Validation of a transdiagnostic psychopathology ecological momentary assessment protocol in a university student sample

Jover Martinez, A.; Lemmens, L.H.J.M.; Fried, E.I.; Guomundsdottir, G.R.; Roefs, A.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

2025, Vol. 37, Nos. 1–2, 46–61
https://doi.org/10.1037/pas0001348

Validation of a Transdiagnostic Psychopathology Ecological Momentary Assessment Protocol in a University Student Sample

Alberto Jover Martínez[1], Lotte H. J. M. Lemmens[1], Eiko I. Fried[2], Guðrún R. Guðmundsdóttir[1], and Anne Roefs[1]

[1] Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University
[2] Department of Clinical Psychology, Faculty of Social and Behavioral Sciences, Leiden University

Ecological momentary assessment (EMA) collects real-time data in daily life, enhancing ecological validity and reducing recall bias. An EMA questionnaire that measures symptoms and transdiagnostic factors was recently developed with network modeling purposes. This study examines this EMA protocol's (a) subjective experience (e.g., burden, item clarity, survey frequency adequacy); (b) compliance, dropout, and predictors thereof; (c) the variability of EMA items across and within participants; and (d) the relations between EMA items and baseline standardized psychopathology questionnaires. University students (n = 262, M_{age} = 21.9, 84.8% females, 17.2% Dutch) completed eight daily momentary surveys (with the first including the morning survey), an evening survey, and a weekly survey during a 4-week EMA protocol. Additionally, a concluding survey examined participants' subjective experiences. Perceived burden was 3.40 on a 7-point scale, and people with higher levels of psychopathology found it more burdensome and more difficult to complete. Moreover, 67% of the surveys were completed, and 16% of the participants dropped out. Baseline psychopathology was not significantly associated with dropout or compliance. Moreover, surveys triggered in later study days, during the weekend, longer surveys, and surveys with lower financial reward were more likely to be missed. Between-subjects and within-subjects variability and correlations with baseline psychopathology varied across EMA items, with most EMA items showing sufficient within-individual variability for network modeling purposes and showing correlations across all types of psychopathology and transdiagnostic factors. The results suggest that the collection of intensive time-series data is feasible, and data quality and characteristics match requirements of different network models.

Public Significance Statement
To understand how psychopathology unfolds in daily life in individuals, relevant symptoms need to be assessed in real time, for which an optimal transdiagnostic measurement protocol is necessary. Our developed protocol was positively experienced by participants, sufficient measurements were completed, and items showed sufficient variability and showed good validity.

Keywords: network modeling, ecological momentary assessment, psychopathology, transdiagnostic

Supplemental materials: https://doi.org/10.1037/pas0001348.supp

The network approach to psychopathology is an alternative framework to the medical model for understanding mental disorders (Borsboom, 2017). This approach states that the symptoms of mental disorders are not provoked by an underlying common cause located in the brain, as the medical model proposes (Bruce, 2009; Deacon, 2013). Instead, it posits that a system of dynamic interactions between symptoms—within an individual—constitutes a mental disorder (Borsboom, 2017). Recently, it was proposed to not only include symptoms in these networks, but also other relevant variables, such as contextual circumstances, and certain behaviors, like the activities an individual engages in or social interactions (Roefs et al., 2022). As such dynamic interactions occur between

Kasey Stanton served as action editor.
Alberto Jover Martínez https://orcid.org/0009-0003-9653-9050
Lotte H. J. M. Lemmens https://orcid.org/0000-0002-2037-1808
Eiko I. Fried https://orcid.org/0000-0001-7469-594X
Guðrún R. Guðmundsdóttir https://orcid.org/0000-0002-0912-6414
Anne Roefs https://orcid.org/0000-0002-9935-1075
The presented article is part of a dissertation. This study was preregistered (#62825) at https://aspredicted.org/79X_XY5. The code used for the analysis is available at https://osf.io/hq6fn/, and data are available upon request. The authors have no known conflicts of interest to disclose. This study is part of the project "New Science of Mental Disorders" (https://www.nsmd.eu), supported by the Dutch Research Council and the Dutch Ministry of Education, Culture and Science (Nederlandse Organisatie voor Wetenschappelijk Onderzoek Gravitation Grant 024.004.016).

This research was funded in whole, or in part, by the Netherlands Organisation for Scientific Research (Grant 024.004.016). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Alberto Jover Martínez played a lead role in data curation, formal analysis, investigation, methodology, project administration, visualization, and writing–original draft and an equal role in conceptualization. Lotte H. J.

continued

symptoms across diagnostic categories (Cramer et al., 2010), the network approach is transdiagnostic in nature.

Taking an individual transdiagnostic perspective requires examining how mental health phenomena evolve over time within an individual. Ecological momentary assessment (EMA) is commonly used to assess these phenomena in psychopathology (Shiffman et al., 2008). EMA encompasses a variety of methods (e.g., time-contingent or event-related sampling) to collect repeated real-time data in participants' natural environments (Shiffman et al., 2008), mostly using smartphones for data collection. EMA studies use items from questionnaires and surveys that were developed for laboratory or retrospective survey studies (Cloos et al., 2023; Schreuder et al., 2020). This is not necessarily problematic, but the psychometric qualities and subjective experience of any EMA protocol and items must be carefully examined.

The present study investigated characteristics of a transdiagnostic EMA protocol, which assesses symptoms across the main categories of mental disorders, disorder-specific symptoms, and other psychopathology-relevant variables such as social context. It was developed based on input from expert clinicians, with the goal of enabling data collection for the estimation of transdiagnostic intraindividual networks (Jover Martínez et al., 2024). Specifically, the present study examined (a) the subjective experience of participants in the EMA protocol; (b) overall compliance, momentary predictors of compliance (predictors that change from moment to moment), dropout, and the predictors of both compliance and dropout; (c) the within-subjects and between-subjects variability of EMA items; and (d) the relationship between the EMA items and standardized questionnaires of psychopathology.

Goal 1: Subjective Experience EMA Protocol

The experience of participants—such as the experienced burden—in EMA studies is known to influence the quantity and quality of data (Eisele et al., 2022; Moskowitz & Young, 2006; Stone et al., 2003). For example, longer questionnaires are related to higher momentary and retrospectively reported burden, but questionnaire frequency is not (Eisele et al., 2022). Given the paucity of evidence available on the burden of EMA studies, and considering that it can affect data quality, dropout, and compliance, assessment of experienced burden and other relevant subjective experiences, such as perceived clarity of the questions, in EMA studies is essential.

Goal 2: Overall Compliance, Momentary Predictors of Compliance, Dropout, and Related Factors

In addition to burden, there are other variables that can influence compliance and dropout (i.e., leaving the study before completion), for example, personal variables such as gender or age, study variables such as length of the questionnaire, and momentary predictors such as the participant's momentary mood. Researchers have mostly explored compliance at the study level, focusing on personal (e.g., gender, age, mental health diagnosis), design (e.g., frequency of assessment, length of surveys, length of study), and time characteristics (e.g., study days, time within a day, weekdays). Regarding personal characteristics, compliance of female participants is typically higher than that of male participants (Eisele et al., 2022; Vachon et al., 2019), but there are no gender differences in dropout (Wrzus & Neubauer, 2023). If and how a clinical diagnosis affects compliance is less clear, with studies providing mixed findings (Jones et al., 2019; Rintala et al., 2020; Vachon et al., 2019). However, it seems like neither psychological nor physical health conditions are related to dropout (Wrzus & Neubauer, 2023). Age does not seem to be significantly related to compliance either (Rintala et al., 2020; Vachon et al., 2019).

Regarding characteristics of the study design, research has not found relations between variables such as study length or compliance reinforcement (i.e., rewarding participants more when they are more compliant) and compliance or dropout (Jones et al., 2019; Vachon et al., 2019; Wrzus & Neubauer, 2023). Other design characteristics can impact compliance. Longer questionnaires are associated with lower compliance rates (Eisele et al., 2022) and longer between-survey intervals and higher incentives with higher compliance rates (Vachon et al., 2019; Wrzus & Neubauer, 2023). Moreover, an experimental study comparing different EMA protocols found that all dropouts occurred in the condition with longer surveys and higher assessment frequency (Eisele et al., 2022). However, results on the relation between assessment frequency and compliance are mixed. Most studies find no significant relation between survey frequency and compliance (Eisele et al., 2022; Jones et al., 2019; Stone et al., 2003), but in one study, a negative relation was found (Vachon et al., 2019). Last, regarding time characteristics, compliance is consistently lower on later study days (Eisele et al., 2022; Rintala et al., 2019, 2020). The findings regarding specific weekdays, or specific times within a day, are mixed and inconclusive (Rintala et al., 2019, 2020; Vachon et al., 2019).

Studying momentary predictors of compliance—predictors of compliance that change from moment to moment—is less common. One study found that being outside of home at the time of a survey, feeling disturbed by a survey, taking medication, and responding to a survey on later study days reduced the likelihood of responding to a specific survey (Rintala et al., 2020). Interestingly, deviation of an individual's mood from their mean, stress-related variables, and physical state variables were not significantly related to the likelihood of answering a survey (Rintala et al., 2020). Last,

M. Lemmens played a supporting role in conceptualization and methodology and an equal role in writing–review and editing. Eiko I. Fried played an equal role in conceptualization, funding acquisition, and writing–review and editing. Guðrún R. Guðmundsdóttir played a supporting role in writing–review and editing and an equal role in data curation. Anne Roefs played a lead role in funding acquisition, project administration, supervision, and writing–review and editing, a supporting role in formal analysis, and an equal role in conceptualization and methodology.

