



HAL
open science

Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments

Tomu Tominaga, Shuhei Yamamoto, Takeshi Kurashima, Hiroyuki Toda

► **To cite this version:**

Tomu Tominaga, Shuhei Yamamoto, Takeshi Kurashima, Hiroyuki Toda. Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments. 18th IFIP Conference on Human-Computer Interaction (INTERACT), Aug 2021, Bari, Italy. pp.3-22, 10.1007/978-3-030-85607-6_1 . hal-04291265

HAL Id: hal-04291265

<https://inria.hal.science/hal-04291265v1>

Submitted on 17 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments

Tomu Tominaga, Shuhei Yamamoto, Takeshi Kurashima, and Hiroyuki Toda

NTT Service Evolution Laboratories, NTT Corporation, Japan
{tomu.tominaga.hm,shuhei.yamamoto.ea,takeshi.kurashima.uf,hiroyuki.toda.xb}
@hco.ntt.co.jp

Abstract. Response rates to Ecological Momentary Assessment (EMA) questions vary among individuals, which inevitably results in a disproportionate distribution of the volume of EMA data to personal characteristics. Previous research has explored how response rates are influenced by subjects' characteristics; however, few studies focused on temporal patterns of responses to EMA questions. In this study, we examined the relationship of personal characteristics with temporal response patterns by using StudentLife dataset. From this dataset, we adopted scales of personality traits (neuroticism, extraversion, openness, agreeableness, conscientiousness) and mental status (depression, stress, loneliness, life satisfaction) as subjects' personal characteristics, and analyzed how these characteristics relate to the subjects' temporal response patterns for EMA questions. Primary results of our analyses using regression models and latent growth curve models indicated that the subjects who reported high stress levels had a higher total number of EMA responses, but the number of responses from these subjects tended to decrease strongly throughout the experiment, especially after a stressful event such as the midterm examination. We also found that the subjects who reported low extraversion had a weak tendency to decrease the number of EMA responses throughout the experiment, but the number of EMA responses of these subjects decreased strongly after a social event such as the school festival. Based on our results, we discussed how the data collected with EMA studies are biased not only in terms of personal characteristics but also with timings.

Keywords: Ecological Momentary Assessment · StudentLife · Human Factors

1 Introduction

Ecological Momentary Assessment (EMA), also known as Experience Sampling Method, is a longitudinal research methodology that asks study subjects in natural environments to continuously report their current thoughts, feelings, behaviors, or experiences in real-time [29]. Unlike traditional questionnaire surveys in a lab setting, EMA captures within-subject fluctuations of perceived states

(e.g., emotion) while improving ecological validity and reducing recall bias. Thus, the EMA data have been widely used for various applications (medical validations [32, 36], mood predictions [22, 27, 28], and so on) as ground truth data of the time-series changes in the values of people’s subjective evaluations of their states.

In order to construct reliable EMA datasets, it is important for EMA to obtain a large number of responses uniformly covering various conditions (e.g., subjects, situations, etc.); however, past studies have shown that it is difficult to collect responses from subjects with specific characteristics [37, 30, 23, 19]. For example, subjects with mental illness [19] or young subjects [37, 20] were found to have low response rates. A response rate refers to the number or ratio of responses to the EMA questions from a subject within some duration. These previous studies show the conditions that bias EMA data against personal characteristics. Describing such conditions is a fundamental issue in EMA research, because it enables us to correctly evaluate and discuss the limitations and reliability of findings drawn from EMA data.

Although prior work has examined the relationships between subjects’ personal characteristics and their EMA response rates, the effect of personal characteristics on temporal changes in response rates (e.g., slopes of EMA response frequency over time) has not been well studied yet. In our assumption, when stressed people experience a tense or anxious event, they may spend less time answering EMA questions because they are focused on dealing with that event. Examining this issue has the fundamental contribution of describing in detail the conditions that cause bias in EMA data, not only in terms of personal characteristics but also in terms of timing. If we understand when subjects with certain characteristics will not respond, it becomes possible to design experiments that encourage those subjects to respond at those times. This will contribute to the original purpose of EMA, which is to collect data uniformly under various conditions.

To understand how EMA data can be biased by the subjects’ personal characteristics and time, this paper analyzes the relationship between subjects’ personal characteristics and their temporal patterns of EMA responses. For the analysis, we used four metrics to characterize the temporal response patterns: (1) *continuity*: how long a subject kept responding to the EMA questions, (2) *frequency*: how many EMA questions a subject responded to per day, (3) *overall slope*: how the number of EMA responses per day for a subject increase/decrease over time, and (4) *period-separated slopes*: how the increase/decrease in EMA response numbers per day over time for a subject vary between time periods. Therefore, this paper addresses the following research questions.

RQ1 *What kind of personal characteristics affect continuity?*

RQ2 *What kind of personal characteristics affect frequency?*

RQ3 *What kind of personal characteristics affect overall slope?*

RQ4 *What kind of personal characteristics affect period-separated slopes?*

The four metrics extracted from the daily frequency of EMA responses are shown in Figure 1. We measured *continuity* as the difference between the first

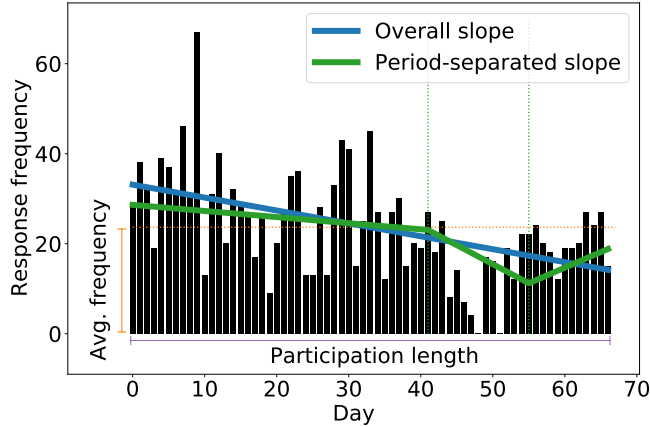


Fig. 1. A sample case of a subject: relationship of *continuity* (participation length), *frequency* (average of daily frequency), *overall slope*, and *period-separated slopes* with daily frequency of EMA responses (the bar colored in black). For period-separated slopes, three periods are given: week #1–#6 (day 0–41), week #6–#8 (day 42–55), and week #8–#10 (day 56–67). Note that overall slope is introduced from daily frequency, and that period-separated slopes are calculated from daily frequency per week.

