Does the Response Format Matter?
The Effects of VAS versus Likert Scales on Positive Affect Dynamics in Experience Sampling Questionnaires

Faculty of Behavioural Management and Social Sciences Department of Positive Psychology and Technology, University of Twente

M12 BSc Thesis PSY (2022-202000384-2A)

Gina L. Haccou

First Supervisor: Dr Jannis T. Kraiss

Second Supervisor: Dr Thomas R. Vaessen

Date: June 27th, 2023

APA 7th Edition

Abstract

Background: The following research consists of an Experience Sampling Method ESM study, which enables the recording of multiple observations per person over a longitudinal period. The influence of scale design in ESM studies has not yet been thoroughly researched, although it could have important implications for study design examining the effect of a Likert scale versus a VAS scale on distinct positive emotions (cheerfulness, enthusiasm, satisfaction, and relaxation) and the analysis of their affect dynamics (emotional variability, instability and inertia)

Methods: The final sample of N = 80 was gathered via convenience sampling. Participants were randomly allocated to either the VAS or Likert condition. They participated for seven days via their mobile phone, whilst receiving ten semi-random prompts per day to fill in ESM questionnaires. For emotional variability and instability, a linear model was investigated, and for inertia, a linear mixed model was used to analyse the effect of condition on these outcomes.

Results: The analysis yielded partially significant results for the interaction effect of condition and inertia for the items cheerfulness, $t = 2.208$, $p = .027$ and relaxation, $t = 2.087$, $p = .037$. No other effects of conditions on affect dynamics for any of the four other positive emotions were significant.

Conclusion: The outcomes indicated that the VAS scale could be more sensitive for measuring inertia for some positive emotions, which should be considered when looking into the variation and fluctuation of emotions. Overall, this study provided insights into scale choices for investigating positive affect and should motivate to do further research into this topic. There were some limitations concerning the study design that could be improved, especially regarding the sample size and its representativeness.

Keywords: ESM, response scales, positive affect, heuristics

Introduction

Emotional states fluctuate throughout the day, meaning that single assessments cannot cover the variety of emotions experienced in daily life (Houben et al., 2015). A solution to this is the use Of Experience Sampling Methods, enabling the recording of multiple observations per day (Csíkszentmihályi & Larson, 1987). An essential aspect of designing an ESM study is the choice of response scales, as they potentially influence the participant's responses (Hasson & Arnetz, 2005; Kieruj & Moors, 2010; Eisele et al., 2021, Chapter 4). Popular response scales in ESM are the Likert scale and the VAS scale (Eisele et al., 2021, Chapter 4). The exact influences of the two scales on response behaviour remain unclear (Eisele et al., 2021, Chapter 4). Therefore, this paper aims to address the research gap concerning the effect of using a VAS scale versus a Likert scale in ESM, with a focus on positive affect items. This study will examine potential biases and relevant implications for choosing the most appropriate response scales in ESM studies, to reduce biases in future research.

The Experience Sampling Method

ESM is an intensive longitudinal data collection method (Myin-Germeys & Kuppens, 2021, Chapter 1). The idea is, that participants can provide real-time answers about their experiences to map out the individuals' daily psychological functioning, which tends to undergo frequent changes throughout the day (Csíkszentmihályi & Larson, 1987; Myin-Germeys & Kuppens, 2021, Chapter 1; Verhagen et al., 2016).

Data is collected over multiple points during the day, ensuring a short recall period of the event of interest. As a result, this limits the so-called retrospective memory bias, which can cause false responses (Dejonckheere & Erbas, 2021, Chapter 3; Verhagen et al., 2016). Thus, increasing the validity of the data. Moreover, ESM includes repeated measurements; this means that multiple observations are recorded within one person. Subsequently, the data can be a good predictor for within-person analyses (Curran & Bauer, 2011). For example, when examining affect dynamics, which will be an important part of the following data analysis.

Response Scales: Likert Scale and VAS Scale

Both the Likert scale and VAS scale are commonly used for the measurement of subjective experiences in self-report studies (Eisele et al., 2020, Hasson & Arnetz, 2005), for example when measuring affect dynamics (Schimmack, 2003). Although findings are contradictory, the majority of conducted research concludes that both scales display high reliability, validity, and responsiveness (Eisele et al., 2021, Chapter 4; Hasson & Arnetz, 2005).

The structure is as follows: the 7-point Likert scale is categorical and includes seven distinct categories, ranging from ‘totally disagree’ to ‘totally agree’. Whereas the VAS scale is a continuous scale, consisting of an analogue slider to indicate agreeableness to the statement on a range from ‘0’ to ‘100’, without any pre-determined categories (Eisele et al., 2021, Chapter 4; Hasson & Arnetz, 2005).

Although interchanging Likert and VAS scales can produce similar results, Hasson and Arnetz (2005) showed in their research that the results are comparable but not identical. According to their study, the responses for VAS and Likert items were highly correlated but also showed significant differences for the absolute levels of half the items. Confirming the idea, that response scales do produce different outcomes. However, not much research has been conducted comparing these different response scales (Eisele et al., 2021, Chapter 4). This raises the question of how these differences occur. One possible explanation could be how the scale might influence a person’s response style.

Response Styles: Extreme Responding

A response style is a personal bias in response behaviour. A common example is the tendency to systematically give extreme responses. This is called ‘Extreme Response Style’’ ERS (Kieruj & Moors, 2010; Weijters et al., 2010). Often extreme responses are triggered by personal style and hence, do not accurately reflect the measured variable (Kieruj & Moors, 2010). Kieruj and Moors (2010) found that the scale format can influence response styles, which was also found by researchers such as Vaerenbergh and Thomas (2013) and Weijters et al. (2010). It is notable, that engaging in ERS is a stable characteristic and is also systematically found within longitudinal studies as ESM (Weijters et al., 2010). It can harm the validity of data (Kieruj & Moors, 2010; Steenkamp, 2001). Possible consequences can be a skewed frequency distribution towards extreme endpoint scores, high variance and low correlation coefficients (Kieruj & Moors, 2010).

There are multiple factors potentially contributing to extreme response styles. Overall cognition, personality and demographic variables can be of importance. When the cognitive demand is too challenging, the respondent might try to simplify their answers (Morsanyi & Handley, 2007). Therefore, the need for significant cognitive effort can facilitate the usage of heuristics when answering survey items and promote ERS (Baumgartner & Steenkamp, 2001). The likelihood of applying heuristics is thought to be directly related to task difficulty (Krosnick, 1991). So, in this case, the ease of scale use.