Correspondence concerning this article should be addressed to Alberto Jover Martínez, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Universiteitssingel 40, 6229 ER Maastricht, the Netherlands. Email: ajovermartinez.psychology@gmail.com

certain weekdays and hours within a day were related to the likelihood of responding to a survey (Rintala et al., 2020). This type of information can be used to optimize compliance in EMA studies, but "little is still known about factors that influence compliance" (Rintala et al., 2020, p. 1). Therefore, investigating our protocol's compliance rates, along with predictors thereof, can determine how feasible it is and provide information on how to improve it.

Goal 3: Within- and Between-Individual Variability

Another consideration for EMA studies is that items need to be specifically designed, for these purposes. EMA studies in psychopathology have frequently used standard questionnaires, which were designed for laboratory or retrospective studies. The tacit assumption was that these questionnaires would also be able to capture the momentary fluctuations of interest (Schreuder et al., 2020), which is one of the core assumptions of EMA studies (Schreuder et al., 2020). However, it is questionable if this assumption is always valid. In case of not meeting this assumption, among others, available statistical models will not be able to capture relations between variables due to the lack of variability. Therefore, studying the variability of the EMA items is crucial.

Goal 4: Relationship Between EMA Items and Standardized Questionnaires of Psychopathology

Moreover, EMA questionnaires need to have construct validity (Cronbach & Meehl, 1955). Here we focus on convergent and divergent validity. Convergent validity is achieved when the measure of interest is correlated to other measures that assess the same or similar constructs, whereas divergent validity is achieved if the measure of interest is uncorrelated with dissimilar measures. Therefore, for our research purposes, items included in an EMA questionnaire should reflect the entire range of psychopathology. Moreover, the different items should be associated with specific subscales of standardized questionnaires of psychopathology and uncorrelated to other subscales (Schreuder et al., 2020). For example, an item that captures body image issues should be highly correlated with subscales that measure eating disorders but uncorrelated with subscales that measure, for example, interpersonal problems.

The Present Study

With this in mind, the present study had four objectives. First, the relations between the subjective experience of participating in the EMA protocol and personal variables, such as level of psychopathology, gender, and diagnosis, were studied. Second, overall compliance and its relation to personal characteristics, such as psychopathology, gender, and past diagnosis, were studied. Moreover, momentary compliance was studied in relation to time variables (i.e., study day and weekend), study variables (i.e., survey type), and personal variables (i.e., positive and negative affect and missingness at the previous time point). Furthermore, dropout was studied in relation to gender, diagnosis, and level of psychopathology. Third, the within- and between-individual variability of the EMA items was investigated. Fourth and last, the relationships between the EMA items and standard questionnaires of psychopathology were examined.

Method

Participants

The inclusion criteria were being university students of Maastricht University, Leiden University, or the University of Amsterdam. They needed to have sufficient English proficiency and own a smartphone. The reward for participating consisted of up to €75 in vouchers or up to €60 in vouchers and two research credits (one research credit equals 1 hr of work or €7.5) depending on their compliance. If a participant chose the €75-voucher option, this participant would get €10.63 for filling in the baseline, €0.26 for each momentary survey, and €0.49 for any survey that was not momentary. For the €60-voucher option, the baseline had the same reward, each momentary item added €0.20, and the other surveys added €0.38. Moreover, the research credits varied based on compliance: 0.5 credits for 0%–25%, 1 credit for 26%–50%, 1.5 credits for 51%–75%, and 2 credits for 76%–100%.

A total of 322 participants showed interest in the study, but 34 did not complete the baseline questionnaire. Of the remaining 288 participants, 26 participants (9.1%) did not start the EMA measurement protocol, leaving a total sample of 262. All analyses were performed with the 262 participants who began the EMA protocol, except the dropout analysis that included the 288 participants who completed the baseline questionnaire. The mean age of the participants was 21.9 years (SD = 2.9), 84.8% were female (n = 218) and came from different countries (i.e., 19.5% German, 17.2% Dutch, 7.3% Italian, 3.4% British, 3.4% Chinese, 3.4% American, 3.1% Spanish, 2.7% Belgian, 2.3 Polish, 2.3 Romanian, 1.5% Danish, 1.5% Hungarian, and 32.4% other nationalities). Moreover, 24.4% (n = 64) had received a diagnosis of a mental disorder at some point in their lives, and four participants were receiving some type of treatment for a mental disorder at the beginning of the study. According to the Depression Anxiety and Stress Scale–21 (Lovibond & Lovibond, 1995), 38 participants (15%) had elevated stress levels, 112 participants (43%) had elevated levels of anxiety, and 83 participants (32%) had elevated levels of depression. The study was approved by the ethical review board of the Faculty of Psychology & Neuroscience of Maastricht University.

Procedure

Participants could join the study between March 2, 2022, and May 31, 2022. The study was advertised on university advertisement boards, on a "research credit" platform for students (i.e., SONA), and on social media (Instagram and Facebook). The advertisements contained a link or a QR code that directed participants to the study website, where they were informed about the study and provided informed consent. The study consisted of a screening and a monitoring phase. To start the study, participants were instructed to download the app "Ethica/Avicenna" from Ethica Data/Avicenna Research (https://avicennaresearch.com/), which was used for data collection.

Screening Phase

First, participants completed an online screening in Ethica/Avicenna to check the inclusion criteria and to provide some personal information (e.g., name, phone number, and email). The following day, participants received a set of questionnaires on

psychopathology (see further), which had to be completed in 4 days. On the fifth day, a practice day, structured like a real study day, was completed to familiarize themselves with the procedure. On one of the following 4 days, participants received an evaluation phone call to verify that they met the inclusion criteria and that the EMA protocol was clear. The following day, the monitoring phase began. Participants could only start if they completed the online screening, the baseline questionnaires, the practice day, and the evaluation phone call. For a summary of the screening phase timeline, see Figure 1.

Monitoring Phase

The monitoring phase consisted of 28 days on which participants were prompted several times per day to answer surveys on their smartphones. Each day, they received a morning survey, eight momentary surveys throughout the day, and an evening survey. In addition, at the end of each week, they received a weekly survey. The day after the last EMA survey, participants received a survey on their experience of the study. The morning, evening, and weekly surveys were rewarded 50% more than the momentary surveys. The participants received emails approximately every seventh day updating them on how much reward they would earn if they kept answering at their current pace. If participants were not compliant, they were called to find out why, and if a problem was identified, a solution was sought.

Measurements

A thorough overview of all measures, including instructions, scoring, means, and variability metrics for the EMA items, and Cronbach's alphas for the baseline measures are available online at https://osf.io/hq6fn/ (Jover Martínez & Guðmundsdóttir, 2024).

Baseline Assessment

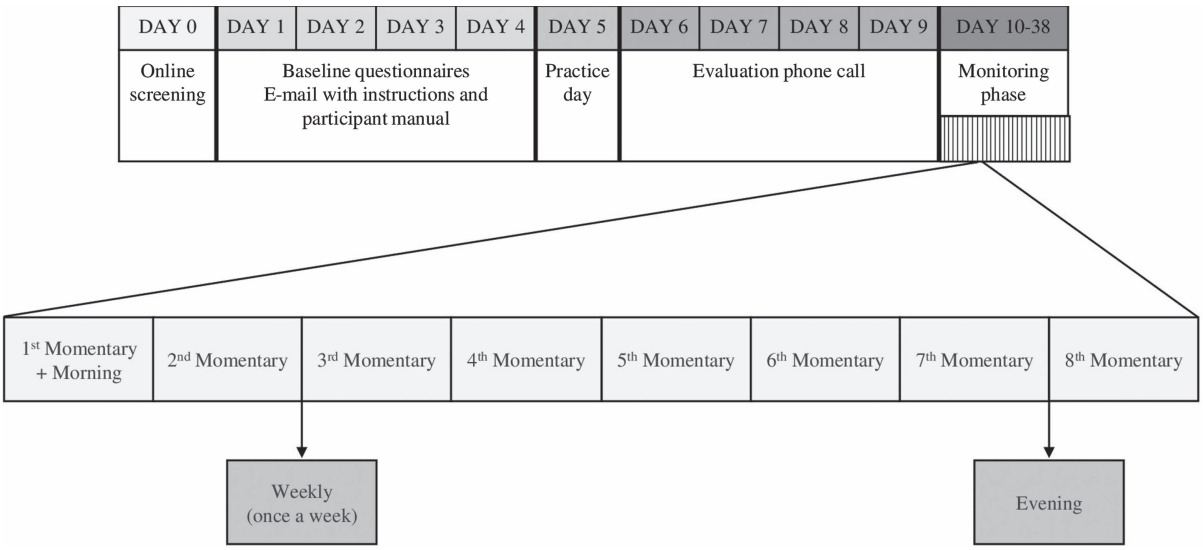
Participants first completed some questions about demographic and personal characteristics (e.g., gender, employment, age, nationality). Next, participants completed 15 standard questionnaires of psychopathology: 11 about a specific type of psychopathology and four about transdiagnostic factors. See Supplemental Table S3 for an overview and the minimum and maximum score of each questionnaire.

Brief Symptom Inventory. The Brief Symptom Inventory (BSI) is a 53-item psychological self-report symptom scale (Derogatis & Melisaratos, 1983), scored on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extremely*). It measures nine psychopathology dimensions: somatization, obsession–compulsion, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Each scale score is computed as the mean of the items that make up the scale. A general BSI score was computed as the average of all items, with higher scores indicating greater severity of symptoms (de Beurs & Zitman, 2006).