and the last response time to the EMA in order to assess how long subjects keep participating in the experiment, and *frequency* as the average number of EMA responses per day to elucidate how active subjects were in making EMA responses. We scaled *overall slope* as a linear trend component of daily frequency of EMA responses throughout the experiment to understand the overall trend in the number of responses to EMA over time. Moreover, we quantified *period-separated slopes* as linear trend components of week-level daily frequency of EMA responses in each of the periods stipulated for discovering how slopes differ between the periods (we set the periods by considering the events of the university (e.g., the midterm examination and the school festival) and the validity of the models used for the analysis).

For our analyses, we used StudentLife dataset [38]. StudentLife [38] followed 48 students at Dartmouth University over 10 weeks to collect responses for EMA questions, answers to questionnaire surveys, and passively-sensed data (e.g., wifi logs). The dataset also contains information about the school events including the examination period and the school festival at the university. In this paper, we used all of the EMA data to measure response rates of subjects. We also used the questionnaire data of personality traits (neuroticism, extraversion, openness, conscientiousness, and agreeableness) and mental status (depression, perceived stress, loneliness, and flourishing) as personal characteristics.

To address RQ1, RQ2, and RQ3, we run a series of regression analyses to explain a participation length (i.e., continuity), an average of daily frequency of

EMA responses (i.e., frequency), and a linear trend component of daily frequency of EMA responses (i.e., overall slope) from personal characteristics (i.e., scales of personality traits and mental status). Concerning RQ4, we used week-level daily frequency of EMA responses to account the influence of the midterm examination and the school festival, which are week-level events. We then adopted latent growth curve models to define the periods so that the trend components in each period (i.e., period-separated slopes) is well fitted to the week-level daily frequency, and to examine how the period-separated slopes are explained by personal characteristics.

Primary results indicated that subjects who reported high stress levels had a higher total number of EMA responses, but those subjects tended to strongly decrease the number of responses throughout the experiment, especially after the midterm examination. We also found that the subjects who reported low extraversion exhibited a slight decrease in the number of EMA responses throughout the experiment, but the number of EMA responses of these subjects decreased strongly after the school festival. Based on these results, we discussed the disproportion of EMA data in terms of subjects and timings, and described implications for understanding the limitations of EMA studies.

2 Related Work

EMA has been recognized as a powerful method for recording people’s subjective states such as emotions, thoughts, and experiences over time. The nature of EMA is that the experimenter intermittently sends questions to the subjects through portable devices, and the subjects answer the questions in real time in the situation. Therefore, EMA has less recall bias and higher ecological validity than conventional questionnaire surveys. The data collected with EMA have been used for various purposes such as identifying behavioral patterns [38, 11, 26], predicting life outcomes [40, 8], and understanding human relationships [10, 13].

As shown above, many studies have been developed on the basis of EMA data, but it has been pointed out that EMA data contain various biases [19]. There are three primary types of factors behind the bias: subject factors, contextual factors, and design factors. Subject factors refer to the characteristics of the subject of the EMA experiment, including demographic attributes [37, 20], psychological state [30, 37], and mental health status [37, 14]. For example, male subjects [37, 20], young subjects [37, 20], subjects with high alcohol consumption [30], subjects with negative emotions [30], and subjects with psychotic disorder [30, 37, 19] are known to have lower response rates. Vachon et al.’s meta-analysis [37] also found that including more males and subjects with psychotic disorders yielded lower response rates. These findings indicate that EMA data is biased by such subject-dependent factors.

Contextual factors, which are properties of the environment in which the experiment is carried out (e.g., time [5, 3], place [5], daily occurrences [3], etc.), also create a bias. According to an investigation of temporal and spatial distri-

butions of EMA responses, subjects were more likely to respond in the middle of the night than in the morning, and more likely to respond in places such as hospitals and restaurants rather than on the move [5]. Moreover, van Berkel et al. [3] examined how the probability of individual EMA questions being answered is influenced by contextual factors (e.g., number of notifications received in the last 15 minutes), and found that high-compliance subjects were likely to respond to EMA questions if they received recent notifications frequently. In this way, environmental properties can distort the distributions of EMA data.

Design factors refer to the characteristics of the experimental design of EMA, including the incentives [37, 20], frequency of questions [30], and methods of data sampling [37]. The experimental conditions under which EMA is carried out can lead to biased EMA data. For example, more EMA responses are collected if EMA questions are sent when the phone screen unlock event is detected than if they are sent randomly or periodically [4]. However, Lathia et al. [17] pointed out that controlling the timing of EMA questions by triggers based on passively-sensed data also causes a bias in that the number or values of EMA answers vary unevenly over time. A previous study that conducted sampling based on information captured by accelerometer, GPS, microphone sensor, smartphone screen usage status, and telephone/email, found that 25% of all responses were recorded between 10:00 and 13:00 [17]. This is different from the other result [5] which found that the response rate is highest in the middle of the night when EMA questions are sent randomly with respect to time. This discrepancy indicates that bias in EMA data can be caused by the experimental design.

These previous works on EMA data bias examined the number (probability) of EMA responses of subjects; however, trajectories of response rates have seldom been analyzed. Additionally, the effect of subject factors on the trajectories have not been well studied. In other words, the issue of which subjects' data are missing at which time has not been adequately examined. By clarifying this, we will be able to understand how EMA data is missing and biased, not only in terms of personal characteristics, but also in terms of time. These findings can also support optimization of EMA experiment design to determine which subjects should be asked to respond and when with the goal of securing more unbiased responses. Therefore, this study aims to examine how personal characteristics relate to the temporal variation of the response rates and to derive new knowledge about the bias of EMA data against time.

