precision of the response formats used which could be assessed by investigating which response format leads to the smallest standard error. Or if the data resulting from one of the response formats are better suited for statistical models when examining the assumptions of certain models. It would also be of interest to consider psychometric properties such as the reliability and validity of items when different response formats are used. Future studies should explore these possibilities in order to open a debate on how to determine what a quality criterion for a good choice of response format is.

Strengths and Limitations

The strength of this study was that it was the first to investigate the effect of response format on emotion dynamics within an ESM context. It, therefore, provides valuable information and feeds possible guidelines for future ESM studies. Furthermore, the study was designed as a randomized controlled trial to ensure the reliability and validity of the study results. Finally, the design of the questionnaire was also based on recommendations of previous ESM studies (Dejonckheere & Erbas, 2022; Myin-Germeys et al., 2018) and a reliable measure of NA was used.

Nevertheless, it should be acknowledged that there were also limitations that might have influenced the validity of the results. Firstly, the obtained sample was not representative of the population due to the used method of convenience sampling. Participants were

primarily German, female, and had a mean age of 29.71 which limits the generalizability of the results. Secondly, NA is differently operationalized in the literature (Kirtley et al., 2019; Wetzel & Greiff, 2018). Different operationalizations might lead to different findings, suggesting that the questions used to measure emotion dynamics might also have an influence on it.

Thirdly, the low compliance rate of participants is another limitation. During data cleaning, 143 participants had to be removed, which also led to the sample being considerably smaller than indicated by the power analysis to be necessary to achieve a power of .8. The low compliance rate could be accounted for by diminished motivation of participants to continue filling out the questionnaires regularly. This may have also resulted in individuals with low motivation being unrepresented in the condition, which may undermine the comparability of the samples and therefore could affect the validity of the results.

Implications

This study makes several potential contributions to the field of emotion dynamics within the context of ESM studies. The findings of this study tentatively suggest that the choice of response format does not seem to significantly impact emotion dynamics in ESM studies. As this would provide evidence that contradicts previous research, this study prompts a re-evaluation of the role of response format in ESM studies and calls for further investigation into the factors that influence emotion dynamics.

If the results were to be verified, they would implicate that researchers can use either a Likert scale or a VAS as response formats without significant consequences for the reported emotional variability, instability, or inertia. This would provide researchers with flexibility in choosing the most suitable response format based on their specific research goals and participant characteristics when researching emotion dynamics. Moreover, understanding that response formats might have minimal impact in an ESM context could inform the design and implementation of future studies, allowing researchers to focus on other crucial aspects of their research, such as participant engagement, compliance, and data quality. Here, future studies could aid in defining what other aspects are important to ensure data quality or suitability of response format to the specific study.

Conclusion

This study investigated the effect of response format on emotion dynamics within an ESM study and found that the choice between a Likert scale and a VAS did not significantly influence emotional variability, emotional instability, or emotional inertia. These results are

partly contradictory to the literature on retrospective studies and highlight the importance of considering the real-time nature of ESM data collection, which may mitigate the influence of response formats. Future research should explore other emotional constructs and incorporate contextual factors to optimize the design and implementation of ESM studies.

References

- Arce-Ferrer, A. J. (2006). An Investigation Into the Factors Influencing Extreme-Response Style. *Educational and Psychological Measurement*, 66(3), 374–392.
<https://doi.org/10.1177/0013164405278575>
- Bernstein, E. E., Curtiss, J. E., Wu, G. W. Y., Barreira, P. J., & McNally, R. J. (2019). Exercise and emotion dynamics: An experience sampling study. *Emotion*, 19(4), 637–644. <https://doi.org/10.1037/emo0000462>
- Dejonckheere, E., & Erbas, Y. (2022). Designing an Experience Sampling Study. In I. Myin-Germeys, P. Kuppens, *The Open Handbook for Experience Sampling Methodology* (pp. 33-70). Retrieved from <https://www.kuleuven.be/samenwerking/real/real-book/index.htm>
- Eisele, G., Kasanova, Z., & Houben, M. (2022). Questionnaire Design and Evaluation. In I. Myin-Germeys, P. Kuppens, *The Open Handbook for Experience Sampling Methodology* (pp. 71-89). Retrieved from <https://www.kuleuven.be/samenwerking/real/real-book/index.htm>
- Ethica Data Services Inc. (2023). *Ethica* (No. 651). <https://ethicadata.com>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Funke, F., & Reips, U. D. (2012). Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. *Field Methods*, 24(3), 310–327. <https://doi.org/10.1177/1525822X12444061>
- Gries, K., Berry, P., Harrington, M., Crescioni, M., Patel, M., Rudell, K., Safikhani, S., Pease, S., & Vernon, M. (2018). Literature review to assemble the evidence for response scales used in patient-reported outcome measures. *Journal of Patient-Reported Outcomes*, 2. <https://doi.org/10.1186/s41687-018-0056-3>
- Hayes, J. E., Allen, A. L., & Bennett, S. M. (2013). Direct comparison of the generalized visual analog scale (gVAS) and general labeled magnitude scale (gLMS). *Food Quality and Preference*, 28(1), 36–44. <https://doi.org/10.1016/j.foodqual.2012.07.012>

- Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. <https://doi.org/10.1037/a0038822>
- Janssens, M., Eshuis, J., Peeters, S., Lataster, J., Reijnders, J., Enders-Slegers, M.-J., & Jacobs, N. (2020). The Pet-Effect in Daily Life: An Experience Sampling Study on Emotional Wellbeing in Pet Owners. *Anthrozoös*, *33*(4), 579–588. <https://doi.org/10.1080/08927936.2020.1771061>
- Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance. *Multivariate Behavioral Research*, *50*(3), 334–349. <https://doi.org/10.1080/00273171.2014.1003772>
- Kincaid, C. (2005). Guidelines for Selecting the Covariance Structure in Mixed Model Analysis. <https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/198-30.pdf>
- Kirtley, O. J., Hiekkaranta, A. P., Kunkels, Y. K., Eisele, G., Schoefs, S., Kemme, N. D. F., Le Grange, J., Simsa, B., Biesemans, T., Van Heck, L., Myin-Germeys, I. (2023, May 22). *The Experience Sampling Method (ESM) Item Repository*. <https://doi.org/10.17605/OSF.IO/KG376>
- Kuhlmann, T., Dantlgraber, M., & Reips, U.-D. (2017). Investigating measurement equivalence of visual analogue scales and Likert-type scales in Internet-based personality questionnaires. *Behavior Research Methods*, *49*(6), 2173–2181. <https://doi.org/10.3758/s13428-016-0850-x>
- Kuppens, P. (2015). Its about Time: A Special Section on Affect Dynamics. *Emotion Review*, *7*(4), 297–300. <https://doi.org/10.1177/1754073915590947>
- Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26. <https://doi.org/10.1016/j.copsyc.2017.06.004>
- Larson, R., & Csikszentmihalyi, M. (2014). The Experience Sampling Method. In *Flow and the Foundations of Positive Psychology* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2
- Lukacz, E. S., Lawrence, J. M., Burchette, R. J., Lubner, K. M., Nager, C. W., & Galen Buckwalter, J. (2004). The use of Visual Analog Scale in urogynecologic research: A psychometric evaluation. *American Journal of Obstetrics and Gynecology*, *191*(1), 165–170. <https://doi.org/10.1016/j.ajog.2004.04.047>

- Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry, 17*(2), 123–132. <https://doi.org/10.1002/wps.20513>
- Myin-Germeys, I., & Kuppens, P. (2022). Sampling Methods, an Introduction. In I. Myin-Germeys, P. Kuppens, *The Open Handbook for Experience Sampling Methodology* (pp. 7-18). Retrieved from <https://www.kuleuven.be/samenwerking/real/real-book/index.htm>
- Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag. <https://doi.org/10.1007/b98882>
- Pinheiro, J., Bates, D., & R Core Team. (2023). *nlme: Linear and Nonlinear Mixed Effects Models* (3.1-162). <https://cran.r-project.org/package=nlme>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.0). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Stratton, S. J. (2021). Population Research: Convenience Sampling Strategies. *Prehospital and Disaster Medicine, 36*(4), 373–374. <https://doi.org/10.1017/s1049023x21000649>
- Studer, R. (2012). Does it matter how happiness is measured? Evidence from a randomized controlled experiment. *Journal of Economic and Social Measurement, 37*(4), 317–336. <https://doi.org/10.3233/JEM-120364>
- Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology, 121*(4), 819–829. <https://doi.org/10.1037/a0027978>
- van der Linden, K., Simons, C., Viechtbauer, W., Ottenheim, E., van Amelsvoort, T., & Marcelis, M. (2021). A momentary assessment study on emotional and biological stress in adult males and females with autism spectrum disorder. *Scientific Reports, 11*(1), 14160. <https://doi.org/10.1038/s41598-021-93159-y>
- Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and Retention With the Experience Sampling Method Over the Continuum of Severe Mental Disorders: Meta-Analysis and Recommendations. *Journal of Medical Internet Research, 21*(12), e14475. <https://doi.org/10.2196/14475>
- von Klipstein, L., Servaas, M. N., Lamers, F., Schoevers, R. A., Wardenaar, K. J., & Riese, H. (2023). Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study. *Journal of Affective Disorders, 334*, 370–381. <https://doi.org/10.1016/j.jad.2023.04.118>