Depression Anxiety Stress Scale. The Depression Anxiety and Stress Scale–21 is a self-report scale composed of three seven-item subscales scored on a 4-point Likert scale ranging from 0 (*did not apply to me at all*) to 4 (*applied to me very much, or most of the time*). The Depression Anxiety and Stress Scale–21 subscales measure depression, anxiety, and stress (Lovibond & Lovibond, 1995). To calculate the subscale scores, the seven items of each subscale are summed, and the sum is multiplied by two. A higher score on a scale indicates a greater level of the measured state.

Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Adult Attention-Deficit/Hyperactivity Disorder and Self-Report Screening Scale (ASRS). The ASRS is a six-item self-report scale that measures the severity of *Diagnostic and Statistical Manual of Mental Disorders, fifth edition* (*DSM-5*) symptoms for the

Figure 1
Study Timeline



Note. The lower part of the figure represents a single day of the monitoring phase. The evening and weekly phases are indicated with an arrow because they were triggered at a fixed time. The weekly survey was triggered only once a week.

diagnosis of adult attention-deficit/hyperactivity disorder scored on a 5-point Likert scale ranging from 0 (*never*) to 4 (*very often*), with higher scores indicating greater symptom severity (Ustun et al., 2017).

Autism Spectrum Quotient. The Autism Spectrum Quotient–10 is a 10-item self-report scale that measures the expression of autism-spectrum traits (Allison et al., 2012) scored on a 4-point Likert scale ranging from 1 (*definitely disagree*) to 4 (*definitely agree*). To compute the Autism Spectrum Quotient–10 total scored, each item above 2 is counted. Higher quotient scores indicate greater autism-spectrum expression. For an explanation of how the quotient is calculated, see Allison et al. (2012).

Eating Disorders Examination Questionnaire–Short. The Eating Disorders Examination Questionnaire–Short is a 12-item self-report scale designed to assess the range, frequency, and severity of behaviors related to eating disorders diagnoses (Gideon et al., 2016) scored on a 4-point Likert scale ranging from 0 (*0 days or not at all*) to 3 (*6–7 days or markedly*). A sum score is used to summarize the scale, with higher scores indicating greater eating pathology.

Posttraumatic Stress Disorder Checklist for DSM-5. The Posttraumatic Stress Disorder (PTSD) Checklist for *DSM-5* is a 20-item self-report scale that assesses the 20 *DSM-5* symptoms of PTSD (Blevins et al., 2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* is answered on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extremely*). A severity score is obtained by summing the items' scores, with higher scores indicating greater PTSD symptom severity.

Sexual Dysfunction Questionnaire. The Sexual Dysfunction Questionnaire is a 19-item self-report scale that assesses sexual problems in psychiatric patients (Infrasca, 2011) scored on a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*). The total score is a sum score of all the items' scores, with higher scores indicating a higher likelihood of experiencing sexual problems.

Insomnia Severity Index. The Insomnia Severity Index is a seven-item self-report scale that measures sleeping difficulties (Bastien et al., 2001) scored on a 5-point Likert scale ranging from 0 (*none or very dissatisfied*) to 4 (*very severe or very satisfied*). The seven items are summed up, and higher sum scores indicate greater sleeping difficulties.

Alcohol Use Disorders Identification Test. The Alcohol Use Disorders Identification Test is a 10-item self-report scale that screens for risky or hazardous alcohol use (World Health Organization, 2001). The first eight items are assessed on 5-point Likert scales. Of those eight, the first item goes from *never* to *4 times a week or more often*, the second goes from *1–2* to *10 or more*, and the other six go from "*never*" to "*daily or almost daily*." The last two items are assessed on a 3-point scale from *never* to *Yes, during the last year*. The items' scores are added up to a total score, with higher scores indicating higher alcohol dependence.

Drug Use Disorders Identification Test. The Drug Use Disorders Identification Test is an 11-item self-report scale that screens for drug-related problems (Berman et al., 2002). The first nine items are assessed on 5-point Likert scales. Of those nine, the first two items range from *never* to *four times a week or more often*, the third ranges from *0* to *17 or more*, and the last six items range from *never* to *daily or almost daily*. The last two items are assessed on 3-point Likert scales that range from *never* to *yes, over the past year*. The items' scores are added up in a total score, with higher scores reflecting higher drug dependence.

Levels of Personality Functioning Scale Brief Form. The Levels of Personality Functioning Scale Brief Form is a 12-item self-report scale that measures the presence and general severity of personality pathology scored on a 4-point Likert scale ranging from 1 (*completely untrue*) to 4 (*completely true*). The items are averaged, and higher scores reflect greater severity of personality pathology (Weekers et al., 2019).

Brief Fear of Negative Evaluation Scale. The Brief Fear of Negative Evaluation Scale is a 12-item self-report scale of a person's tolerance for the possibility they might be judged disparagingly or hostilely by others (Duke et al., 2006) scored on 5-point Likert scales ranging from 1 (*not at all characteristic of me*) to 5 (*extremely characteristic of me*). A total score is calculated by adding the scores of the items, with higher scores indicating greater distress related to social situations.

Dichotomous Thinking Inventory. The Dichotomous Thinking Inventory (DTI) is a 16-item self-report scale that measures black-and-white cognitive thinking style (Byrne et al., 2008) scored on a 4-point Likert scale ranging from 1 (*not at all true of me*) to 4 (*very true of me*). The DTI consists of two subscales: one six-item subscale measuring dichotomous thinking related to food, eating, dieting, and weight and a 10-item subscale measuring general dichotomous thinking. Moreover, a global measure is computed by summing all items. On all these scales, higher scores indicate a greater tendency to think in black or white terms.

Self-Control Questionnaire. The Self-Control Questionnaire is a 36-item self-report scale that measures an individual's ability to control impulses, modulate cognitive and affective processes, and intervene on undesirable behavioral tendencies while refraining from their execution (Brandon et al., 1990) scored on a 5-point Likert scale ranging from 1 (*not at all like me*) to 5 (*very much like me*). The total score is computed by summing up the items, with higher scores indicating a higher degree of self-control.

Ten-Item Personality Inventory. The Ten-Item Personality Inventory is a 10-item self-report brief version of the Big Five personality dimensions (Gosling et al., 2003) scored on a 7-point Likert scale from *disagree strongly* to *agree strongly*. Pairs of items represent each of the Big Five elements: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Averages of these item pairs are used to measure each of the Big Five elements, with higher scores reflecting higher levels of the measured personality element.

EMA Protocol

Details of the EMA protocol have been described elsewhere (Jover Martínez et al., 2024) and will only be summarized here. The transdiagnostic EMA protocol covered the whole spectrum of psychopathology and included four types of surveys: a morning survey (five items), a momentary survey (up to 35 items; eight times per day), an evening survey (up to 27 items), and a weekly survey (four items). See Figure 1 for a graphical representation of the EMA protocol. The morning survey coincided with the first momentary survey of the day. Some items were not triggered for everyone (e.g., Did you smoke since the last beep?), and other items were triggered based upon the answer to a previous item (e.g., "What do you crave?" was triggered only if the answer to "At this moment, I experience cravings" was not "Not at all"). Therefore, the specific number of items could differ across people and measurement moments. Most

items were quantitative and were answered on a 7-point Likert scale. The meaning of the different scores varied depending on the question. The questionnaires also included qualitative items asking about, for example, company, location, and substance use.

Momentary surveys were triggered semirandomly in time windows of 1:37:30, following a normal distribution to increase the likelihood of surveys being triggered in the middle of the time window. The evening and weekly surveys were triggered at a specific moment to increase compliance (Eisele et al., 2022). The morning and evening surveys expired in 45 min, the weekly survey expired in 12 hr, and the momentary surveys expired in 20 min. The time at which measurements started each day was adapted to participants' usual waking time. Participants were asked whether they usually woke up before 9 a.m., between 9 and 11 a.m., or after 11 a.m. Depending on the answer, the first survey (i.e., morning and first momentary survey) was triggered between 07:30:00 a.m. and 09:07:30 a.m., 09:07:30 a.m. and 10:45:00 a.m., or 10:45 a.m. and 12:22:30 p.m. respectively. Subsequently, the seven remaining momentary surveys were triggered within the semirandom time windows specified above. The evening surveys were triggered between 08:30:00 p.m. and 9:30:00 p.m., 9:30:00 p.m. and 10:30:00 p.m., or 10:30:00 p.m. and 11:30:00 p.m. depending on each participant's waking preference. The weekly survey was triggered at 12:00:00 p.m. for everybody.

Subjective Experience EMA Protocol

After the last survey of the monitoring phase was completed, participants received a survey about their subjective experiences participating in the EMA protocol. This survey consisted of seven items, assessed on 7-point Likert scales. The survey assessed whether participants thought the study period was a good representation of their lives, how difficult it was to complete the surveys, how burdensome the study was, how adequate the frequency of the surveys was, how clear the questions were, how difficult it was to know the answer to the items, and how much participation impacted on their daily lives. For an overview of this survey, see Table 1.

Analysis

Goal 1: Subjective Experience EMA Protocol

Intercorrelations between the subjective experience survey items and scores on the general BSI were computed. Additionally, the differences between participants who had received a diagnosis and those who had not, and between genders, in the evaluation-of-study survey questions were analyzed using independent-samples *t* tests.

Goal 2: Overall Compliance, Momentary Predictors of Compliance, Dropout, and Related Factors

The relationship between overall compliance and gender, past diagnosis, survey type, and different weekdays was studied by paired-samples *t* tests and one-way within-subjects analyses of variance. Last, the relationship between compliance and psychopathology was examined by correlation analysis.