3 Dataset

To accomplish our research goals, we used the StudentLife dataset [38]. StudentLife [38] conducted traditional questionnaire surveys to identify personal characteristics, adopted EMA to observe ongoing situations or experiences, and used built-in sensors of smartphones to measure activities automatically. The obtained data was published as the StudentLife dataset for third parties. From the dataset, we used the records that paired questionnaire survey answers with

EMA responses to assess how personal characteristics are associated with EMA responses. Details of these records are explained below.

3.1 EMA Data

The StudentLife dataset has subjects' responses to EMA questions concerning emotions, stress, sleep, and so on. We used all of the responses to introduce continuity, frequency, overall slope, and period-separated slopes. In total, 20,027 EMA responses were used for our analyses.

3.2 Personal Characteristics

This study describes personal characteristics with personality traits and mental status. The scales of personality traits follow the Big Five personality traits [21], defining personality from five dimensions: *neuroticism* (sensitive / nervous vs. resilient / confident), *extraversion* (outgoing / energetic vs. solitary / reserved), *openness* (inventive / curious vs. consistent / cautious), *conscientiousness* (efficient / organized vs. extravagant / careless), and *agreeableness* (friendly / compassionate vs. challenging / callous) [24]. The scales of depression, stress, loneliness, and life satisfaction were, respectively, measured using PHQ-9 [15], Perceived Stress Scale [6], UCLA Loneliness Scale [25], and Flourishing Scale [9]. In the StudentLife project, these surveys were conducted twice: at the beginning and the end. We used the scales obtained at the beginning (pre-survey).

For our analysis, we excluded subjects who did not answer all questionnaire items; 42 subjects remained. Table 1 shows the descriptive statistics of the measured values of the 42 subjects.

4 Analysis

We conducted a series of regression analyses for RQ1, RQ2, and RQ3, assessing how personal characteristics influence continuity, frequency, and overall slope. We also used latent growth curve models to examine how personal characteristics affect period-separated slopes. These procedures are explained below.

4.1 Personal Characteristics vs. Continuity, Frequency, & Overall Slope

In the series of regression analyses, continuity, frequency, and overall slope were used as objective variables, while personal characteristics are introduced as explanatory variables. The explanatory variables were standardized so that the mean is 0 and the standard deviation is 1. We took the partial regression coefficients yielded by these analyses as the effects of personal characteristics on the continuity, frequency, and overall slope of EMA responses. We considered the statistically significant partial regression coefficients in discussing the impact of personal characteristics.

Table 1. Left: descriptive statistics of scales of personal characteristics. Right: the number of questions (#) and the scales of answer options (answer scales) of each questionnaire. Concerning depression scale, the score ranges from 0 to 27 and is transferred into five levels of severity: 1. minimal (0-4), 2. minor (5-9), 3. moderate (10-14), 4. moderately severe (15-19), and 5. severe (20-27). We used the levels in our analyses.

Measure	Mean	S.D.	Min.	1st Q.	Med.	3rd Q.	Max.	#	Scales
Mental status									
Depression scale	2.25	0.98	1	2.00	2.00	3.00	5	9	0 – 3 ^a
Perceived stress scale	20.60	3.99	9	18.75	20.00	24.00	28	10	0 – 4 ^b
Flourishing scale	41.20	10.90	25	32.00	39.50	50.25	64	8	1 – 7 ^c
Loneliness scale	42.40	7.87	16	37.00	45.00	48.00	54	20	1 – 4 ^d
Personality traits									
Neuroticism	23.78	5.63	14	20.00	23.00	27.25	36	8	1 – 5 ^e
Extraversion	23.40	6.33	13	18.75	23.00	27.25	36	8	1 – 5 ^e
Openness	35.88	4.85	25	33.00	36.00	38.25	46	10	1 – 5 ^e
Agreeableness	32.83	5.62	12	30.00	34.00	37.00	41	9	1 – 5 ^e
Conscientiousness	30.85	6.45	17	26.00	31.00	34.25	43	9	1 – 5 ^e

^anot at all – very often, ^bnever – very often, ^cstrongly disagree – strongly agree
^dnever – often, ^edisagree strongly – agree strongly

4.2 Personal Characteristics vs. Period-separated Slopes

To assess how personal characteristics affect the period-separated slopes, we adopted latent growth curve (LGC) models. In this analysis, we used the week-level daily frequency of EMA responses. LGC models enable us to examine normative and individual-level changes in EMA responses over the weeks covered by the EMA experiment. In the following part, we firstly described LGC models. We then explained how to use LGC models for assessing the normative and individual-level changes.

Latent Growth Curve (LGC) Models LGC modeling is a statistical technique based on the structural equation modeling framework to estimate growth trajectories of observed data. There are two types of the estimated variables: intercept and slopes. Intercept refers to the initial value of the growth trajectory. Slopes represent the linear trends of the growth trajectory. To represent the observed data, LGC models introduce several latent variables, which are equivalent to the intercept and slopes of the observed data.

To capture the trajectory more finely, piecewise LGC modeling is adapted. This approach separates the observation duration into multiple periods and fits linear models to the observed data in each period. The separation approach can be based on visual inspection or interpretability of results. For example, if the trajectory explicitly shows different trends before and after at a time point or if some influential events occur at multiple time points, the observation duration can be separated by the point(s).

According to the separation, LGC models automatically determine the number of latent variables. If the duration is separated into two periods, the number

of the latent variables is three: the intercept and the slopes in each period. When inferring the latent variables, LGC models generally assume that values of the observation data grow or decline linearly. Based on this assumption, LGC models determine effects of the slopes on the observation data in the corresponding periods.

Afterward, LGC models infer the latent variables so that the models fit the observed data. Overall model fit is measured by RMSEA (Root Mean Square Error of Approximation) [31], CFI (Comparative Fit Index) [2], and TLI (Tucker-Lewis Index) [35].

LGC Models for Normative Changes To model normative dynamics of EMA responses, we used piecewise LGC models. In this study, we separated the observation duration by week #6 and #8 into three periods: period I (week #1-#6), period II (week #6-#8), and period III (week #8-#10). We then adopted the LGC model according to this separation, leading the model to infer four latent variables: the initial status at week #1 (intercept) and the slopes in the period I, II, and III. There are two reasons for this separation.