Heuristics include ‘satisficing’ and ‘anchoring’. First, satisficing occurs when the participant does not want to apply the required effort and time to give accurate answers and

uses a mental shortcut to provide an answer that is satisfactory (Kieruj & Moors, 2010; Krosnick, 1991). For anchoring, the respondent orients their answers towards certain anchor points on the scale, this is especially the case for wide-ranged scales that possess anchor points towards their extreme ends (Kieruj & Moors, 2010; Weathers et al., 2005). Based on this, previous research also supports that heuristics, especially ERS, are more prevalent for VAS scales. For instance, Hasson and Arnetz's (2005) found systematic biases in an individual's response style when comparing the VAS and Likert scales. For Likert scales, participants tended to choose mid-range responses and for VAS scales the scores were more towards the extreme ends of the scale (Hasson & Arnetz, 2005).

Affect Dynamics

Emotions are very complex and fluctuate during the day. These fluctuations can be interconnected and may follow systematic patterns (Houben et al., 2015; Kuppen et al., 2010; Pirla et al., 2022; Schimmack, 2003). These individual differences in emotional experiences are known as affect dynamics and play an important role in health and well-being (Pirla et al., 2022), as well as psychopathology (Houben et al., 2015; Pirla et al., 2022) and life satisfaction (Schimmack, 2003).

For measurement, affect dynamics can be split up into three main components. These are emotional variability, emotional instability and emotional inertia (Houben et al., 2015; Pirla et al., 2022), which are constructs that will be analysed within this research. Emotional variability describes the range of the person's affective states over time, indicating how much it tends to change during the day (Houben et al., 2015; Pirla et al., 2022). High levels of variability imply more extreme responses (Houben et al., 2015) and negatively affect subjective well-being (Pirla et al., 2022). It can be calculated by looking at the within-person standard deviation SD (Houben et al., 2015; Pirla et al., 2022). Emotional instability represents how extreme the changes in emotional experiences are when comparing and quantifying the changes from one instant to the next (Houben et al., 2015; Pirla et al., 2022). High levels are associated with poor mental health (Pirla et al., 2022) and mean that emotional shifts are more extreme (Houben et al., 2015). It is operationalised as the root mean square of successive differences RMSSD (Houben et al., 2015; Pirla et al., 2022). The level of resistance towards these changes is called emotional inertia. It describes how well an emotional experience can be predicted by a previous moment (Houben et al., 2015; Pirla et al., 2022). A high score infers that emotional states persist longer. If this is experienced for PA, it can be beneficial for subjective well-being. Inertia is investigated by the autocorrelation

by regressing the current emotional state on the previous state (Houben et al., 2015; Pirla et al., 2022).

Distinct Emotions and Scales

This study will focus on four distinct emotions, cheerfulness, enthusiasm, satisfaction and relaxation. As beforementioned, the variance and fluctuations of emotions can be examined by looking into affect dynamics. However, this is also highly dependent on adequate measurement scales (Lucas et al., 2009, pp. 139–156). In this study, the focus is on the intensity of emotions. For Likert, the participants choose a category that corresponds to the perceived strength of their emotions. For VAS, the participants can choose freely.

The different response scales could vary in their ability to capture the dynamic character of positive affect (Lucas et al., 2009, pp. 139–156). To investigate which scale is more accurate, VAS or Likert, the positive affect items can be compared by looking at the scales' sensitivity to emotional changes, regarding variability, instability and inertia.

Current Study

This study will focus on the different outcomes and biases related to response formats on positive affect items. Participants are randomly allocated to a VAS or Likert condition. The VAS condition will likely display different absolute levels in affect dynamics. Possible explanations could be that the VAS scale promotes heuristics and facilitate the tendency to select more extreme values. Choosing extreme responses in VAS may take up less time and effort than evaluating the meaning of the specific ranges within the continuous scale. Also, the extreme points might appear clearer in their meaning and less ambiguous than the scores in between. This is expected to be reflected by finding significant differences in affect dynamics between the two conditions. The more extreme responses may lead to higher emotional variability, instability, and autocorrelation.

Based on these assumptions, the following research question and hypotheses are formulated: *What is the effect of using a Likert scale compared to a VAS scale on emotional variability, emotional instability and emotional inertia of four positive distinct emotions commonly used in ESM studies (cheerfulness, enthusiasm, satisfaction and relaxation)?*

Hypothesis 1: Emotional Variability

Participants in the VAS condition demonstrate significantly higher emotional variability of distinct positive emotions compared to the Likert condition.

Hypothesis 2: Emotional Instability

Participants in the VAS condition demonstrate significantly higher emotional instability of distinct positive emotions compared to the Likert condition.

Hypothesis 3: Emotional Inertia

Participants in the VAS condition demonstrate significantly higher emotional inertia of distinct positive emotions compared to the Likert condition.

Methods

This research was approved under request number 22144 by the Ethics Committee of Behavioural, Management and Social Sciences (BMS) of the University of Twente.