Compliance at the momentary level was examined in a multilevel logistic regression model with the lme4 package (Bates et al., 2015) using whether a survey was answered or not as the dependent variable. Table 2 includes a summary of the predictors. Random

Table 1
End of the Study Survey Analysis

Item	*M* (*SD*)	Correlation with BSI	Gender	Diagnosis
In the past 4 weeks, to what extent was your day-to-day life representative of how it normally is? (1 = *an extremely bad representation*, 7 = *a perfect representation*)	5.06 (0.96)	**$r(232) = -0.13, p < .05$**	$t(49.75) = 1.75, p = .09$	$t(94.48) = 0.30, p = .76$
How did you experience filling out all notification surveys? (1 = *It was extremely difficult*, 7 = *It was extremely easy*)	3.85 (1.36)	**$r(232) = -0.22, p < .001$**	$t(49.75) = 0.65, p = .52$	$t(91.94) = 0.005, p = .99$
How burdensome was answering the notification surveys? (1 = *not burdensome at all*, 7 = *extremely burdensome*)[a]	3.50 (1.53)	**$r(232) = 0.19, p < .01$**	$t(55.36) = 0.04, p = .97$	$t(89.25) = 1.25, p = .22$
How was the frequency of the surveys (i.e., the number of times you were asked to fill in a survey every day)? (1 = *extremely high*, 7 = *extremely low*)[a]	2.89 (0.87)	$r(232) = -0.04, p = .52$	$t(57.71) = 0.85, p = .40$	$t(76.27) = 0.004, p = .99$
Were the questions clear? (1 = *extremely unclear*, 7 = *extremely clear*)	5.46 (1.10)	$r(93) = -0.14, p = .17$	$t(23.36) = 0.01, p = .99$	$t(72.14) = 1.20, p = .24$
How difficult was it for you to know the answers to the questions? (1 = *extremely difficult*, 7 = *extremely easy*)	4.74 (1.31)	**$r(93) = -0.43, p < .001$**	$t(29.26) = 0.60, p = .55$	$t(46.39) = 0.93\ p = .36$
How influential was your participation in this study on your day-to-day life? (1 = *not influential at all*, 7 = *extremely influential*)	4.96 (1.43)	$r(232) = -0.08, p = .25$	**$t(55.50) = 2.08, p = .04$**	$t(86.04) = 0.25, p = .81$

Note. The fourth and fifth columns contain the results of the independent *t* tests comparing the answers with each item between gender and diagnosis levels, respectively. Results in bold indicate significant results. *r* = correlation with levels of psychopathology; BSI = Brief Symptom Inventory.
[a] Item scale was the opposite in the original questionnaire; for the current analysis, the coding was reversed as described in the table.

intercepts were included per participant, and the predictors were first analyzed with univariate models and then with a multivariate model that included all predictors simultaneously. Positive affect was conceptualized as the average of happy, calm, and energetic and negative affect as the average of sad, guilty, hopeless, anxious, stressed, overwhelmed, angry, lonely, and paranoid.

To study dropout, a logistic regression model was used with dropout (0 = not dropout, 1 = dropout) as the dependent variable. A participant was considered a dropout if they did not start the last week of the EMA protocol. To predict dropout, the personal characteristics of Table 2 (i.e., gender, diagnosis, and BSI score) and the average level of positive and negative affect were used as predictors. All variables were introduced in the model simultaneously.

Goal 3: Within- and Between-Individual Variability

The variability of the EMA items was studied by computing the between- and within-individuals variance ($std_{between}$ and std_{within} respectively) and the intraclass correlation (ICC) of the quantitative items. The between-individuals variance for each item was computed by calculating the standard deviation of the participants' mean scores (collapsed over time points) on that item. The within-individual variance was computed by first calculating the standard deviation of each item per individual across time points and then calculating the average standard deviation across participants for each item. The ICC was calculated by taking the between-individuals variance per item and dividing it by the sum of the within-individuals variance and the between-individuals variance. A higher ICC value reflects a higher proportion of between-individuals variance relative to within-individuals variance. That is, scores on that item varied more across people than within an individual.

The qualitative items needed a different analytical approach. The variances of the qualitative items were reflected by Shannon's entropy (Shannon, 1948) estimated as seen in Equation 1. The range of Shannon's entropy can go from 0 up to $\log_2(k)$, with k being the number of categories of the studied variables. In the case of dichotomous outcomes, the entropy coefficient can range from 0 to 1. An entropy coefficient of 0 represents no uncertainty of the outcome, which in our case translates to no variability.

$$H(x) = -\sum P(x_i) \log P(x_i), \quad (1)$$

where $P(x_i)$ is the probability of an outcome, which is calculated by dividing the occasions that an event was present by the number of times that an event could have been present. The entropy of each qualitative item was calculated per participant, and then these entropy values were averaged across participants for each item.

Goal 4: Relationship Between EMA Items and Standardized Questionnaires of Psychopathology

To study the relation of the EMA items to baseline measures of psychopathology, correlations were computed. Specifically, the mean levels of the EMA items across time were correlated with the scores on the subscales of the baseline questionnaires.

Transparency and Openness

The reported analyses were not preregistered, but the study was preregistered on AsPredicted with registration number 78277 and can be accessed at https://aspredicted.org/ej6jp.pdf (Jover Martínez et al., 2023). Data are available upon reasonable request. Study materials and code used in this analysis are publicly available on the Open Science Framework at https://osf.io/hq6fn/ (Jover Martínez & Guðmundsdóttir, 2024). All analyses were performed using R 4.2.2 (R Core Team, 2016).

Results

Goal 1: Subjective Experience EMA Protocol

Table 1 contains a summary of the associations between the subjective experience of the EMA protocol, BSI score ($M = 0.81$,

Table 2
List of Predictors and Coding for Compliance at the Momentary Level

Variable	Level	Type	Coding
Personal characteristics			
Psychopathology level (BSI score)	Person	Continuous	0–4
Age	Person	Continuous	18–33
Gender	Person	Dichotomous	0 = Male 1 = Female
Time variables			
Day number	Day	Continuous	1–28
Weekend	Day	Dichotomous	0 = Weekday 1 = Weekend
Design variables			
Survey type	Survey	Categorical	Dummy coded with momentary surveys as reference
Momentary variables			
Missed survey at $t - 1$	Momentary	Dichotomous	0 = Answered survey 1 = Missed survey
PA at $t - 1$	Momentary	Continuous	1–7
NA at $t - 1$	Momentary	Continuous	1–7

Note. BSI = Brief Symptom Inventory; PA = positive affect; NA = negative affect.

SD = 0.57), gender, and diagnosis. People with higher levels of psychopathology perceived that the period of the study was a worse representation of their lives and more burdensome. Moreover, completing all surveys, and knowing the answer to the questions was perceived as more difficult. Last, the study was perceived as less influential on their day-to-day life for male (M = 5.35) than for female (M = 4.87) participants.

Goal 2: Overall Compliance, Momentary Predictors of Compliance, Dropout, and Related Factors

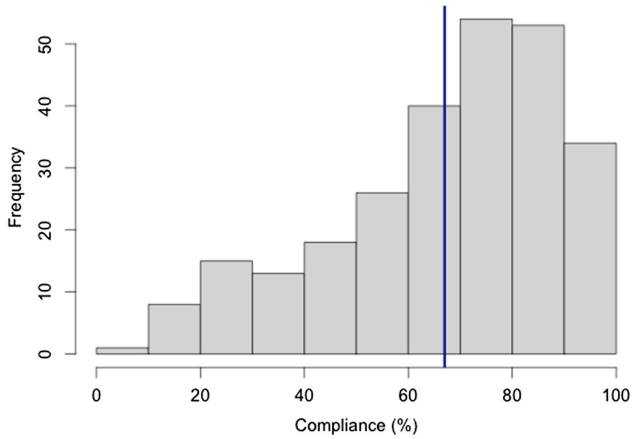
Compliance

Each participant received an average of 236 (*SD* = 57) surveys of which on average 162 were answered (*SD* = 65.7). On average, participants took 1 min and 46 s to complete a survey of any kind. The average compliance level was 67% (*SD* = 21.80) for the whole sample. Figure 2 shows the distribution of compliance per participant for the whole sample. Participants who answered at least 50% of the momentary surveys (n = 192), giving enough data to reliably estimate network models, had an average compliance of 75.77 (*SD* = 12.12). It took 14 min and 20 s on average to open a survey from the moment it was triggered. Compliance did not differ significantly between genders $t(63.50) = 0.22$, $p = .83$, or between people who had versus who had not received a diagnosis of a mental disorder in the past, $t(96.92) = 1.30$, $p = .19$. Last, BSI score was not significantly correlated with compliance $r(260) = -0.04$, $p = .13$.

The compliance dropped across study days as displayed in Panel A of Figure 3. Weekends had lower compliance than weekdays, $t(256.00) = 4.65$, $p < .001$. Moreover, different days of the week had different levels of compliance, $F(61,476) = 13.28$, $p < .01$. Post hoc pairwise t tests showed that Mondays and Sundays were the days with lowest compliance (see Panel B of Figure 3).

A within-subjects comparison with survey as a factor revealed that different surveys also had different levels of compliance $F(3,741) = 321.86$, $p < .01$. The weekly survey had the highest compliance, followed by the evening survey, the morning survey, and the momentary survey. Follow-up paired-samples t tests comparing each type of survey with all other types of survey revealed that compliance for every type of survey differed from all other types of surveys (see Panel C of Figure 3).

Figure 2
Compliance Distribution



Note. The blue line denotes the average compliance. See the online article for the color version of this figure.