First, we hypothesized that the way of responding to EMA questions is influenced by the midterm examination and the school festival, as indicated by prior work [11, 39]. The examination was conducted from week #3 to #5 and the festival was held during week #7. To capture differences in EMA responses before and after these events, we separated the duration by week #6 and #8. The scheme of the separation enables us to reasonably interpret our results because we can discuss how slopes are different according to such influential events.

Second, we confirmed that the LGC model separating the duration by week #6 and #8 fits the observed data the best. Other than the #6-and-#8 separation, we have candidates of separation by 1 knot (e.g., separating by week #4), 2 knots (e.g., separating by week #5 and #7), or 3 knots (e.g., separating by week #3, #6, and #8). To avoid over-fitting, we excluded the strategies of 3-knots separations. For models based on 1-knot and 2-knots separations, we calculated the model fit measures. As a result, the model separating the duration by week #6 and #8 was the best for all model fit measures among those models.

For the above reasons, we adopted the LGC model separating the observation duration by week #6 and #8. For simplicity, the LGC model for normative changes of EMA responses is called unconditional model later in this paper.

LGC Models for Individual-level Changes To build LGC models for individual-level change, we introduced the factor of interest (personal characteristics here) into the unconditional model. LGC models can use the factor of interest to explain variances in each latent variable with regression. This model is called the conditional model.

Our analyses considered nine personal characteristics, thus we built nine kinds of conditional models. As for the unconditional model, the model fit of the conditional model is calculated using RMSEA, CFI, and TLI.

Table 2. Regression models explaining (1) continuity, (2) frequency, and (3) overall slope of EMA responses from personal characteristics. Note that β is a partial regression coefficient, S.E. is standard error, and p is p-value (**... $p < 0.001$, **... $p < 0.01$, *... $p < 0.05$).

Measure	Continuity			Frequency			Overall slope		
	β	S.E.	p	β	S.E.	p	β	S.E.	p
(Intercept)	55.91	2.83	***	436.65	47.82	***	-0.120	0.012	***
Mental status									
Depression scale	-2.56	3.43		-107.60	57.94		-0.004	0.014	
Perceived stress scale	4.82	3.37		119.23	57.06	*	-0.033	0.015	*
Flourishing scale	1.71	3.58		15.82	60.57		0.013	0.015	
Loneliness scale	1.03	4.21		77.63	71.15		-0.025	0.018	
Personality traits									
Neuroticism	0.20	4.28		55.25	72.36		-0.011	0.018	
Extraversion	-1.74	3.73		11.06	63.01		-0.035	0.016	*
Openness	1.04	3.18		-3.63	53.85		0.026	0.014	
Agreeableness	-3.75	3.96		45.98	67.05		-0.001	0.016	
Conscientiousness	1.25	3.49		25.07	59.08		-0.034	0.015	*

5 Results

5.1 Personal Characteristics vs. Continuity, Frequency, & Overall Slope

Table 2 shows the partial regression coefficients of personal characteristics for continuity, frequency, and overall slope. The intercept in the models represents values of the objective variables of a subject whose all explanatory variables are 0 (i.e., average). Thus, on average, the subjects made 436.65 EMA responses, participated in the experiment for 55.91 days, and reduced the number of responses by 0.12 per day. Betas mean the effect of personal characteristics on the continuity, frequency, and overall slope of EMA responses.

We found a significantly positive coefficient of the perceived stress scale for the frequency of responses to EMA questions ($\beta = 119.23^*$). This means that subjects with higher perceived stress make more responses in total than those with lower perceived stress. Furthermore, we found that perceived stress scale, extraversion, and conscientiousness have significantly negative effects on the overall slope of the daily frequency of EMA responses ($\beta = -0.033^*$, -0.035^* , -0.034^*). These results indicated that the higher the subject's perceived stress, extraversion, and conscientiousness, the stronger the decrease in EMA frequency for the total duration. We did not find significant coefficients of personal characteristics for continuity, implying that personal characteristics were not effective in explaining variances of the participation length.

5.2 Personal Characteristics vs. Period-separated Slopes

Table 3 shows results obtained from LGC models. The unconditional model

Table 3. Unconditional and conditional latent growth curve models for EMA responses (note that all predictors are scaled so that the mean is 1 and standard deviation is 0). M_i is the mean value of the latent variables (intercept, slopes in period I, II, and III) of the unconditional model. V_i represents the variance of the latent variables of the unconditional model. β and S.E. mean, respectively, the partial regression coefficient and its standard error of predictors in the conditional models (***... $p < 0.001$, **... $p < 0.01$, *... $p < 0.05$).

Model/Predictor	Model Fit Measures			Intercept			Slope in period I			Slope in period II			Slope in period III						
	RMSEA	CFI	TLI	M_i/β_0	$V_i/S.E.$		M_i/β_1	$V_i/S.E.$		M_i/β_2	$V_i/S.E.$		M_i/β_3	$V_i/S.E.$					
Unconditional model	0.149	0.923	0.915	10.82	***	25.79	***	-0.96	***	0.61	**	-0.84	**	2.03	**	-0.79	**	2.26	**
Mental status																			
Depression scale	0.140	0.922	0.909	0.14		0.90		-0.16		0.15		-0.25		0.26		0.25		0.27	
Perceived stress scale	0.142	0.921	0.908	1.88	*	0.85		-0.04		0.15		-0.68	**	0.24		0.26		0.27	
Flourishing scale	0.142	0.920	0.906	-0.54		0.89		0.18		0.15		0.20		0.26		-0.13		0.27	
Loneliness scale	0.139	0.923	0.910	1.50		0.86		-0.29		0.15		0.16		0.26		-0.14		0.27	
Personality traits																			
Neuroticism	0.146	0.916	0.901	0.71		0.89		0.02		0.15		-0.11		0.26		-0.12		0.27	
Extraversion	0.146	0.916	0.902	0.13		0.90		-0.27		0.15		0.03		0.26		0.52	*	0.26	
Openness	0.138	0.923	0.910	-0.29		0.89		0.11		0.15		0.17		0.26		-0.10		0.27	
Agreeableness	0.145	0.918	0.904	-1.48		0.86		0.31	*	0.15		0.00		0.26		0.33		0.26	
Conscientiousness	0.151	0.911	0.896	1.11		0.88		-0.19		0.15		-0.17		0.26		-0.24		0.27	

shows significantly negative slopes for all periods ($M. = -0.96^{***}$ (period I), -0.84^{**} (period II), -0.79^{**} (period III)). This is consistent with the result that the overall slope of daily EMA frequency is negative, which was obtained in answering RQ3. Moreover, the variances in the intercept and slopes in each period were significantly large ($V. = 25.79^{***}$ (intercept), 0.61^{**} (period I), 2.03^{**} (period II), 2.26^{**} (period III)). These results mean that overall, the frequency of EMA responses decreased over time, but the values of initial status (i.e., intercept) and the slopes significantly varied among the subjects.