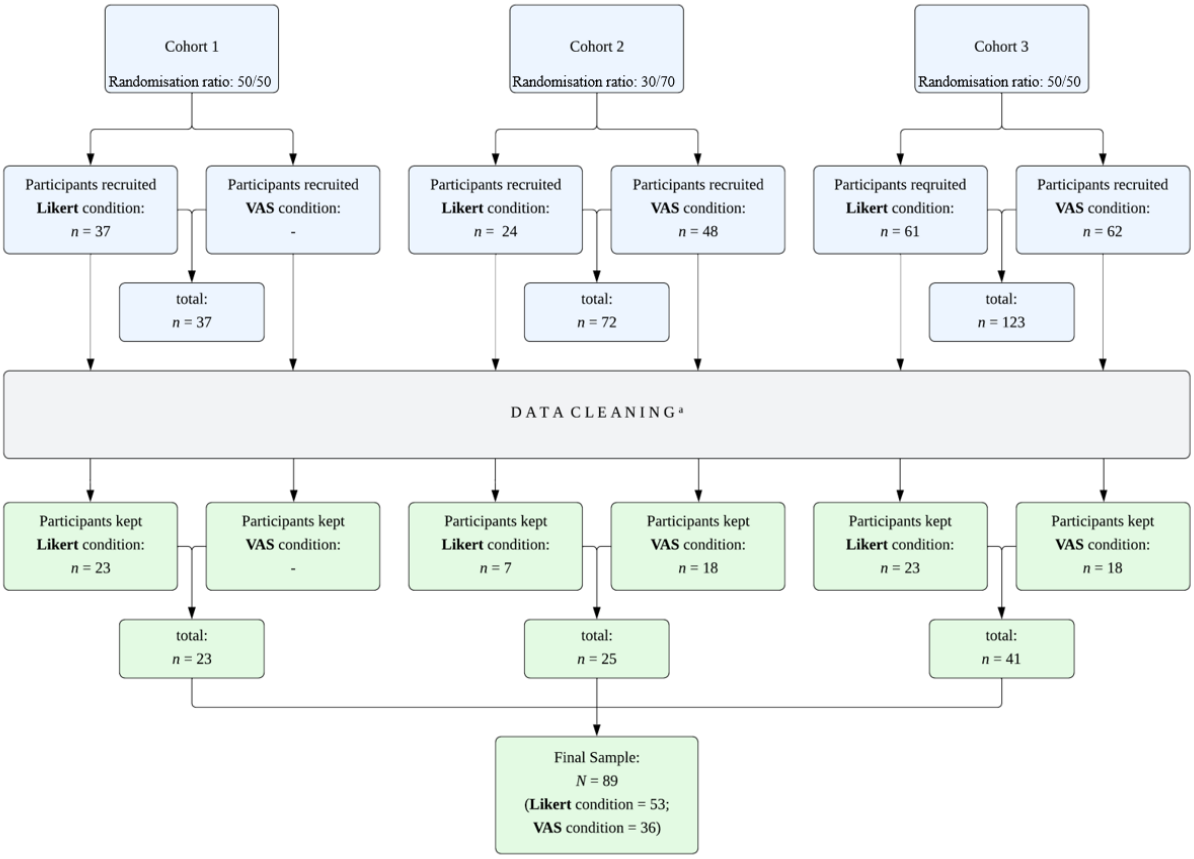
Participants

Participants were gathered via the online tool SONA test subject tool of the University of Twente (https://utwente.sona-systems.com/) and the personal network of each participating researcher. This implies the use of convenience sampling and volunteer sampling. This method was chosen because of its time and cost-efficacy. Furthermore, this meant that the participants were voluntarily motivated to engage in the data collection.

Requirements for participation were 18 years or older, English reading skills and possession of a mobile phone. As shown in Figure 1, participants were recruited in three different cohorts. Overall, 232 people were enrolled, distributed over three different cohorts. They were divided into multiple conditions. For this study, the relevant conditions were the Likert condition (N=122) and the VAS condition (N=110). An a priori power analysis in G*Power (Faul et al., 2009) showed that at least 64 participants per condition were required. The parameters for this test were set for a linear model with an effect size of $f = .25$ and $\alpha = .05$, aiming to reach a statistical power of .8.

Figure 1

Flowchart of Data Collection Process



Note. Figure 1 provides an overview of the data collection process and the steps for obtaining the final sample size that was used for this study. The relevant data were extracted from each cohort and followed by a data cleaning process, during which participants with a response rate below 33% were deleted. Leaving the data set with a final sample size of N = 89.

Randomisation

A randomised controlled trial was conducted. The participants were randomly allocated to two conditions, using stratified randomisation. The control group condition received the study with a Likert scale as a response format and for the experimental group, the participants received the same questionnaires but with a VAS scale. Randomisation was done within separate lists created in Excel per researcher and for SONA using a pattern from www.sealedenvelope.com. The goal was to ensure that the groups are not biased, by accounting for environmental influences per researcher and maximising the variability of the participants (Lafit, 2021, Chapter 11). Since this study was part of a larger project, data

collection took place in multiple cohorts. The three different cohorts collected data at different time periods and with different randomisation ratios (Figure 1).

Design and Procedure

To conduct the study in an ESM format, the application Ethica was used (https://www.ethicadata.com). The participants were initially briefed on the relevant information of the study. After giving informed consent (Appendix E) the participants were able to start the study following the starting date.

Data collection always occurred over a time span of seven days from Monday to Sunday. The time periods for the different cohorts were as follows. The first cohort started Monday 7th of November 2022 and ended Sunday 12th of November 2022. The second was from the 13th of February 2023 to the 20th of February 2023. The latest period took place from the 17th of April 2023 to the 23rd of April 2023.

The first activity was filling out a baseline questionnaire. After, Ethica instructed the user to participate ten times a day, with a short questionnaire. In total, one general questionnaire was answered plus 70 short ESM questionnaires. The short questionnaires were spread over the day and triggered at random during pre-defined 90-minute time intervals from 7:30 to 21:00. At such a point the participant received a notification to inform them that a response was required within the next 15 minutes. The questionnaire expired if there was no response within this time frame. The study schedule is displayed in Table 5 (Appendix A).

Measures

Since this study was part of a larger project, several baseline and ESM questionnaires about mental health and emotion regulation were administered. In the following, only questionnaires will be described that are relevant to the current study.

Baseline Questionnaires

For the demographics, a baseline questionnaire was conducted. Giving information about the age, gender, nationality, occupation, and education of the participant.

ESM Questionnaires

Positive emotions were measured with four items from the Positive and Negative Affect Schedule (PANAS), which is according to Watson et al. (1988) a reliable construct. These items include cheerfulness, enthusiasm, satisfaction and relaxation. With the question: '*How cheerful/ enthusiastic/satisfied/relaxed do you feel right now?*'. These items were also used in a previous study for looking into the variability of emotions, intention, self-efficacy and physical complaints (Maes et al., 2022). Another study by Watson and Naragon (2009)

also included the item cheerfulness to assess positive affect. Whereas the question style was suggested by Lucas et al. (2009, pp. 139–156) to make simple assessments about PA.

The conditions received different response scales. The Likert scale consisted of seven answer categories ranging from (1) strongly disagree to (7) strongly agree. For the VAS scale condition, there was an analogue slider with a range from 0 to 100.

Data Analysis

Pre-Processing

The data for the two conditions were downloaded separately and merged, whilst adding a condition variable indicating whether people were allocated to the VAS or Likert condition (0= Likert scale, 1=VAS scale). After, a time variable was added indicating the assessment point. The next step contained data cleaning, during which all irrelevant participants, for example from the pilot or other conditions, were excluded. This was followed by deleting all participants with a response rate below 33% which meant having less than 23 observations within one person. Subsequently, 143 participants were excluded, leaving a final sample of N=89. The final step was a z-score standardisation of the scores of the four positive affect items, to ensure the comparability of results between conditions.

For the first hypothesis on emotional variability, the within-person standard deviations for all four PA items were calculated and added as time-invariant variables to the data. Concerning emotional stability, the analysis included the calculation of the mean square successive difference (mssd) between emotion scores for each item, which was also added as a time-invariant variable to the data frame. For the third hypothesis on emotional inertia, a lag-1 variable for each positive affect item was added to the data.

Statistical Analyses

The statistical analysis was done in R (https://www.r-project.org) and RStudio. First, the sample characteristics were analysed and visualised within a table of demographics (Table 1). Secondly, the distribution of the data was visualised with histograms for each item and per condition (Figure 2) with the help of the ggplot package in R (Wickham, 2009). Then the hypotheses were tested by creating appropriate models per affect dynamics item. For the analysis of emotional variability and emotional instability, a new data set was created. This included a linear model with a fixed effect for emotional variability and one for emotional instability. Only one row per participant was included to account for nested observations and avoid a unit-of-analysis error.