- Viechtbauer, W. (2022a). Statistical Methods for ESM Data. In I. Myin-Germeys, P. Kuppens, *The Open Handbook for Experience Sampling Methodology* (pp. 153-183). Retrieved from <https://www.kuleuven.be/samenwerking/real/real-book/index.htm>
- Viechtbauer, W. (2022b). Structuring, Checking and Preparing the Data. In I. Myin-Germeys, P. Kuppens, *The Open Handbook for Experience Sampling Methodology* (pp. 137-150). Retrieved from <https://www.kuleuven.be/samenwerking/real/real-book/index.htm>
- Voutilainen, A., Pitkäaho, T., Kvist, T., & Vehviläinen-Julkunen, K. (2016). How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing*, 72(4), 946–957. <https://doi.org/10.1111/jan.12875>
- 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>
- Wetzel, E., & Greiff, S. (2018). The World Beyond Rating Scales. *European Journal of Psychological Assessment*, 34(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>

Appendix A

Enrolment Information Sheet: Likert Condition

Dear participant,

Thank you for your participation in the study on mental health in daily life. We are contacting you because you kindly agreed to participate in this study for the bachelor psychology at the University of Twente.

Brief summary of the project

The study you are participating in is a daily diary study. With this study we want to investigate how people feel and react to events in their day-to-day lives. By asking a few questions at several moments throughout the day, we get an insight in behavior of people in their everyday environment, which is necessary if we want to understand how people behave and feel in daily life. You will receive a notification on 10 random moments a day to answer a short questionnaire which will take about 1 minute to complete. We ask you to do this for 7 days in a row. The first questionnaire will be send on Monday morning, the 17th of April. Of course, there are situations in which it is not possible to fill it out (such as when you are driving), but to get a good overview of your daily life it is important that you fill out as many of these questionnaires as possible. In addition to these short questionnaires, you will receive one questionnaire in the beginning of the study that takes about 20 minutes to complete. It's important that you complete this questionnaire as well.

How to get ready to participate

Before continuing, make sure to download the Ethica application on your smartphone.

Clicking on the following links on your smartphone will bring you the app store.

Android:

https://play.google.com/store/apps/details?id=com.ethica.logger&hl=en_US&gl=US&pli=1

IOS: <https://apps.apple.com/nl/app/ethica/id1137173052>

Then follow the these steps:

- Open the Ethica application on your phone. **Please make sure to allow push notifications for the Ethica app on your phone!**
- Click on “Sign up” and create an account.

- After you signed up in Ethica, login in to the Ethica application using your username and password.
- After logging in, click on the following link on your phone:

<https://ethicadata.com/study/2349/>

- Alternatively, you can also directly enter the registration code **2349** in the Ethica application.
- On the next window click on “Register” to enroll in the study.
- The study should now be set up and you will receive the first questionnaire next Monday.

Contact details

This study is part of a larger project with many researchers involved. If you have any questions, you can contact one of the following students who are involved in data collection or the supervisors. The contact details can be found below.

Students

Simon Brune

Nick Delventhal

Jan Derksen

Gina Haccou

Samuel Pietsch

Aleksandra Popovic

Lea Staudigel

Nina Zarrin Tigh

Supervisors

Jannis Kraiss

Thomas Vaessen

Thank you for participating in this study. Your contribution is greatly appreciated.

Kind regards, also on behalf of the whole study team,

Appendix B

Enrolment Information Sheet: VAS Condition

Dear participant,

Thank you for your participation in the study on mental health in daily life. We are contacting you because you kindly agreed to participate in this study for the bachelor psychology at the University of Twente.