The conditional models can explain the variations of the initial values and the slopes among the subjects. The perceived stress scale was found to have a positive effect on the initial value ($\beta_0 = 1.88^*$) and a negative effect on the slope in period II ($\beta_2 = -0.68^{**}$). The positive effect is consistent with the result that subjects with higher perceived stress respond to EMA questions more frequently, which was obtained in answering RQ2. The negative effect means that the higher the perceived stress, the more strongly the EMA response rate decreased after the end of the examination period.

We also found that extraversion had a positive effect on the slope in period III ($\beta_3 = 0.52^*$). It means that the lower the extraversion (i.e., the higher introversion), the more strongly the EMA response rate decreased after the school festival.

Moreover, our results showed that agreeableness had a positive effect on the slope in period I ($\beta_1 = 0.31^*$). It indicates that, in the initial duration of the experiment, subjects with higher agreeableness decreased the EMA response rates more strongly.

6 Discussion

6.1 Influential Factors

As indicated by our results, perceived stress, extraversion, conscientiousness, and agreeableness are influential factors for EMA response rates. Here, we discussed why these factors were significant.

Effects of Perceived Stress Indicated by our results, subjects with higher perceived stress entered EMA responses at higher levels of daily frequency. According to the previous research [18], stressful individuals prefer disclosing themselves to machines (e.g., virtual agents) rather than to real people. In the context of EMA studies, answering EMA questions can be an activity of disclosing themselves to machines such as entering their emotions with the portable devices. Therefore, the EMA response rates of subjects with higher perceived stress were generally higher.

Subjects with higher perceived stress were also found to have stronger negative overall slopes. As mentioned above, such subjects tended to have a larger volume of EMA responses. This corresponds to a large amount of decrease margin, which may simply yield stronger decreasing trends.

The slope of response rates of subjects with higher perceived stress was found to be strongly negative particularly after the examination period. As discussed in prior work [11], the midterm examination might create extreme pressure for the students. Such stress often makes people disclose themselves to machines than to real people [18]. Considering this background, subjects with higher perceived stress may have disclosed themselves less frequently in the EMA questions after the examination, because they were released from their stress imposed by the tense event.

Effects of Extraversion (Introversion) We found that subjects with higher extraversion showed a stronger negative overall slopes. It was also seen that subjects with higher extraversion showed a stronger positive slope after the school festival (i.e., in period III). These results seem to be contradictory. We start by explaining how to interpret these results.

As shown in Table 3, the period-separated slopes of the conditional model with extraversion in period I, II, and III are -0.27 , 0.03 , and 0.52 . From these values, we can derive that an overall effect of extraversion on the slope at week level is -0.25 ($-0.27 \cdot 5 + 0.03 \cdot 2 + 0.52 \cdot 2$). As shown in Table 2, the effect of extraversion on the overall slope at a day level is -0.035 . Both are consistently negative. Therefore, extraversion was found to have a significantly positive effect on the slope in period III; however, it totally has a negative effect considering the period-separated slopes and the duration of each period.

We here explain why subjects with lower extraversion (i.e., higher introversion) showed a stronger negative slope in period III. Introverted people prefer reflecting on their internal aspects [1, 34]. Answering EMA questions such as telling their mood or behaviors can be an activity indicative of reflection. According to the prior work [33], such reflection needs their psychological energy. However, the psychological energy is limited: it is consumed by social interactions [12]. Considering this, we assumed that introverted subjects depleted their psychological energy through social interaction such as the school festival, which led to a decrease in the EMA response rate.

To see whether introverted subjects engaged in social interaction, we examined the duration of subjects' daily conversations. The conversation durations were estimated by audio classification of microphone data [22, 16]. They are stored in the StudentLife dataset. Following the prior work [11], we used the data as indicators of social activities.

Figure 2 shows the relationship between the conversation duration (the y-axis) and the extraversion (the x-axis) for each period. Extraversion and daily conversation time show slightly positive correlations in period I and III, but a slightly negative correlation in period II. Considering that highly extroverted people prefer social interaction, we can expect to see the same trend in period II as in period I and III. Contrary to this expectation, the association in period II shows a negative correlation. This means that the introverted subjects engaged in social interaction more than expected. However, it should be noted that these

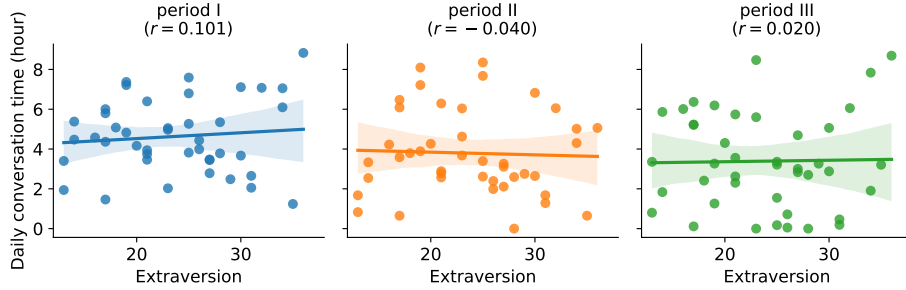


Fig. 2. Extraversion and daily averages of conversation time in each period. Note that each subject has three points (conversation time in period I, II and III). r represents correlation coefficients between extraversion and conversation time in each period.

correlations are weak and not statistically significant. Therefore, our assumption was not completely supported, though it was confirmed in a weak trend.