To check the first hypothesis on emotional variability, the linear fixed-effect model was used to compare the two conditions and check for a significant difference between them

(Table 2). For each item, the dependent variable was the within-person standard deviation, and the independent variable was the condition, which was set as the fixed effect. This was done for all four individual positive affect items separately. For the second hypothesis about emotional instability, the statistical analysis was also based on a linear fixed-effect, the dependent variable being mssd and the independent variable. The third hypothesis was tested on significance by using four separate linear mixed models (Table 3). For this, the effect of the condition was examined by including an interaction between the condition and the lagged variables of the four positive affect items, respectively. The interaction of the two independent variables, lag-1 variable and condition, serves as an indication of the effect of condition on emotional inertia. The models also include the participants to account for the multilevel structure of the data. For LMMs, the lme4 package was used (Bates et al., 2015).

Results

Sample Characteristics

Overall, 232 participants were gathered for data collection. However, after data cleaning only 53 participants remained for the Likert condition and 36 for VAS, of which nine participants were later deleted in order to conduct regression analyses. Leaving a sample of N = 80. Causes for the dropout are missing data from application errors, technical issues during the data collection period in Ethica, and many participants with a response rate below 33% of all the questionnaires. Concerning the characteristics (Table 1), the majority of respondents were female (Likert = 60.71%, VAS = 65.71%). Also, the average age was approximately 30 for both conditions. For nationality, the sample consisted mostly of Germans (Likert = 96.43, VAS = 97.14).

Table 1

Sample Characteristics of the Baseline Questionnaire from Cohort 3

Variable	Category	Likert ($n_{Baseline}$ = 28)		VAS ($n_{Baseline}$ = 35)	
Age (mean, *SD*)		29.9 (13.7)		30 (14)	
		n	%	n	%
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 (completed)					
	High School	9	32.14	16	45.71
	Bachelor	13	46.63	11	31.43
	Master	5	17.86	8	22.86
	Other	1	3.57	0	0

Note. $N_{Baseline}$ is lower than N_{ESM} as the baseline questionnaire was not triggered for all cohorts.

Visualisation of the Data

To compare the central tendencies of responses for each condition, a table containing relative frequencies was created (Table 2). Most observations were allocated in category five, corresponding to ‘somewhat agree’ in Likert and range 57 to 70 in VAS. Additionally, histograms of the standardised responses can be found in Figure 2. Both conditions show slightly positively skewed distributions with altogether, a similar kurtosis. The central tendencies and overall distributions appear very similar.

Table 2

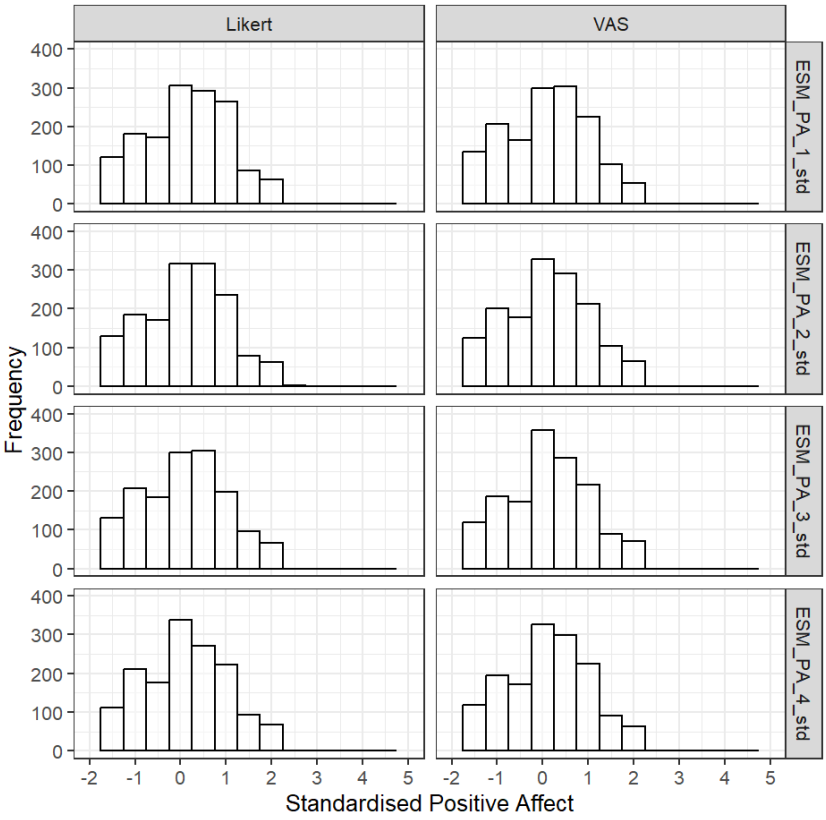
Relative Frequency of Responses

Category	Likert				VAS			
	PA1	PA2	PA3	PA4	PA1	PA2	PA3	PA4
Relative frequency (frequency)								
1	5.1 % (144)	7.8 % (221)	5.6 % (157)	3.9 % (109)	6.1 % (133)	7.2 % (157)	6.0 % (130)	4.3 % (94)
2	9.6 % (271)	13.8 % (390)	8.2 % (231)	8.4 % (235)	12.7 % (277)	15.8 % (345)	12.1 % (263)	10.7 % (233)
3	17.9 % (505)	19.8 % (560)	15.8 % (446)	19.3 % (544)	15.4 % (336)	19.0 % (414)	13.5 % (295)	15.1 % (330)
4	22.4 % (691)	**23.4 % (660)**	23.4 % (660)	20.4 % (575)	19.0 % (407)	18.8 % (411)	19.7 % (429)	15.5 % (339)
5	**25.7 % (726)**	21.2 % (599)	**24.6 % (722)**	**24.6 % (692)**	**26.1 % (566)**	**20.7 % (451)**	**22.4 % (489)**	**24.9 % (543)**
6	12.0 % (339)	9.5 % (268)	14.9 % (420)	14.5 % (409)	16.1 % (349)	13.3 % (291)	18.9 % (412)	21.0 % (459)
7	5.3 % (151)	4.4 % (125)	6.5 % (184)	8.9 % (250)	5.2 % (114)	5.2 % (113)	7.5 % (164)	8.4 % (184)

Note. The central tendencies are highlighted in bold. For interpretation, the corresponding ranges for the VAS condition can be found in Appendix B (Table 6). Please consider that the relative frequencies were calculated after the initial data cleaning (N = 89), later, further data was omitted due to missing values throughout the conducted statistical analyses (N = 80).

Figure 2

Histograms of the Singular Standardised Positive Affect Items. Divided by the Conditions Likert and VAS.