Momentary Predictors of Compliance

Table 3 shows a summary of both the univariate and multivariate results of the multilevel logistic regression model predicting momentary compliance. Regarding time-related variables, surveys triggered on later study days and surveys triggered during the weekend were significantly more likely to be missed. These effects remained significant in the multivariate model. Of the personal variables, higher positive affect and missing a survey at the previous time point significantly increased the likelihood of missing a survey, but these effects did not remain significant in the multivariate model. Concerning personal variables, neither gender, age, nor BSI score was significantly associated with the likelihood of missing a survey in either the univariate or the multivariate models. Last, surveys that were more rewarded, contained fewer items, and were triggered less frequently (i.e., morning surveys, evening surveys, and weekly surveys) were less likely to be missed than the momentary survey. However, the effect of the morning surveys did not remain significant in the multivariate model.

Dropout

Figure 4 shows a flowchart of the study's dropout. No significant associations between age (*OR* = 1.07, 95% CI [0.97, 1.18]), gender (*OR* = 2.45, 95% CI [0.89, 8.76]), diagnosis (*OR* = 0.74, 95% CI [0.33, 1.59]), level of psychopathology (*OR* = 1.30, 95% CI [0.72, 2.28]), and dropout were found, and the logistic model was not significant $\chi^2(4) = 4.85$, $p = .29$.

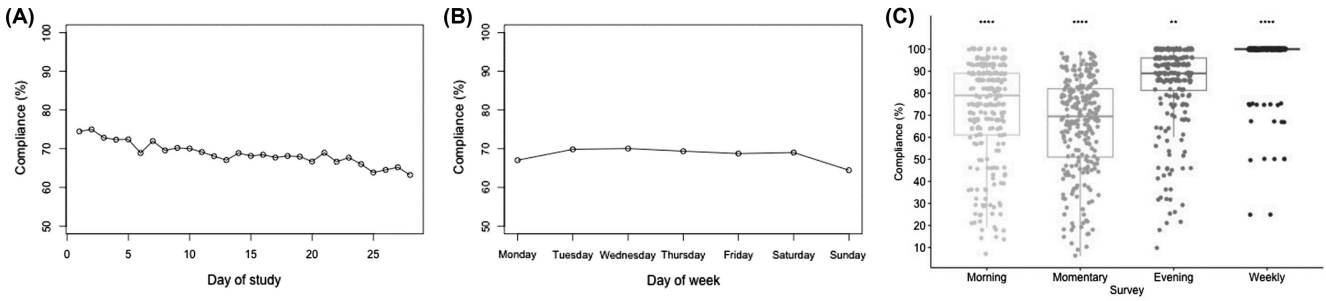
Goal 3: Within- and Between-Individual Variability

Figure 5 provides a visual representation of the items' scores on each metric (mean, between-individual variance, within-individual variance, and ICC) for the quantitative items. Supplemental Table S1 contains tables with the specific scores on these metrics. Note that standard deviations must be interpreted in the context of the scale that was used. A 7-point Likert scale with $M = 4$ and 1 *SD* means that about 68% of the scores fall between 3 and 5, while 95% of the scores fall between 2 and 6.

The means of $std_{between}$ (M = 0.95; range = 0.30–1.59), std_{within} (M = 0.91; range = 0.09–2.42), and the ICC (M = 0.57; range = 0.17–0.93) show that items vary to some extent and a bit more at the between-individuals level. Moreover, the ranges of values show that there is heterogeneity in variability across items. Items with higher std_{within} concerned contextual factors, such as the number of people (1.66), enjoyment of company (2.42), and day appraisal (1.33). Lower std_{within} were found for items about extreme behaviors that were asked with lower time frequency, like the use of laxatives (0.09), purging (0.13), and self-harm (0.17). Interestingly, the items with lowest std_{within} also had the lowest $std_{between}$, that is, the use of laxatives (0.30), purging (0.42), and self-harm (0.38). Items with the highest $std_{between}$ were avoidance of objects (1.59), avoidance of intimacy (1.56), and

Figure 3

Average Compliance per Day of Study, Day of the Week, and Survey



Note. (A) Average compliance per day of the study. (B) Average compliance per day of the week. (C) *t* test comparisons of compliance for the different survey types. Each survey was compared with all other surveys. The compliance of the morning survey also reflects the compliance of the first momentary survey of the day as these coincided. Each dot reflects one participant.
** $p < .01$. **** $p < .0001$.

sex satisfaction (1.56). In general, items about avoidance had high $std_{between}$. Last, the items with low std_{within} and low $std_{between}$ had high ICCs (i.e., use of laxatives, 0.93; purging, 0.91; and self-harm, 0.84).

Figure 6 shows a visual representation of the entropy scores of each item, and Supplemental Table S2 provides tables with the specific scores. The average entropy was 0.41, showing that on average, these items had low variability. However, some items scored above 0.5 on entropy, showing more variability. Similar to the quantitative items, contextual items such as food-related items, people-related items, and activity type had the highest entropy, showing more within-individual variability. The items with the lowest entropies were smoking and taking medication. These items were just triggered for individuals who stated at baseline that they engaged in this type of behavior. Other items with low entropies concerned drug consumption and control loss when eating. The mean standard deviation of the entropies was 0.14, showing that on average the items' entropies did not vary much across participants.

Goal 4: Relationship Between EMA Items and Standardized Questionnaires of Psychopathology

Supplemental Table S3 provides descriptive values for the baseline questionnaires. A graph showing the correlations between the scores on the standard questionnaires of psychopathology and the scores on the EMA items averaged across time points per individual is provided in Figure 7. In general, EMA items correlated with scores on most baseline measures (e.g., feeling rested, concentration, social support, and obsessions). Other EMA items were more disorder specific, such as the avoidance items, sex-related items, or enjoying one's own company.

Notably, some EMA items that were thought to be disorder specific by clinical experts (Jover Martínez et al., 2024) were correlated with many disorder-related baseline measures. For example, body checking was included to capture eating disorder-related psychopathology, and checking information on the internet was included to

Table 3

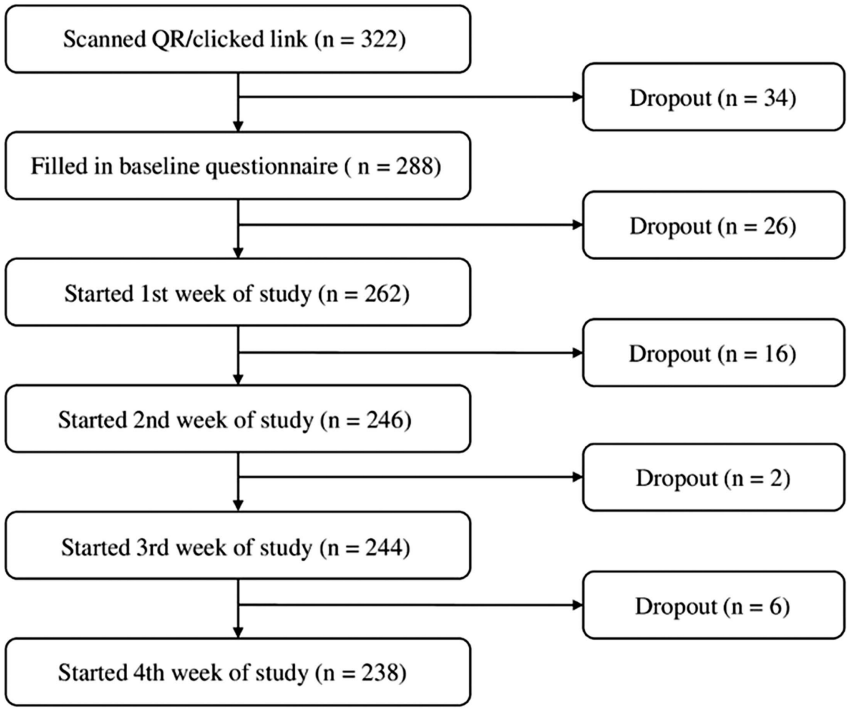
Results of the Momentary Predictors of Compliance

	Individual predictor model			Multivariate model		
Predictor	*OR*	CI	*p*	*OR*	CI	*p*
Intercept				0.17	[0.04, 0.65]	<.009
Study day	1.03	[1.02, 1.03]	**<.001**	1.02	[1.02, 1.02]	**<.001**
Weekend	1.22	[1.17, 1.27]	**<.001**	1.09	[1.02, 1.16]	**.008**
PA $t-1$	1.05	[1.01, 1.08]	**.005**	1.04	[0.997, 1.08]	.052
NA $t-1$	0.96	[0.92, 1.01]	.108	0.98	[0.93, 1.03]	.420
Missed survey $t-1$	2.06	[1.97, 2.14]	**<.001**	1.88	[0.74, 4.76]	.185
Gender	0.87	[0.60, 1.27]	.479	0.89	[0.61, 1.29]	.533
Age	1.00	[0.96, 1.05]	.976	1.02	[0.97, 1.07]	.472
BSI score	1.15	[0.90, 1.48]	.256	1.26	[0.98, 1.62]	.068
Survey						
Morning	0.65	[0.61, 0.69]	**<.001**	0.82	[0.60, 1.13]	.228
Evening	0.26	[0.24, 0.28]	**<.001**	0.17	[0.15, 0.20]	**<.001**
Weekly	0.03	[0.02, 0.05]	**<.001**	0.00	[0.0001, 0.05]	**.015**
Random effects						
σ^2				3.29		
τ_{00}				1.11		
ICC				0.25		

Note. Results in bold indicate significant results. CI = confidence interval; PA = positive affect; NA = negative affect; BSI = Brief Symptom Inventory; ICC = intraclass correlation.