Effects of Conscientiousness Subjects with higher conscientiousness were found to have stronger negative slopes of the frequency of EMA responses, but conscientiousness did not have any significant effect on continuity and frequency. In general, people with high conscientiousness are known to show self-discipline, act dutifully, and aim for achievement [7]. Thus, subjects with high conscientiousness are expected to make many responses continuously, which is inconsistent with our results.

To understand this, we extracted top-5 subjects and bottom-5 subjects in terms of conscientiousness, and compared their week-level daily frequencies. Figure 3 shows the week-level daily frequency of these subjects. The daily frequency at week #1 of the top-5 subjects is higher than that of the bottom-5 subjects. Considering that continuity was not associated with conscientiousness, this means that subjects with a conscientious personality entered EMA responses in the early stage of the experiment, but not in the following period.

Moreover, the top-5 subjects showed a strong decrease in the EMA frequency from week #5 to #6, which is included in the examination period. After the event, the daily frequency of these subjects also gradually decreased. According to the previous work, the subjects with higher conscientiousness had higher GPA scores [39]. This implies that these subjects gave more weight on the midterm examination than the EMA experiment. This might lead them to make fewer responses around this event.

To sum up, the highly conscientious subjects made more EMA responses in the early stage of the experiment but not in the following period. They also avoided answering EMA questions during (and after) a part of the examination period because they concentrated on the examination. As a result, conscientious personality had negative influence on the overall slope, but did not significantly affect frequency or continuity.

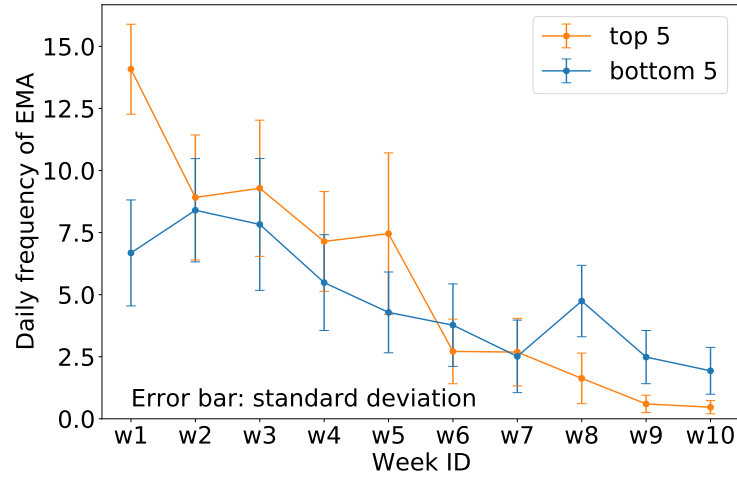


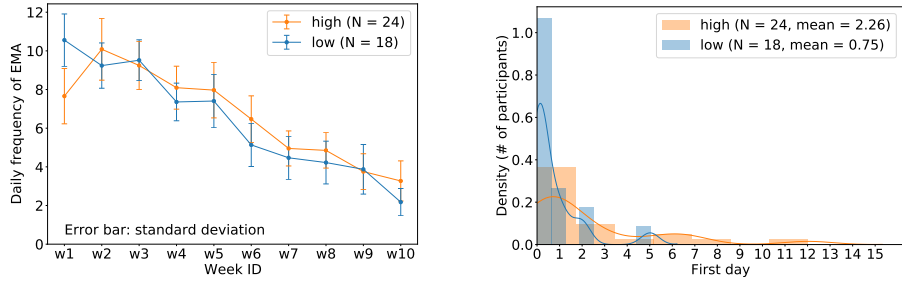
Fig. 3. Week-level daily frequency of top-5 and bottom-5 subjects in terms of conscientiousness.

Effects of Agreeableness Agreeableness was found to weaken the slope in period I. To understand this more deeply, we first separated the subjects into two groups by the mean value of agreeableness: high ($N = 24$) and low ($N = 18$) groups. We then compared the week-level daily EMA frequency of these groups. As shown in Figure 4(a), the high group shows a strong increase from week 1 to week 2. We assumed that the subjects in the high group tended to delay making their initial EMA response.

To assess our assumption, we compared the time from the start of the experiment to the first time of making EMA response across these groups. Figure 4(b) shows the distributions of the number of subjects for the time between the start of the experiment and the first EMA response. For these distributions, we conducted Welch’s t-test to see whether the mean values of the time are statistically different across these groups. As a result, we confirmed the mean value of the high group was higher than that of the low group ($t = 2.245, p < 0.05$). Therefore, our assumption was confirmed. However, we could not understand why the subjects with higher agreeableness delayed making their initial responses to EMA questions. Assessing this issue is a future task.

6.2 Theoretical Implications

Based on the above results and discussion, we discussed here how EMA data is biased according to personal characteristics and timing. In this discussion, the bias in EMA data can be divided into two types: *overall bias* and *local bias*. Overall bias refers to the fact that the EMA responses collected are biased toward subjects with certain characteristics. Local bias refers to the fact that the collected EMA responses are biased toward subjects with certain characteristics



(a) Daily EMA frequency in each week. (b) Distributions and density curves of time from the start of the experiment to the first EMA response.

Fig. 4. Comparison of week-level EMA frequency and initial response time by high and low groups of agreeableness.

at a particular period or timing. The implications of overall bias and local bias are summarized in Table 4.

Among the personal characteristics in this study, perceived stress can be a factor behind overall bias, because it yields differences in the frequency of daily EMA responses among subjects. More specifically, it means that the EMA data is more likely to contain responses from subjects with higher perceived stress levels. The other personal characteristics were not found to cause overall bias, indicating that the overall proportion of EMA data is not biased by the personal characteristics other than perceived stress.

Local bias refers to the fact that the collected EMA responses are biased toward subjects with certain characteristics at a particular period or timing. The extent of local bias can be explained by the steepness of overall slopes of response rates. Our results indicated that subjects with higher perceived stress, extraversion, and conscientiousness had stronger negative overall slopes. This means that the EMA responses of subjects with these personal characteristics decrease over time strongly. In other words, the percentage of responses from subjects with a high degree of these personal characteristics is high in the early period of the survey, compared to the percentage in the later period of the survey. In this manner, the distribution of these personal characteristics in the EMA data is skewed for time.