Hypothesis Testing

H1: Participants in the VAS condition will demonstrate significantly higher emotional variability of distinct positive emotions as compared to the Likert condition.

A regression analysis on the linear mixed model of emotional variability showed that there was no statistically significant difference between the two conditions for any of the positive affect items. The values are displayed in Table 3.

H2: Participants in the VAS condition will demonstrate significantly higher emotional instability of distinct positive emotions as compared to the Likert condition.

The linear mixed model for emotional stability also did not generate significant p-values for the predictor variable condition for any of the positive emotions. Meaning that the second hypothesis could also be rejected based on the p-values extracted from Table 3.

Model Results: Linear Fixed-Effect Models

Table 3

Regression Analysis of Linear Fixed-Effect Models for Emotional Variability and Emotional Instability per Positive Affect Item

Model	*Estimate*	*SE*	*df*	*t*	95% CI	p.value
Model 1: Variability PA1	-0.075	0.048	78	-1.563	[-0.171, 0.021]	0.122
Model 2: Variability PA2	-0.031	0.047	78	-0.654	[-0.125, 0.063]	0.515
Model 3: Variability PA3	-0.061	0.047	78	-1.287	[-0.155, 0.033]	0.202
Model 4: Variability PA4	0.009	0.048	78	0.196	[-0.086, 0.104]	0.845
Model 5: Instability PA1	-0.245	0.140	78	-1.750	[-0.523, 0.034]	0.084
Model 6: Instability PA2	-0.116	0.130	78	-0.894	[-0.376, 0.143]	0.374
Model 7: Instability PA3	-0.103	0.121	78	-0.852	[-0.345, 0.138]	0.397
Model 8: Instability PA4	-0.100	0.136	78	-0.737	[-0.371, 0.171]	0.464

Note. N = 80, *SE* = Standard Error, *df* = degrees of freedom, CI = Confidence Interval, Independent Variable = Condition (0 = Likert, 1 =VAS). Also, emotional variability includes

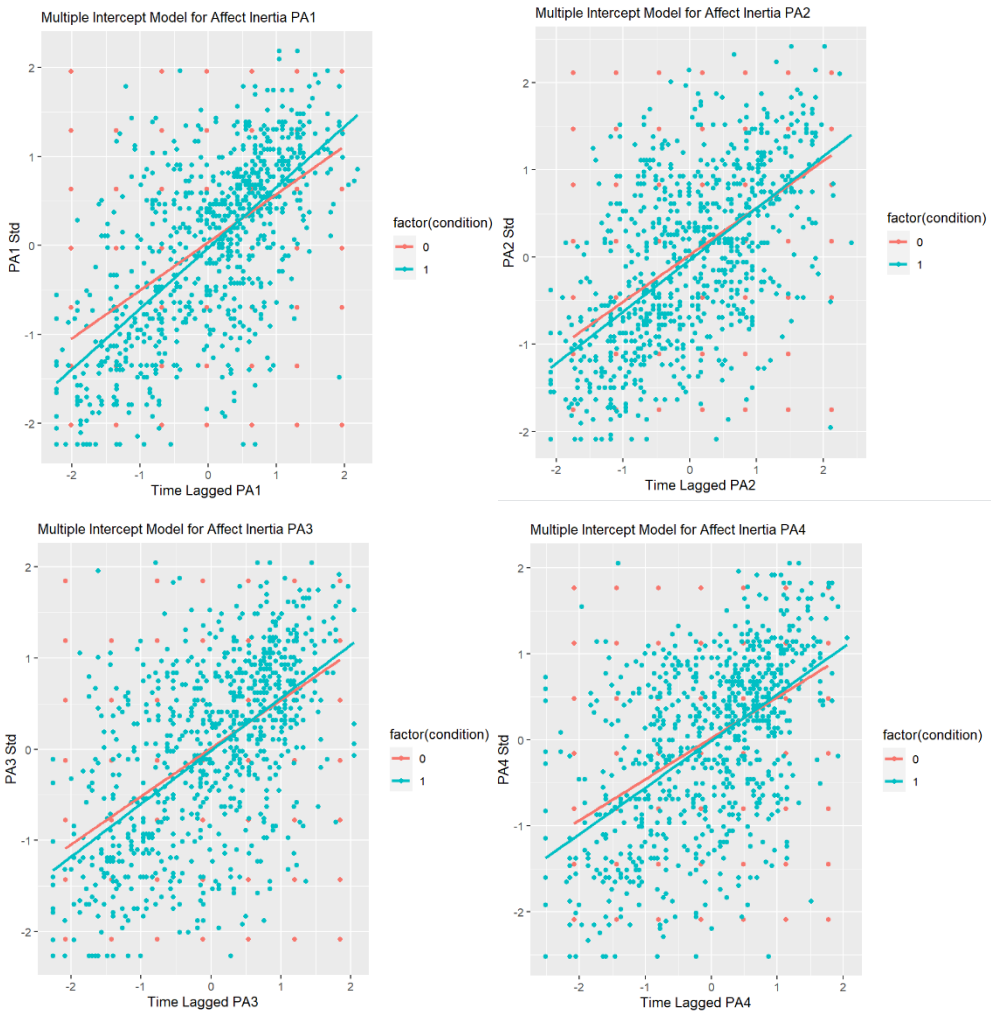
four rows for each positive affect item and thus, includes models 1 to 4 and emotional instability includes models 5 to 8.

H3: Participants in the VAS condition will demonstrate significantly higher emotional inertia of distinct positive emotions as compared to the Likert condition.

The linear mixed model provided partially significant results. The interaction of condition: lagged PA was significant for the first PA item, cheerfulness t = 2.208, p = .027 and the fourth PA item, relaxation, t = 2.087, p = .037.

Figure 3

Linear Mixed Model – Emotional Inertia



Note. LMM contains multiple intercepts and one slope per condition. PA 1: Cheerfulness, PA 2: Enthusiasm, PA 3: Satisfaction, and PA 4: Relaxation

Model Results: Linear Mixed Model

Table 4

Regression Analysis of Linear Mixed Model for Emotional Inertia per Positive Affect Item

Model	Variable	*Estimate*	*SE*	*df*	*t*	95% CI	p.value
Model 1: Inertia PA1	Condition	-0.052	0.088	78	-0.589	[-0.227, 0.124]	0.557
	Lagged PA1	0.381	0.025	1980	14.946	[0.331, 0.431]	0.000
	Lagged PA1*condition	0.093	0.042	1980	2.208	[0.010, 0.176]	0.027
Model 2: Inertia PA2	Condition	-0.054	0.083	78	-0.650	[-0.221, 0.122]	0.518
	Lagged PA2	0.414	0.026	1976	15.830	[0.363, 0.465]	0.000
	Lagged PA2*condition	0.020	0.042	1976	0.492	[0.061, 0.102]	0.623
Model 3: Inertia PA3	Condition	-0.006	0.089	78	-0.069	[0.183, 0.171]	0.945
	Lagged PA3	0.384	0.026	1970	15.063	[0.334, 0.435]	0.000
	Lagged PA3*condition	-0.016	0.042	1970	-0.377	[-0.098, 0.066]	0.706
Model 4: Inertia PA4	Condition	-0.029	0.077	78	-0.373	[-0.182, 0.125]	0.710
	Lagged PA4	0.377	0.026	1962	14.299	[0.325, 0.429]	0.000
	Lagged PA4*condition	0.086	0.041	1962	2.087	0.167]	0.037

Note. N = 80 (level 2), N = 70 (level 1), *SE* = Standard Error, *df* = degrees of freedom, CI = Confidence Interval.