Figure 4

Participants' Dropout Rates per Study Phase

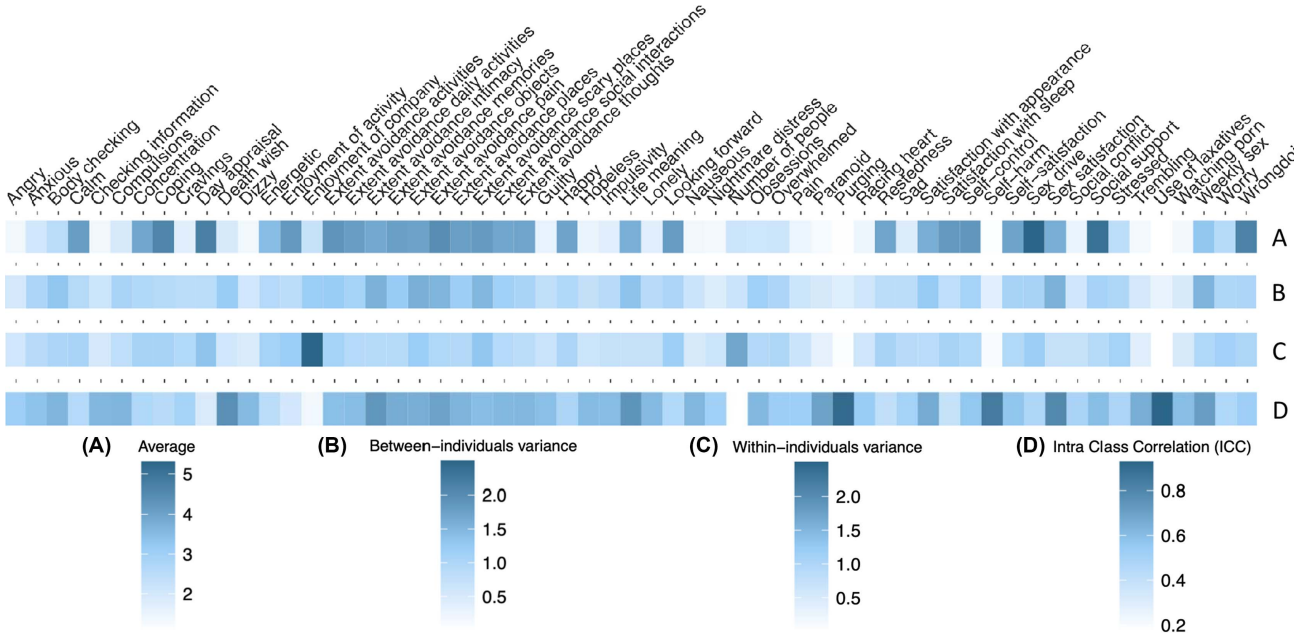


capture anxiety-related psychopathology. However, these items were correlated with most other types of psychopathology.

The EMA items also correlated with scores on most transdiagnostic baseline measures, that is, fear of negative evaluation scale (Brief Fear of Negative Evaluation Scale), dichotomous thinking (DTI), self-control (Self-Control Questionnaire), and personality (Ten-Item Personality Inventory). Similar to the correlations with the psychopathology measures, the avoidance, sex-related, or

Figure 5

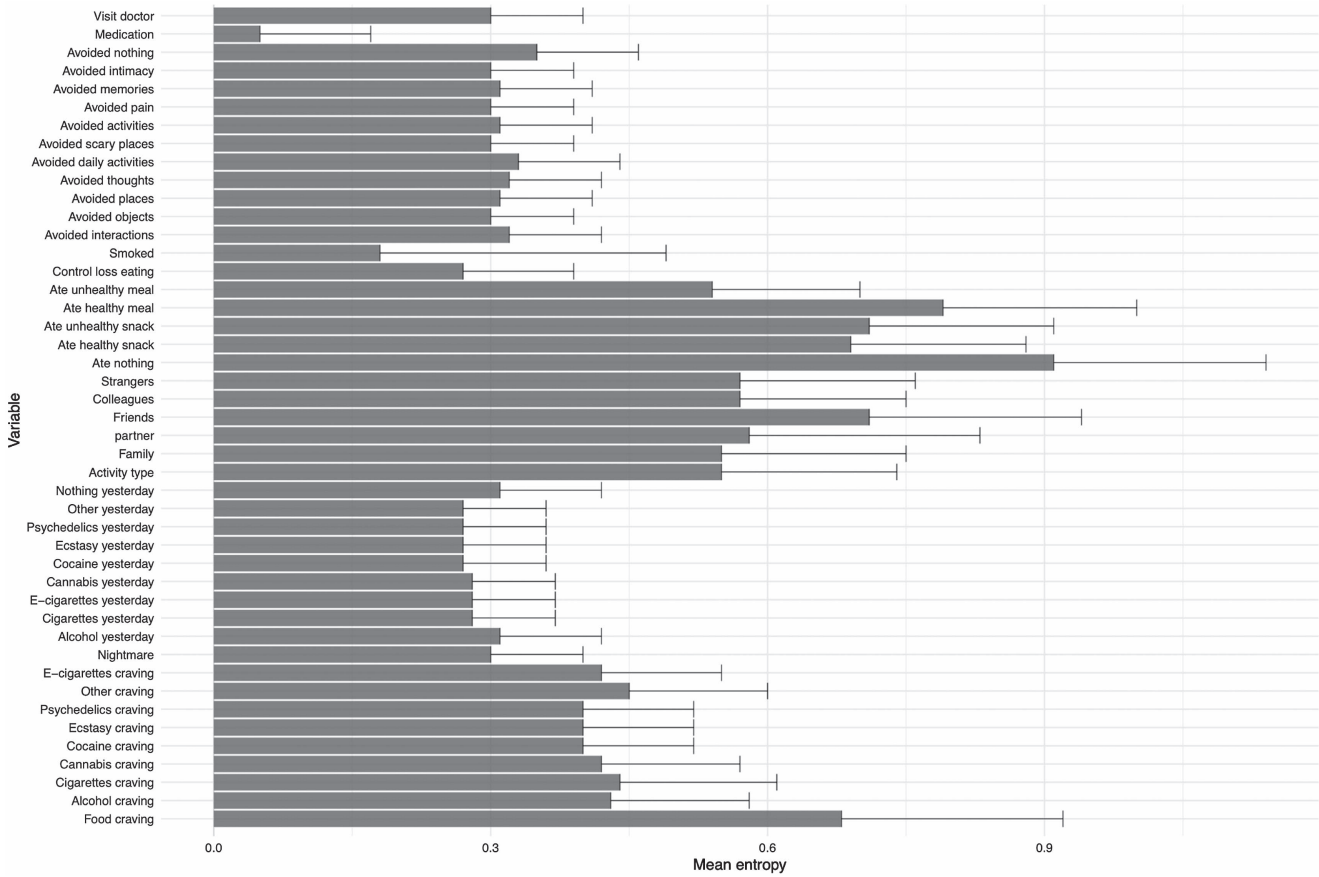
Visual Representation of Average and Variability Metrics of the Quantitative Items



Note. Letters in the rows A–D in the upper part of the figure refer to scales in the columns in the lower part of the figure. See the online article for the color version of this figure.

Figure 6

Average Entropy Levels of Qualitative Items



Note. Bars represent entropy levels and whiskers reflect 1 *SD*.

items regarding enjoying the company of other displayed fewer correlations. While the direction of the correlations was the same for all the baseline measures, it was the opposite for the personality trait measures (i.e., extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience). For example, all the negative mood items correlate positively with all the baseline measures, but negatively with the personality measures, and this pattern is consistent across all EMA items.

Discussion

In the present study, we thoroughly investigated the subjective experience and validity of a transdiagnostic psychopathology EMA protocol. Findings will be discussed per goal.