The extent of local bias can be also explained by period-separated slopes, that is, the slopes of response rates in specific period. In the early stage of the EMA study, subjects with higher agreeableness were found to delay making an initial EMA response. Therefore, the EMA responses from subjects with high agreeableness tend to be missing in the early stage. After tense events such as the midterm examination, the proportion of EMA responses from subjects with high perceived stress may decrease. Similarly, the EMA responses of subjects with low extraversion are likely to be missed after they have social experiences.

Table 4. EMA data bias and personal characteristics.

Bias type	Personal characteristic	Description (In the EMA data collected, ...)
Overall bias	Perceived stress	% responses from subjects with high degree of it is large.
Local bias	Perceived stress	% responses from subjects with high degree of it decreases over time.
	Extraversion	% responses from subjects with high degree of it decreases over time.
	Conscientiousness	% responses from subjects with high degree of it decreases over time.
Local bias	Perceived stress	% responses from subjects with a high degree of it strongly decrease after tense events.
	Extraversion	% responses from subjects with a low degree of it strongly decreases after social experience.
	Agreeableness	% responses from subjects with a high degree of it strongly increases in the early stage (they are late for making an initial EMA response).

The results of the overall bias are similar to those of previous studies [37, 30, 23, 19]. This is because they show the relationship between the characteristics of the subjects and the frequency of their responses throughout the experiment. However, the present study certainly shows a new finding, since previous studies did not verify the stress level of the subjects. More importantly, the results for local bias describe how the heterogeneity of EMA data due to individual characteristics varies by timing. The results differ from the findings of previous studies and are the main contribution of this study.

6.3 Recommendations

As shown by the present study and previous works, EMA data is inevitably biased in terms of personal characteristics and timing. However, the data collected by EMA are widely used for both academic and practical purposes. This is because they are accepted as true data that represent changes in people's subjective states.

Therefore, we encourage prospective EMA studies to describe the variance of EMA response rates as determined by individual characteristics and time, and to discuss the biases present in the data. These studies include, for example, studies that use EMA data to predict people's future moods, and studies that compare methods for increasing response rates to EMA.

In EMA studies involving human subjects, determining the nature of the study sample and assessing its impact are important to confirm the reproducibility of the findings. Such individual-level analyses will become increasingly im-

portant in the future development of HCI research. However, it is challenging for EMA studies to control study subjects and timing. Therefore, future EMA studies should share their findings on the bias of their respective EMA data and discuss, in depth, the limitations and generality of their work.

6.4 Limitations

Our study is a case study that points out bias in EMA data in a specific situation. Thus, our findings will be valid when one has similar conditions as to events, subjects, study design, and environments, particularly when subjects experience influential events in social contexts in the community (e.g., final examinations for students, salary increase interview for working adults, etc.).

We considered such influential events (a midterm examination and a school festival) that most of the subjects encountered, and did not focus on purely personal experiences. For example, if a subject had a negative experience such as troubles with family or traffic accidents, the subject would lose his/her energy regardless of his/her degree of perceived stress or introverted personality. Our research did not take this point into account, because our focus is the effect of personal characteristics.

Moreover, our results are based entirely on the StudentLife dataset. However, as discussed above, prospective EMA studies need to share their findings on the biases of their EMA data to understand the limits and the generality of EMA studies. The present study certainly contributes to these issues.

6.5 Future Work

To understand the effects of such personal experiences on response rates, it might be useful to examine passively-sensed data such as locations, mobility, phone logs, and so on. This contextual information can be used for inferring subjects' experiences. Our future work will investigate how response rates are associated with this contextual information to better understand individual differences in response rates from the viewpoint of personal ongoing experience.

To achieve the general goal of understanding how EMA data is biased, it is necessary to carry out experiments individually under various conditions in which each factor is controlled, and to perform a meta-analysis that aggregates the data of each experiment. For example, EMA experiments will be conducted in different cultures (e.g., Western vs. Eastern) and with different social statuses (e.g., students vs. workers). It is expected that generalized findings will be derived by comparing the results of the present study with those results, or by integrated analyses based on the experimental data of the present study.

7 Conclusion

EMA data is inherently heterogeneous. However, it is currently recognized and utilized as true data that represents the changes in people's subjective state over

time. Therefore, it is important to discuss how the EMA data is biased. In this study, unlike previous studies, we analyzed the temporal changes in the response rate of EMA data.

Primary results indicated that the subjects with high perceived stress had a higher frequency of EMA responses, but the frequency of EMA responses from these subjects tended to decrease strongly throughout the experiment, especially after the midterm examination. We also found that the subjects with low extraversion had a weak tendency to decrease the number of EMA responses throughout the experiment, but the number of EMA responses of these subjects decreased strongly after the school festival. In addition, conscientiousness had a negative effect on the overall slope of EMA frequency. Subjects with high agreeableness were found to show a strong increase in EMA frequency at the initial stage of the experiment. We discussed the reasons behind these significant results, and provided new insights into the biasing of EMA data.

In the future, we will consider individual contextual information to understand the effect of purely personal experience on data bias, and conduct meta-analysis using multiple datasets to derive more general findings. Future EMA research should describe how the collected EMA data are biased with regard to individual characteristics and time, and advance the discussion on the limitations and generality of EMA research. We hope that this study will be one of such contributions.