Discussion

Overall, it was hypothesised that the VAS and Likert scales would have different absolute levels for the three affect dynamics on cheerfulness, enthusiasm, satisfaction, and relaxation. Based on theories on cognitive load significantly higher levels of affect dynamics for condition one, VAS, were expected. Even so, the study generated non-significant results for emotional variability and emotional instability, disproving the first two hypotheses. The third hypothesis about emotional inertia was only partially confirmed, indicating significant patterns for an effect by the interaction of condition and a lagged positive affect item. This suggests that the effect of response scales on affect dynamics could be mediated by other variables, for example by interaction effects or by item-related characteristics.

Main Findings

The Effect of Labelling on Emotional Variability and Instability in Positive Affect Items

This research did not find significant evidence to confirm the first or second hypothesis. These checked whether the condition could influence the intensity of the individuals' scores on distinct positive emotions, cheerfulness, enthusiasm, satisfaction and relaxation. Emotions consist of multiple components (Lucas et al., 2009, pp. 139–156), for this study the focus was on how emotions change and persist over time. Therefore, affect dynamics were used to assess what scale can most accurately capture the complexity of each of the four positive emotions. emotional variability and emotional stability.

The hypotheses were based on the assumption, that participants would engage in heuristics to decrease cognitive load. Consequently, participants could exhibit specific response styles, like ERS. It was expected that participants would perceive the continuous VAS scale as not having obvious answer options, causing them to take mental shortcuts, like anchoring (Kieruj & Moors, 2010; Weathers et al., 2005) to reduce necessary cognitive load (Morsanyi & Handley, 2007, Naemi et al., 2009). For Likert, the categories serve as anchors, meaning that responding would be easier and not promote heuristics like ERS (Lucas et al., 2009, pp. 139–156). Based on previous literature, it was assumed that more end-point responses would occur in the VAS condition, meaning more extreme values (Hasson & Arnetz, 2005). The differences in the hypotheses could then have potentially been explained by answers tending towards the vicinity of the highest endpoint. Causing higher values and indicating differences in absolute levels between the conditions. This would also mean, that

for VAS the distribution of responses would be negatively skewed. However, when looking at the histograms (Figure 2), the distributions were positively skewed, also refuting the initial idea. Moreover, the skewness was the same for both conditions, with no apparent differences. This shows disagreement with the findings of Hasson and Arnetz (2005), who found significantly different distributions when comparing VAS scales to Likert scales.

Furthermore, Table 2 displays the relative frequencies of responses per item and for each condition. The table clearly shows that, although the response patterns might vary, most observations were recorded within category five (somewhat agree) for both conditions. There was only one exception for PA2 enthusiasm, for which most observations were recorded in category four (neither agree or disagree). Therefore, none of the conditions triggered a majority of extreme responses. Moreover, the conditions were very similar, except for one item of the Likert condition being even less intense compared to the other items. Also in line with the non-significant results on emotional variability and instability, these results would imply no substantial difference in absolute levels. This would mean, that both scales captured the spectrum and fluctuations of positive emotions, similarly, indicating a similar degree of sensitivity towards these components of emotions. However, other mediating variables, such as sample characteristics cannot yet be excluded.

The Effect of Condition and Lagged Variables on Emotional Inertia in Positive Affect Items

The regression analysis of the linear mixed model provided partially significant results. It showed that for the current state of the positive affect items cheerfulness and relaxation were substantially influenced by the combined effect of the condition and the lagged variable of the current mood, meaning that the condition can be ruled out as the sole predictor. Nevertheless, the significant results suggest that participants that used the VAS scale were more likely to maintain their state of cheerfulness, and relaxation over a longer time period than those of the Likert scale. This could be interpreted that the VAS scale could positively influence the persistence of emotions over time if the differences in absolute levels were explained by an effect of the response scale.

The results also indicated a significant effect of the lagged variable on the dependent variable, which makes sense as it is derived from the existing responses and is a key aspect of inertia. Regarding the fact that VAS could trigger heuristics, this connection, but also the differences between the scales could be explained with ERS. Heuristics are tightly connected to self-control. Hence, it is more likely to trigger heuristic thinking if it was triggered before and because of high cognitive load, this is known as the carry-over effect (Wan & Agrawal,

2011). Thus, the carry-over effect could be a possible mediator (Wan & Agrawal, 2011) that, combined with VAS, triggers different responses, for instance, ERS. Additionally, the anchoring effect could play a role in combination with the carry-over effect (Wan & Agrawal, 2011; Weathers et al., 2005). If the participants chose an anchor point for previous responses this could become a reference point for other responses as well. For example, extreme-endpoint anchors would become consistent reference points and facilitate extreme personal patterns.

This analysis confirms that response scales can have a possible effect on responses. Nevertheless, the third hypothesis was only partially confirmed for two out of four positive affect items. The differences could depend on the nature of emotions. Enthusiasm and satisfaction could be more constant emotions and less susceptible to fluctuation. In some scales, satisfaction is defined as a slowly changing emotion (Diener et al., 2012). This could, in turn, result in lower absolute values. Creating a possible explanation, as to why the relative frequency of enthusiasm was mainly rated in a lower category when compared to others. Other explanations could be that the VAS and Likert scales are both similarly sensitive for measuring the inertia of these two emotions or that sample characteristics could have played a role.

Possible Explanations for Partially Insignificant Results

As already mentioned in the findings of both the linear models and linear mixed model, sample characteristics could have played a substantial role. The results could have been affected by the bias of sample characteristics. Morsanyi and Handley (2007) conclude that heuristic responding, like ERS, is positively related to increases in cognitive burden. This is also directly related to demographic variables, such as age (Morsanyi & Handley, 2007). According to Kieruj and Moors (2010) and Vaerenbergh and Thomas (2013), ERS grew more likely with older age, which is also confirmed by the research of Morsanyi and Handley (2007). Another demographic variable related to ERS was education (Kieruj & Moors, 2010; Vaerenbergh & Thomas, 2013), which could also be directly related to cognitive demand. Meaning that low education could be related to lower IQ and struggles with cognitive ability, leading the participants to use a heuristic system (Vaerenbergh & Thomas, 2013). Kokis et al. (2002) support this, by claiming that substantial cognitive ability is required to override heuristic systems. This was also supported by Naemi et al.'s (2009) claim that ERS is a consequence of simplistic thinking.