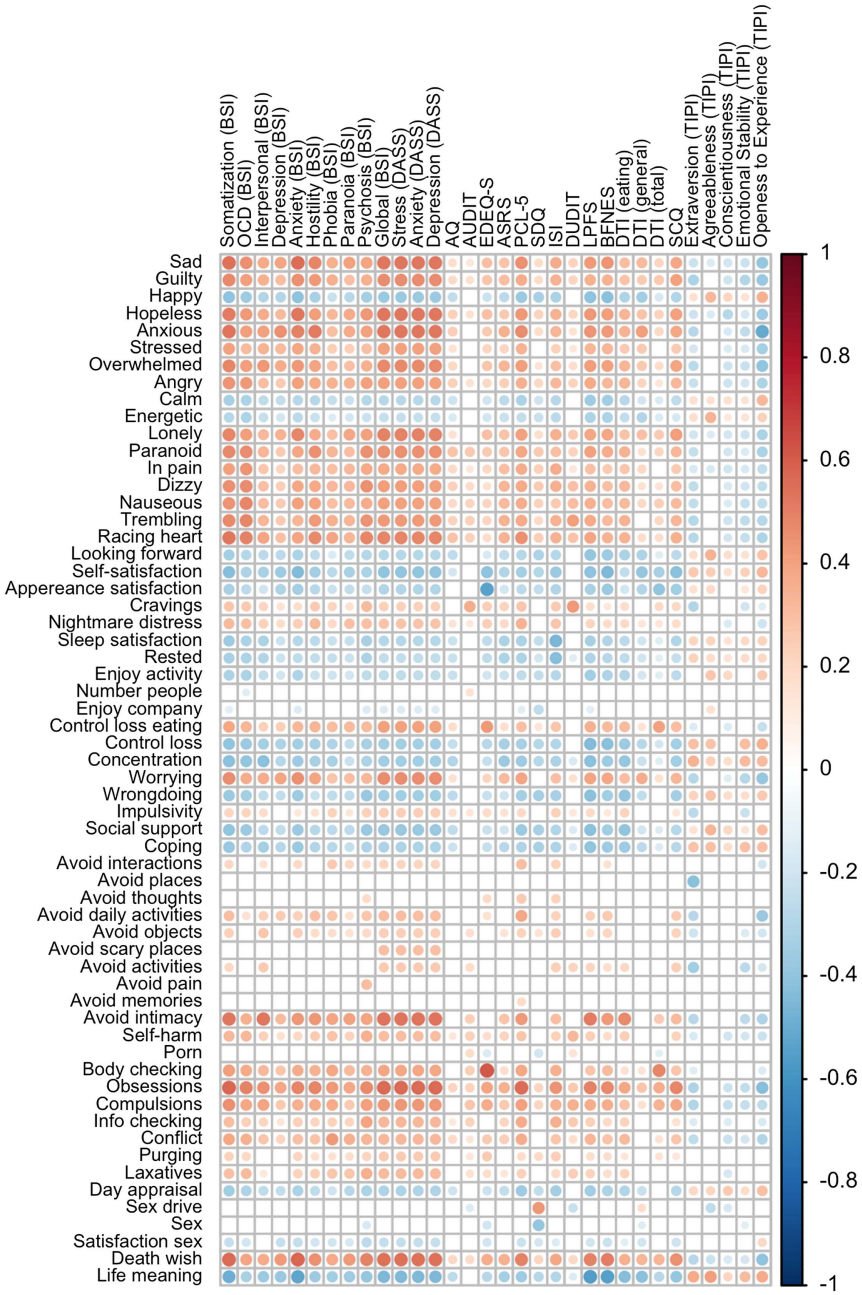
Goal 1: Subjective Experience EMA Protocol

Participants found the surveys neither easy nor difficult to complete, quite burdensome, and a bit frequent. Moreover, the items were perceived as clear, but it was neither easy nor difficult to know the answer. The study was perceived as a fairly good representation of their normal lives and as being fairly influential on it (mostly for female participants). The burden scores in the present study were a bit higher than in a previous study (Eisele et al., 2022), possibly due to our different item set and the longer duration of our study.

People who had received a diagnosis during their lives did not significantly differ from people without a diagnosis on burden-related variables. However, participants with higher levels of psychopathology perceived the study period as a worse representation of their lives and more burdensome and found it more difficult to complete and know the answer to all surveys. These findings align with studies that found that people with a mental health diagnosis are less compliant (Jones et al., 2019; Rintala et al., 2020; Vachon et al., 2019) because lower compliance rates in people with higher levels of psychopathology could be an attempt to reduce burden (Eisele et al., 2022; Stone et al., 2003). However, more research is needed on what specific diagnoses lead to a lower compliance because certain diagnoses, such as major depressive disorder, have been linked to higher compliance (Rintala et al., 2020). Considering that EMA studies will always have some impact on participants, it is concluded that the implemented EMA protocol's burden was acceptable. Future research with populations that are more prone to feel burdened, such as people with higher levels of psychopathology,

Figure 7

Correlations Between Ecological Momentary Assessment Items Averaged Across Time Points per Individual (y-Axis) and Baseline Questionnaires (x-Axis)



Note. Color coding of strength and direction of correlations is displayed in the vertical bar on the right. BSI = Brief Symptom Inventory; DASS = Depression, Anxiety and Stress Scale; OCD = obsessive compulsive disorder; AQ = Autism Quotient; AUDIT = Alcohol Use Disorders Identification Test; EDEQ-S = Eating Disorder Examination Questionnaire–Short; ASRS = Adult Attention-Deficit/Hyperactivity Disorder; PCL-5 = Posttraumatic Stress Disorder Checklist for *DSM-5*; SDQ = Sexual Dysfunction Questionnaire; ISI = Insomnia Severity Index; DUDIT = Drug Use Disorder Identification Test; LPFS = Levels of Personality Functioning Scale; BFNES = Brief Fear of Negative Evaluation Scale; DTI = Dichotomous Thinking Inventory; SCQ = Self-Control Questionnaire; TIPI = Ten-Item Personality Inventory; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, fifth edition. See the online article for the color version of this figure.

should consider alternative ways of reducing such burden, such as increasing the study duration while reducing the frequency of EMA prompts during each day.

Goal 2: Overall Compliance, Momentary Predictors of Compliance, Dropout, and Related Factors

Dropout was not significantly predicted by any of the included variables (gender, age, past diagnosis, and level of psychopathology). Most dropouts (66%) occurred in the first week, suggesting that there may be certain mechanisms at play that were not investigated in this study (e.g., participants who have a low tolerance for repetitive tasks may drop out early). Compliance was quite high, with 67% of the surveys being completed. There are no guidelines about what compliance rate is acceptable in EMA studies. Some studies' compliance rates are as high as 94% (Stone et al., 2003). Note that study characteristics, such as survey frequency, can influence compliance rates (Eisele et al., 2022), but they also determine the type of phenomena that can be examined. Phenomena that unfold more slowly require a lower frequency of assessment, and the opposite is true for faster unfolding phenomena (Wichers et al., 2021). Researchers who wish to use a protocol with a higher assessment frequency must think about the phenomena of interest and if such a high frequency is required for the phenomena of interest.

Personal characteristics such as gender, diagnosis in the past, or level of psychopathology were not significantly associated with dropout or compliance, neither at the study level nor at the momentary level. Females were not significantly more compliant than males, which does not align with a number of studies that observed higher compliance rates in females (Rintala et al., 2019; Vachon et al., 2019; Wrzus & Neubauer, 2023). Note that our sample size was relatively low compared with these previous studies and our sample was predominantly female. Some authors hypothesize that the inconsistent findings regarding gender and compliance might be due to complex interactions between gender and design variables such as assessment schedule and incentives (Wrzus & Neubauer, 2023).

Diagnosis and level of psychopathology do not align with previous research that found different levels of compliance for participants with psychosis or depression in comparison to healthy participants (Rintala et al., 2019, 2020; Vachon et al., 2019). However, the present study did not include diagnosed participants, only information on lifetime diagnosis and BSI global scores was available. This aligns with the results of a meta-analysis of EMA studies, which found that participants with any clinical diagnosis did not have significantly different levels of compliance than healthy participants.

Time variables like study day and weekend *did* have a significant effect on compliance. Specifically, later study days had lower compliance at the study and momentary level, and Mondays and Sundays had lower compliance at the study level, which aligns with earlier research (Rintala et al., 2019, 2020), regardless of the study duration (Vachon et al., 2019; Wrzus & Neubauer, 2023). Moreover, earlier research found Sundays to be the second least compliant day (Rintala et al., 2019). Furthermore, the univariate model at the momentary level revealed that surveys triggered on weekends were more likely to be missed. This is not in line with a study in which weekend days were positively associated with compliance (Rintala et al., 2020). These incongruent findings might be due to the nature of the participants who were included in these studies. For example, people of different ages may spend their weekends differently, affecting compliance. For example, the average age in Rintala et al.'s (2020) study was 32 years old, which is 10 years older than the age of our study's sample.

Survey type also had an effect on compliance at both the study and momentary levels. First, shorter surveys were less likely to be missed, which is in line with earlier research (Eisele et al., 2022). Second, surveys with higher rewards were more likely to be answered, except for the morning survey, possibly because it coincided with the first momentary survey, making it the longest survey. A recent meta-analysis also showed that higher rewards increase compliance (Vachon et al., 2019). Therefore, monetary resources can be used to increase response rates for populations or moments with reduced compliance.

Last, regarding momentary predictors of compliance, in the univariate model, positive affect at the previous time point was significantly and positively associated with the likelihood of missing a survey, which contradicts a previous study in which the opposite was observed (Rintala et al., 2020). Missing a survey at the previous time point was positively associated with the likelihood of missing a survey in the univariate model as well, which is in line with previous research (Rintala et al., 2020). Unlike in the previous study (Rintala et al., 2020), our participants received reminders to answer a survey before it expired and if a survey was missed. These reminders may have tackled the detrimental effects that missing a survey may have on compliance at the momentary level and may have caused effects not showing up anymore in the multivariate models.

Goal 3: Within- and Between-Individual Variability

Overall, our EMA protocol successfully captured experiences that change over time within individuals. All EMA items showed a degree of variability, although there were differences in the extent and type of variability between the items. Some items had very low variability, such as "use of laxatives," "purging," and "self-harm." Differences in item variability may show that items operate at different timescales (e.g., items with low variability may be operating at a slower pace). Therefore, if an item with low variability captures a relevant variable, it may still be interesting to include such an item in intensive longitudinal studies—but at a lower frequency—to see how it relates to other variables despite its low variability.

Network theory (Borsboom, 2017) is silent regarding the timescales at which different phenomena unfold. Importantly, vector autoreggressive models, which are often used in this field of research, can only include variables that are assessed with the same frequency, limiting the usefulness of this statistical method. Recently, Wichers et al. (2021) proposed the momentary affective dynamics network theory, differentiating between macrolevel (i.e., symptoms) and microlevel (i.e., momentary affects) networks, with macrolevel phenomena operating at a slower pace than microlevel phenomena. Wichers et al. (2021) differentiated between symptoms and

momentary states, although some momentary states are listed as symptoms in the *DSM-5*. Note that the items with lowest within-individual variability in our set of EMA items were *DSM-5* symptoms. Moreover, momentary affective dynamics network theory states that persistent interactions between microlevel phenomena can become interactions at the macrolevel (i.e., repeated interactions between momentary affects become symptoms). It would be interesting to statistically model this type of multiple-layer networks.