References

1. The Myers & Briggs Foundation. Extraversion or Introversion. Retrieved May 24th, 2021, from <https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/extraversion-or-introversion.htm> (2014)
2. Bentler, P.M.: Comparative Fit Indexes in Structural Models. *Psychological Bulletin* **107**(2), 238–246 (1990)
3. van Berkel, N., Goncalves, J., Hosio, S., Sarsenbayeva, Z., Velloso, E., Kostakos, V.: Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies* **134**, 1–12 (2020)
4. van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., Kostakos, V.: Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* **125**, 118–128 (2019)
5. Boukhechba, M., Cai, L., Chow, P.I., Fua, K., Gerber, M.S., Teachman, B.A., Barnes, L.E.: Contextual Analysis to Understand Compliance with Smartphone-based Ecological Momentary Assessment. In: *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. pp. 232–238. ACM, New York, NY, USA (2018)
6. Cohen, S., Kamarck, T., Mermelstein, R.: A Global Measure of Perceived Stress. *Journal of Health and Social Behavior* **24**(4), 385 (1983)
7. Costa, P.T., MacCrae, R.R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual. *Psychological Assessment Resources, Incorporated* (1992)

8. Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., Tuerlinckx, F.: Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour* **3**(5), 478–491 (2019)
9. Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.w., Oishi, S., Biswas-Diener, R.: New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research* **97**(2), 143–156 (2010)
10. Eagle, N., Pentland, A., Lazer, D.: Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* **106**(36), 15274–15278 (2009)
11. Harari, G.M., Gosling, S.D., Wang, R., Chen, F., Chen, Z., Campbell, A.T.: Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior* **67**, 129–138 (2017)
12. Helgoe, L.A.: *Introvert Power: Why Your Inner Life is Your Hidden Strength*. Sourcebooks, Inc. (2013)
13. Hsieh, H.P., Li, C.T.: Inferring social relationships from mobile sensor data. In: *Proceedings of the 23rd International Conference on World Wide Web*. pp. 293–294. ACM, New York, New York, USA (2014)
14. Jones, A., Remmerswaal, D., Verweer, I., Robinson, E., Franken, I.H.A., Wen, C.K.F., Field, M.: Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* **114**(4), 609–619 (2019)
15. Kroenke, K., Spitzer, R.L., Williams, J.B.W.: The PHQ-9. *Journal of General Internal Medicine* **16**(9), 606–613 (2001)
16. Lane, N., Mohammad, M., Lin, M., Yang, X., Lu, H., Ali, S., Doryab, A., Berke, E., Choudhury, T., Campbell, A.: BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In: *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*. IEEE (2011)
17. Lathia, N., Rachuri, K.K., Mascolo, C., Rentfrow, P.J.: Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods. In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. pp. 183–192. ACM, New York, NY, USA (2013)
18. Lucas, G.M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., Morency, L.P.: Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers in Robotics and AI* **4**(51), 1–9 (2017)
19. Messiah, A., Grondin, O., Encrenaz, G.: Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and Alcohol Dependence* **114**(2-3), 153–158 (2011)
20. Morren, M., Dulmen, S., Ouwkerk, J., Bensing, J.: Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain* **13**(4), 354–365 (2009)
21. Mount, M.K., Barrick, M.R.: The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in personnel and human resources management* **13**(3), 153–200 (1995)
22. Rabbi, M., Ali, S., Choudhury, T., Berke, E.: Passive and In-Situ assessment of mental and physical well-being using mobile sensors. In: *Proceedings of the 13th International Conference on Ubiquitous Computing*. pp. 385–394. ACM, New York, New York, USA (2011)
23. Rintala, A., Wampers, M., Myin-Germeys, I., Viechtbauer, W.: Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment* **31**(2), 226–235 (2019)

24. Roccas, S., Sagiv, L., Schwartz, S.H., Knafo, A.: The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin* **28**(6), 789–801 (2002)
25. Russell, D.W.: UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure. *Journal of Personality Assessment* **66**(1), 20–40 (1996)
26. Sadri, A., Salim, F.D., Ren, Y., Shao, W., Krumm, J.C., Mascolo, C.: What Will You Do for the Rest of the Day? An Approach to Continuous Trajectory Prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2**(4), 1–26 (2018)
27. Saha, K., Chan, L., De Barbaro, K., Abowd, G.D., De Choudhury, M.: Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**(3), 1–27 (2017)
28. Servia-Rodríguez, S., Rachuri, K.K., Mascolo, C., Rentfrow, P.J., Lathia, N., Sandstrom, G.M.: Mobile Sensing at the Service of Mental Well-being: a Large-scale Longitudinal Study. In: *Proceedings of the 26th International Conference on World Wide Web*. pp. 103–112 (2017)
29. Shiffman, S., Stone, A.A., Hufford, M.R.: Ecological Momentary Assessment. *Annual Review of Clinical Psychology* **4**(1), 1–32 (2008)
30. Sokolovsky, A.W., Mermelstein, R.J., Hedeker, D.: Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine and Tobacco Research* **16**(3), 351–358 (2014)
31. Steiger, J.H.: Statistically based test for the number of common factors. In: *The annual meeting of the Psychometric Society*. (1980)
32. Thewissen, V., Bentall, R.P., Oorschot, M., à Campo, J., van Lierop, T., van Os, J., Myin-Germeys, I.: Emotions, self-esteem, and paranoid episodes: An experience sampling study. *British Journal of Clinical Psychology* **50**(2), 178–195 (2011)
33. Thompson, E.R.: Development and Validation of an International English Big-Five Mini-Markers. *Personality and Individual Differences* **45**(6), 542–548 (2008)
34. Toegel, G., Barsoux, J.L.: How to become a better leader. *MIT Sloan Management Review* **53**(3), 51–60 (2012)
35. Tucker, L.R., Lewis, C.: A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* **38**(1), 1–10 (1973)
36. Udachina, A., Varese, F., Myin-Germeys, I., Bentall, R.P.: The role of experiential avoidance in paranoid delusions: An experience sampling study. *British Journal of Clinical Psychology* **53**(4), 422–432 (2014)
37. Vachon, H., Viechtbauer, W., Rintala, A., Myin-Germeys, I.: 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) (2019)
38. Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., Campbell, A.T.: StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. pp. 3–14. ACM, New York, New York, USA (2014)
39. Wang, R., Harari, G., Hao, P., Zhou, X., Campbell, A.T.: SmartGPA : How Smartphones Can Assess and Predict Academic Performance of College Students. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. pp. 295–306 (2015)

40. Yao, H., Lian, D., Cao, Y., Wu, Y., Zhou, T.: Predicting academic performance for college students: A campus behavior perspective. *ACM Transactions on Intelligent Systems and Technology* **10**(3), 1–21 (2019)