For example, a large percentage of participants were students, mediating variables here could be the familiarity with working with scales or higher levels of education which

could mean that less cognitive load was required, reducing the likelihood of ERS (Kieruj & Moors, 2010; and Vaerenbergh & Thomas, 2013). Also, for age, ERS is more likely to occur for younger and older people, however, most were neither young nor old (Vaerenbergh & Thomas, 2013). In addition, gender differences may play a role. The majority was female, and some studies claim that females are more likely to engage in ERS (Crandall, 1973), but according to Marshall and Lee (1998), men are more likely to exhibit ERS for items of affective nature. Thus, maybe the lack of ERS was due to the minority of male participants. This means, that demographics should also be considered in future research and that the effect of personality on response scales should also be considered with similar studies like Naemis et al.'s (2007) investigative research.

Implications

The question of whether the response format influences the results of affect dynamics has already produced many different outcomes. Therefore, choosing a certain scale can be very hard. There are different opinions on choosing appropriate scales, however, the choice likely differs for each situation. According to Henninger et al. (2022), the Likert scale is a reliable choice and easy to administer, which finds agreement by many other researchers (Eisele et al., 2020, Hasson & Arnetz, 2005, Schimmack, 2003). On the other hand, both Eisele et al. (2021, Chapter 4) and Hasson and Arnetz (2005) claim that continuous scales, such as the VAS scale show greater precision. Hasson and Arnetz (2005), even go as far as saying that continuous scales are more sensitive to detecting minor differences, indicating higher levels of responsiveness. All things considered, adjusting the scale to the goal can be very valuable in clinical settings (Hasson & Arnetz, 2005) and depends on the required sensitivity of the scale.

Different objectives could be, detailed looking at the fluctuations and variations of emotions or just the intensity of emotions. As the VAS scale is thought to be more precise, this would be a good choice for detailed observations (Eisele et al., 2021, Chapter 4; Hasson & Arnetz, 2005). The Likert scale could be used for more general levels due to its absolute categories, as well as, because of its easy application (Henninger et al., 2022). All in all, this would mean that the response format matters when examining positive affect dynamics and that the choice should be made whilst considering the goals of the questionnaire and analysis.

Strengths and Limitations

As the ESM design was chosen because of its high validity in a daily life setting (Myin-Germeys & Kuppens, 2021, Chapter 1), this can be viewed as an important strength of this study. The longitudinal character of the study provided real-time data. Comprehensive

analysis showed which variables had an effect. Another important strength of this design was using a randomised control trial. Using RCT in research on methodological choices in ESM is a very unique and strong tool for investigating causality (Hariton & Locascio, 2018).

Regardless, there are still some mentionable limitations. The first one is, that the analysis focused on explaining differences in positive affect dynamics with ERS, however, when looking at the histograms (Figure 2) and relative frequency table (Table 2), it was not clear whether there were even patterns for ERS. This could be improved by tracking response style (Henninger et al., 2022) or looking into the variance of the distributions (Kieruj & Moors, 2010).

In addition, there were two issues regarding the final sample N. An a priori power analysis calculated the relevant sample size to accurately identify possible effects (Faul et al., 2009). However, there was a high number of unexpected dropouts, which might have limited the possibility of finding possible effects and also reduced the generalisability of the sample size and may have resulted in bias because only a specific group of people with specific characteristics may have been included in the final analyses. Furthermore, the sample included mainly female and relatively highly educated young people, limiting the generalisability of the findings.

Another factor could have been the choice of distinct positive emotions. Enthusiasm and satisfaction may be rather constant emotions (Diener et al., 2012), which would make it unnecessary to look into their fluctuations and variations of emotions, as the general tendency might be more time-efficient and still be meaningful.

Future Directions

All in all, this study provided valuable insights into ESM design and for investigating which response scales would be more appropriate for different objectives, like precise or rather simple measurement of positive emotions. The study also found partially significant results. Future research with a similar study design could benefit from a higher sample size. Other ideas for the future could be the analysis and comparison of response scale effects in different study designs and different methodologies. This could also help with finding out more about possible mediating variables on response scale and style. Other designs could be a within-person design, where respondents receive the same questions multiple times but with different response scales in combination with at least a singular personality assessment. Here, mediating variables like personality could be analysed by examining whether specific response styles are stable across different response scales and can be traced back to personal characteristics.

Another possible change could be the content of the study. During this research, the focus lay on cheerfulness, enthusiasm, satisfaction, and relaxation. However, other distinct emotions could also be used to inspect the effect of response scales and possible personal response styles. Also, cognitive load likely plays a mediating role in displaying certain response styles. Finding out more about this relationship could be interesting in future research as well. Tourangeau et al. (2007) found that respondents have problems with paying attention to and working with scales, which can lead to changes in response styles by making random and minor changes to a scale. These changes then promote certain heuristics (Tourangeau et al., 2007). Hence, it could also be interesting to create multiple conditions that differently affect cognitive load. Vaerenbergh and Thomas (2013) suggested that the situational factor is a determining factor for exhibiting specific response styles. These factors could be manipulated, for example by decreasing the time limit for the experimental condition whilst using the same scale, which could trigger more heuristic thinking in the experimental group and thus, extreme response styles.

Conclusion

The question of whether response formats can influence responses to distinct positive emotions has produced unclear results. Regarding the current study, which examined the VAS and the Likert scale's effect on the positive affect dynamics of cheerfulness, enthusiasm, satisfaction and relaxation, there were also, only partially significant results. The study found an interaction between condition and inertia influencing the absolute level of positive affect items, cheerfulness, and relaxation. Indicating that the VAS scale could be more sensitive in detecting variations and changes of emotions, especially over an extended time as in inertia. Nonetheless, the characteristics of the chosen positive emotions could also be relevant for explaining these differences. All in all, future research from all domains, especially mental well-being, would benefit from further analyses.

References

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumgartner, H., & Steenkamp, J. E. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, *38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Crandall, J. E. (1973). Sex Differences in Extreme Response Style: Differences in Frequency of Use of Extreme Positive and Negative Ratings. *Journal of Social Psychology*, *89*(2), 281–293. https://doi.org/10.1080/00224545.1973.9922601

Csíkszentmihályi, M., & Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease*.