Brief summary of the project

The study you are participating in is a daily diary study. With this study we want to investigate how people feel and react to events in their day-to-day lives. By asking a few questions at several moments throughout the day, we get an insight in behavior of people in their everyday environment, which is necessary if we want to understand how people behave and feel in daily life. You will receive a notification on 10 random moments a day to answer a short questionnaire which will take about 1 minute to complete. We ask you to do this for 7 days in a row. The first questionnaire will be send on Monday morning, the 17th of April. Of course, there are situations in which it is not possible to fill it out (such as when you are driving), but to get a good overview of your daily life it is important that you fill out as many of these questionnaires as possible. In addition to these short questionnaires, you will receive one questionnaire in the beginning of the study that takes about 20 minutes to complete. It's important that you complete this questionnaire as well.

How to get ready to participate

Before continuing, make sure to download the Ethica application on your smartphone.

Clicking on the following links on your smartphone will bring you the app store.

Android:

https://play.google.com/store/apps/details?id=com.ethica.logger&hl=en_US&gl=US&pli=1

IOS: <https://apps.apple.com/nl/app/ethica/id1137173052>

Then follow the these steps:

- Open the Ethica application on your phone. **Please make sure to allow push notifications for the Ethica app on your phone!**
- Click on “Sign up” and create an account.

- After you signed up in Ethica, login in to the Ethica application using your username and password.
- After logging in, click on the following link on your phone:

<https://ethicadata.com/study/1296/>

- Alternatively, you can also directly enter the registration code **1296** in the Ethica application.
- On the next window click on “Register” to enroll in the study.
- The study should now be set up and you will receive the first questionnaire next Monday.

Contact details

This study is part of a larger project with many researchers involved. If you have any questions, you can contact one of the following students who are involved in data collection or the supervisors. The contact details can be found below.

Students

Simon Brune

Nick Delventhal

Jan Derksen

Gina Haccou

Samuel Pietsch

Aleksandra Popovic

Lea Staudigel

Nina Zarrin Tigh

Supervisors

Jannis Kraiss

Thomas Vaessen

Thank you for participating in this study. Your contribution is greatly appreciated.

Kind regards, also on behalf of the whole study team,

Appendix C

Informed Consent

Dear participant,

Thank you for your participation in this study.

Brief summary of project

The study is using the Experience Sampling Method (ESM) to obtain data. This means that 10 times a day there will be a prompt to answer a questionnaire containing about 20 items, which will take about 1 minute to complete. The questions regard your psychological well-being in the specific moment you are receiving the questionnaire and the time in-between questionnaires. It is important to fill out as many questionnaires as possible to ensure the success of the project.

To participate in this study, we need to ensure that you understand the nature of the research, as outlined in the participant information sheet. Please confirm at the bottom of the page to indicate that you understand and agree to the following conditions:

- I confirm that I have read the participant information sheet for this study. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily
- I understand that to take part in this study, I should
- Be at least 18 years old
- Possess a basic level of English
- I understand that personal data about me will be collected for the purposes of the research study including age, gender, nationality, level of education, current studies, and primary occupation, and this data will be processed completely anonymous and in accordance with data protection regulations.
- I understand that taking part in this study involves that I will be filling in 10 questionnaires every day for one week.
- I am voluntarily taking part in this research, and I know that I can stop the research at any time without giving any reason, without my rights being affected
- I don't expect to receive any benefit or payment for my participation.

- I understand that I am free to contact the researchers or supervisor with any questions I may have in the future.
- I understand that the data collected in this study will be anonymized, and only be used for academic purposes i.e., writing a thesis for the bachelor and/or master.
- I understand that personal data that will be collected within this study will not be shared with anyone other than the study team.
- I agree to take part in this study.

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

Study contact details for further information:

Students

Simon Brune

Nick Delventhal

Jan Derksen

Gina Haccou

Samuel Pietsch

Aleksandra Popovic

Lea Staudigel

Nina Zarrin Tigh

Supervisors

Jannis Kraiss

Thomas Vaessen

Appendix D
Value Ranges Categories Relative Frequency

Table D1

Ranges of Values for Each Category for the Likert and VAS Condition

Category	Range	
	Likert	VAS
1	1.0 - 1.25	0 - 14
2	1.5 - 2.25	15 - 28
3	2.5 - 3.25	29 - 42
4	3.5 - 4.25	43 - 57
5	4.5 - 5.25	58 - 71
6	5.5 - 6.25	72 - 85
7	6.5 - 7.0	86 - 100