Goal 4: Relationship Between EMA Items and Standardized Questionnaires of Psychopathology

The EMA items clearly captured psychopathology, but only some items were correlated with disorder-specific types of psychopathology (e.g., "Today my sexual desire/drive was" correlated strongly with the Sexual Dysfunction Questionnaire), reflecting convergent validity. Strikingly, most EMA items were correlated with many scales of psychopathology; the main exception being the Autism Questionnaire and the Alcohol Use Disorders Identification Test, which had fewer and weaker correlations with the EMA items. This may suggest poor divergent validity of the EMA items. A different interpretation is that DSM categories are not as distinct as they are often presented, with many symptoms present in many diagnoses. A recent study investigated criteria repetition in 202 *DSM* diagnoses (Forbes et al., 2024). Only 62 (30.7%) of the included diagnoses did not have any symptom overlap with other diagnoses. Moreover, the symptoms that repeat do so 4.4 times on average and a total of 1,022 instances. Some disorders, such as bipolar and related disorders, are exclusively characterized by symptoms included in other diagnoses too. That means that it is impossible to capture bipolar psychopathology without items included in other disorders.

This overlap between *DSM* categories casts doubts on their validity, especially given the heterogeneity of symptoms among people with the same diagnosis. Such heterogeneity is evident in the vast number of criteria combinations, ranging from almost 24,000 for panic disorder to up to 270 million for comorbid cases like PTSD and major depressive disorder (Allsopp et al., 2019). Whereas using categories that are heterogeneous can be useful for clinicians to diagnose cases that do not fit neatly in homogeneous diagnoses, it may obscure relevant causal pathways. Such pathways could be cross-cutting symptoms or psychological processes, which tap multiple disorders (Allsopp et al., 2019; Forbes et al., 2024).

The EMA items also correlated with the baseline transdiagnostic measures, such as fear of negative evaluation (Brief Fear of Negative Evaluation Scale), dichotomous thinking (DTI), self-control (Self-Control Questionnaire), and personality (Ten-Item Personality Inventory), underlining the transdiagnostic nature of the current set of EMA items and showing the clinical validity of the included transdiagnostic measurement scales. In line with what was observed for the disorder-specific standard questionnaires, it also suggests that these transdiagnostic measures may not be specific enough to distinguish between individuals' psychological problems. Overall, the usefulness of general (*DSM-5*) categories or dimensions might be limited, suggesting that a more fruitful way forward lies in ideographic and transdiagnostic approaches. More specifically, future research may need to focus more on interactions of symptoms at the individual level, and exploring specific constellations of such networks, rather than relying on common disorder categories.

Constraints on Generality

The findings presented in this study may have some constraints regarding generalizability (Simons et al., 2017). First, the generalizability of our findings is constrained to university students. Second, although our sample displayed some level of psychopathology, these findings are not generalizable to clinical samples. Last, although our sample represented a range of nationalities, the majority were European. Therefore, generalizations to non-European populations should be made with caution.

Conclusion

The perceived burden, overall compliance, momentary predictors thereof, and dropout levels indicated that our EMA protocol is a good method to collect time-series data for estimating intraindividual networks, as well as for other analysis methods. However, this protocol still needs to be improved and adapted to different populations (e.g., clinical populations who might need less workload). Moreover, the EMA items showed considerable within-subjects and between-subjects variability, though that was not the case for all items. Measurement frequency should be adapted to the expected timescale at which a phenomenon fluctuates. Future research would benefit from statistical methods that enable the modeling of variables that develop at different timescales. The EMA items' correlation patterns with multiple types of psychopathology reflect the heterogeneity of *DSM* categories and suggest that a transdiagnostic approach might be a better representation of psychopathology.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: The short autism spectrum quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202–212.E7. https://doi.org/10.1016/j.jaac.2011.11.003

Allsopp, K., Read, J., Corcoran, R., & Kinderman, P. (2019). Heterogeneity in psychiatric diagnostic classification. *Psychiatry Research*, *279*, 15–22. https://doi.org/10.1016/j.psychres.2019.07.005

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Medicine*, *2*(4), 297–307. https://doi.org/10.1016/S1389-9457(00)00065-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2002). *Drug use disorders identification test* [Database record]. APA PsycTests. https://doi.org/10.1037/t02890-000

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, *28*(6), 489–498. https://doi.org/10.1002/jts.22059

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Brandon, J. E., Oescher, J., & Loftin, J. M. (1990). The Self-Control Questionnaire: An assessment. *Health Values: Health Behavior, Education & Promotion*, *14*(3), 3–9.

Bruce, S. E. (2009). The NIMH professional coalition for research progress 2008 report. *The Behavior Therapist*, *32*(2), 41–42.

Byrne, S. M., Allen, K. L., Dove, E. R., Watt, F. J., & Nathan, P. R. (2008). The reliability and validity of the dichotomous thinking in eating disorders scale. *Eating Behaviors*, *9*(2), 154–162. https://doi.org/10.1016/j.eatbeh.2007.07.002

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment*, *35*(3), 189–204. https://doi.org/10.1037/pas0001200

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150. https://doi.org/10.1017/S0140525X09991567

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

de Beurs, E., & Zitman, F. G. (2006). De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90 [The Brief Symptom Inventory (BSI): Reliability and validity of a practical alternative to SCL-90]. *Maandblad Geestelijke Volksgezondheid*, *61*(2), 120–141.

Deacon, B. J. (2013). The biomedical model of mental disorder: A critical analysis of its validity, utility, and effects on psychotherapy research. *Clinical Psychology Review*, *33*(7), 846–861. https://doi.org/10.1016/j.cpr.2012.09.007

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, *13*(3), 595–605. https://doi.org/10.1017/S0033291700048017

Duke, D., Krishnan, M., Faith, M., & Storch, E. A. (2006). The psychometric properties of the Brief Fear of Negative Evaluation Scale. *Journal of Anxiety Disorders*, *20*(6), 807–817. https://doi.org/10.1016/j.janxdis.2005.11.002

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). 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

Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., Twose, M., & Dras, M. (2024). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychological Medicine*, *54*(5), 886–894. https://doi.org/10.1017/S0033291723002544

Gideon, N., Hawkes, N., Mond, J., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and psychometric validation of the EDE-QS, a 12 item short form of the Eating Disorder Examination Questionnaire (EDE-Q). *PLOS ONE*, *11*(5), Article e0152744. https://doi.org/10.1371/journal.pone.0152744

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Infrasca, R. (2011). Sexual Dysfunction Questionnaire: Scale development and psychometric validation. *Giornale Italiano di Psicopatologia/Italian Journal of Psychopathology*, *17*(2), 253–260. https://psycnet.apa.org/record/2013-05331-012

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction*, *114*(4), 609–619. https://doi.org/10.1111/add.14503

Jover Martínez, A., & Guðmundsdóttir, G. R. (2024). *validationEMAprotocol*. https://osf.io/hq6fn

Jover Martínez, A., Lemmens, L. H. J. M., Fried, E. I., & Roefs, A. (2024). Developing a transdiagnostic ecological momentary assessment protocol for psychopathology. *International Journal of Methods in Psychiatric Research*, *33*(3), Article e2028. https://doi.org/10.1002/mpr.2028

Jover Martínez, A., Roefs, A., Lemmens, L. H. J. M., & Fried, E. I. (2023). *The student mapping study*. https://aspredicted.org/ej6jp.pdf

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry & Neuroscience*, *31*(1), 13–20.

R Core Team. (2016). *R: A language and environment for statistical computing (4.2.2)*. R Foundation for Statistical Computing. https://www.R-project.org/

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, *31*(2), 226–235. https://doi.org/10.1037/pas0000662

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research*, *286*, Article 112896. https://doi.org/10.1016/j.psychres.2020.112896

Roefs, A., Fried, E. I., Kindt, M., Martijn, C., Elzinga, B., Evers, A. W. M., Wiers, R. W., Borsboom, D., & Jansen, A. (2022). A new science of mental disorders: Using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. *Behaviour Research and Therapy*, *153*, Article 104096. https://doi.org/10.1016/j.brat.2022.104096

Schreuder, M. J., Groen, R. N., Wigman, J. T. W., Hartman, C. A., & Wichers, M. (2020). Measuring psychopathology as it unfolds in daily life: Addressing key assumptions of intensive longitudinal methods in the TRAILS TRANS-ID study. *BMC Psychiatry*, *20*(1), Article 351. https://doi.org/10.1186/s12888-020-02674-1

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A Proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain*, *104*(1), 343–351. https://doi.org/10.1016/S0304-3959(03)00040-X

Ustun, B., Adler, L. A., Rudin, C., Faraone, S. V., Spencer, T. J., Berglund, P., Gruber, M. J., & Kessler, R. C. (2017). The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for *DSM-5*. *JAMA Psychiatry*, *74*(5), 520–527. https://doi.org/10.1001/jamapsychiatry.2017.0298

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research*, *21*(12), Article e14475. https://doi.org/10.2196/14475

Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale–Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, *13*(1), 3–14. https://doi.org/10.1002/pmh.1434

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021). A narrative review of network studies in depression: What different methodological approaches tell us about depression. *Frontiers in Psychiatry*, *12*, Article 719490. https://doi.org/10.3389/fpsyt.2021.719490

World Health Organization. (2001). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care*.

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A Meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Received January 29, 2024
Revision received August 24, 2024
Accepted August 29, 2024 ■