Dejonckheere, E., & Erbas, Y. (2021). *The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies.* (I. Myin-Germeys & P. Kuppens, Eds.).

Diener, E., Inglehart, R., & Tay, L. (2012). Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, *112*(3), 497–527. https://doi.org/10.1007/s11205-012-0076-y

Eisele, G., Kasanova, Z., & Houben, M. (2021). *The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies.* (I. Myin-Germeys & P. Kuppens, Eds.).

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (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

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research. *BJOG*, *125*(13), 1716. https://doi.org/10.1111/1471-0528.15199

Hasson, D., & Arnetz, B. (2002). Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements. *International Electronic Journal of Health Education*, *8*, 178–192.

Henninger, M., Plieninger, H., & Meiser, T. (2022). The effect of response formats on response style strength: An experimental comparison. *PsyArXiv*. https://doi.org/10.31234/osf.io/5jxg7

Home. (n.d.). Ethica Data. https://www.ethicadata.com/

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

Kang, H. (2021). Sample size determination and power analysis using the G*Power software. *Journal of Educational Evaluation for Health Professions*, *18*, 17. https://doi.org/10.3352/jeehp.2021.18.17

Kieruj, N. D., & Moors, G. (2010). Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research*, *22*(3), 320–342. https://doi.org/10.1093/ijpor/edq001

Kokis, J., Macpherson, R., Toplak, M., West, R., & Stanovich, K. (2000). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *J. Experimental Child Psychology*, *83*, 26–52.

Krosnick, J. (1989). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *John Wiley & Sons, Ltd.*

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Lafit, G. (2021). *The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies.* (I. Myin-Germeys & P. Kuppens, Eds.).

Lucas, R. E., Diener, E., & Seidlitz, L. (2009). Assessing Well-Being. *Social Indicators Research Series*, *39*, 139–156. https://doi.org/10.1093/ijpor/edq001

Maes, I. H. L., Mertens, L., Poppe, L., Crombez, G., Vetrovsky, T., & Van Dyck, D. (2022). The variability of emotions, physical complaints, intention, and self-efficacy: an ecological momentary assessment study in older adults. *PeerJ*, *10*, e13234. https://doi.org/10.7717/peerj.13234

Marshall, R., & Lee, C. (1998). A Cross-Cultural, Between-Gender Study of Extreme Response Style. In B. Englis & A. Olofsson (Eds.), *E - European Advances in Consumer* (Vol. 3, pp. 90–95). https://www.acrwebsite.org/volumes/11158/volumes/e03/E-03/full

Morsanyi, K., & Handley, S. (2007). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *J. Of Experimental Child Psychology*, *99*, 18–36.

Myin-Germeys, I., & Kuppens, P. (2021). *The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies.*

Naemi, B., Beal, D., & Payne, S. (2007). Personality Predictors of Extreme Response Style. *Journal of Personality*, *77*(1).

Pirla, S. F., Taquet, M., & Quoidbach, J. (2022). Measuring affect dynamics: An empirical framework. *Behavior Research Methods*, *55*(1), 285–300. https://doi.org/10.3758/s13428-022-01829-0

R: The R Project for Statistical Computing. (n.d.). https://www.r-project.org/

Schimmack, U. (2003). Affect Measurement in Experience Sampling Research. In *Journal of Happiness Studies* (Issue 4, pp. 79–106).

Sealed Envelope | Randomisation (randomization) and online databases for clinical trials. (n.d.). https://www.sealedenvelope.com/

Test subject pool BMS. (n.d.). https://utwente.sona-systems.com/

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Color, Labels, and Interpretive Heuristics for Response Scales. *Public Opinion Quarterly*, *71*(1), 91–112. https://doi.org/10.1093/poq/nfl046

Uy, M. A., Foo, M., & Aguinis, H. (2010). Using Experience Sampling Methodology to Advance Entrepreneurship Theory and Research. *Organizational Research Methods*, *13*(1), 31–54. https://doi.org/10.1177/1094428109334977

Van Vaerenbergh, Y., & Thomas, T. (2013). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. https://doi.org/10.1093/ijpor/eds021

Verhagen, S., Hasmi, L., Drukker, M., Van Os, J., & Delespaul, P. (2015). Use of the experience sampling method in the context of clinical trials. *Evid Based Mental Health*, *19*(3).

Wan, E. W., & Agrawal, N. (2011). Carryover Effects of Self-Control on Decision Making: A Construal-Level Perspective. *Journal of Consumer Research*, *38*(1), 199–214. https://doi.org/10.1086/658471

Watson, D., & Naragon, K. (2009). *Positive affectivity: The disposition to experience positive emotional states*. Oxford University Press.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Weijters, B., Geuens, M., & Schillewaert, N. (2009). The Stability of Individual Response Styles. *American Psychological Association*, *15*(1), 96–110. https://doi.org/10.1037/a0018721

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer Science & Business Media.

Appendix A

Table 5

Daily Study Schedule Including Type of Questionnaire, Variables of Interest, Time Points and Expiration Time

Day	Questionnaire	Variables of interest	Time points	Expiration time
1	Baseline	All	At the beginning of the study	No
1-7	ESM	PA1: cheerfulness PA2: enthusiasm PA3: satisfaction PA4: relaxation	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	After 15 minutes

Appendix B

Table 6

Categorisation of VAS Scale Ranging from 0 to 100

Category	1	2	3	4	5	6	7
Range	0-14	15-28	29-42	43-56	57-70	71-84	85-100
Number of Values	15	14	14	14	14	14	16

Appendix C

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.

A 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 into 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 (e-mail)

Nick Delventhal (e-mail)

Jan Derksen (e-mail)

Gina Haccou (e-mail)

Samuel Pietsch (e-mail)

Aleksandra Popovic (e-mail)

Lea Staudigel (e-mail)

Nina Zarrin Tigh (e-mail)

Supervisors

Jannis Kraiss

(e-mail)

Thomas Vaessen

(e-mail)

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

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

Appendix D

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 (e-mail)

Nick Delventhal (e-mail)

Jan Derksen (e-mail)

Gina Haccou (e-mail)

Samuel Pietsch (e-mail)

Aleksandra Popovic (e-mail)

Lea Staudigel (e-mail)

Nina Zarrin Tigh (e-mail)

Supervisors

Jannis Kraiss

(e-mail)

Thomas Vaessen

(e-mail)

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

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

Appendix E

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 (e-mail).

Students

Simon Brune (e-mail)

Nick Delventhal (e-mail)

Jan Derksen (e-mail)

Gina Haccou (e-mail)

Samuel Pietsch (e-mail)

Aleksandra Popovic (e-mail)

Lea Staudigel (e-mail)

Nina Zarrin Tigh (e-mail)

Supervisors

Jannis Kraiss

(e-mail)

Thomas Vaessen

(e-mail)