Generating Feedback Reports for Ecological Momentary Assessment Data

Rimpler, Aljoscha

Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen

Generating Feedback Reports for Ecological Momentary Assessment Data

Aljoscha Rimpler

Research Master Thesis *Clinical and Health Psychology*
Date: 05.08.2022
Student number: 2819872
Supervisor: Eiko I. Fried
Second reader: Niki Antypa

Abstract

Ecological Momentary Assessment (EMA) is a data collection method that utilizes phone apps to gather data in daily life. EMA has many advantages, such as ecological validity. However, data collection protocols are often intense, with multiple measurements per day, which can interrupt participants' everyday activities and place a burden on them. This can reduce compliance. One way to tackle this is to provide participants with personalized data reports as an intrinsic reward. However, current frameworks to generate such reports are focused on single individuals in treatment, and not suitable for large-scale studies. Here we introduce a software to fill this gap, FRED (Feedback Reports on EMA Data), and showcase FRED by generating reports for 428 participants who took part in the WARN-D study. Participants were followed for 85 consecutive days, and received four daily and one weekly survey, resulting in up to 352 observations. We provided feedback to participants in the form of downloadable HTML-files, which were generated using the R programing environment. Reports included descriptive statistics, time-series visualizations, and network analyses on selected variables. Furthermore, we assessed participants' perceptions of the created reports (n=54), who judged reports mostly as understandable, insightful, and that reports resonated well with them. Given that FRED is flexible and can be adjusted to the needs of a particular research project, it provides a good basis to generate large numbers of personalized data reports.

Keywords: Ecological Momentary Assessment, Experience Sampling Method, Personalized Feedback

Laymen's Abstract

It gets increasingly more common to collect data through multiple smartphone surveys per day over a prolonged time frame. While this has various advantages, such as observing behavior, emotions, and situations in a natural setting, it places a burden on participants. This is challenging since researchers want to collect as much data as possible to get a good understanding of the participants. A way to motivate participants to complete surveys is to reward them with personalized feedback on their own data. However, software is needed to do this efficiently for hundreds of participants. In the current research, we developed software that provides researchers with the possibility to generate large numbers of personalized data reports. We followed 428 students in the context of the larger WARN-D project over 85 consecutive days. Participants received 4 daily surveys plus 1 weekly survey. This resulted in up to 352 observations for each participant. The software we developed enables researchers to create downloadable reports. These reports entailed summary statistics, the development, and relationships between selected variables. We also assessed how participants perceived their personal report in an additional online survey. In total, 54 participants completed this survey. Participants perceived the personalized data reports mostly as understandable, insightful, and resonated well with the generated reports. Based on this research we conclude that the software we developed provides a good basis to generate large numbers of personalized data reports. Furthermore, it is flexible and highly adjustable to the needs of other research projects.

Generating Feedback Reports for Ecological Momentary Assessment Data

With the rise of digitalization and new technologies, new ways of data collection are becoming more feasible. One common, state-of-the-art way of data collection are ecological momentary assessment (EMA) methods. Characteristic for these methods is that data is collected by sending surveys to participants´ smartphones while participants are in their everyday environment (Shiffman et al., 2008). Typical collected data are self-reported moods, thoughts, situational factors, behaviors, or symptoms (Ebner-Priemer & Trull, 2009). EMA is used in research settings (Shiffman et al., 2008) as well as therapeutic settings (Bringmann et al., 2021; von Klipstein et al., 2022). One particular challenge EMA studies face is participant compliance since repeated measurements can interrupt everyday activities and place a burden on participants (Rintala et al., 2019). There are multiple ways to increase compliance by EMA study designs (e.g., random vs fixed survey schedules). However, the choice of study design is usually based on characteristics of the situations, behaviors, and moods that the study intends to capture (Doherty et al., 2020). Therefore, the current research focuses on a different approach- how to give participants insights into their own collected data to potentially increase compliance. This is not trivial as there are many methodological, ethical, and technical challenges to overcome, and no guidelines are yet available.

In the current work we present FRED. FRED is an acronym for Feedback Reports on EMA Data. It is a tool we developed to generate personalized feedback reports for 428 research participants. We used FRED to generate reports for an observational, ongoing study. Participants received reports in a non-therapeutic and uncontrolled setting. Thus, we erred on being too cautious in terms of variable selection, and not inflicting harm on participants.

Widespread use of EMA

EMA is an umbrella term that encompasses a variety of methods that differ in observation frequency, and the type of collected data e.g., self-report measures vs. sensory data (Shiffman et al., 2008). The application of EMA is wide fold and not limited to specific research settings. Examples of EMA use include, as mentioned earlier, psychotherapy (Bringmann et al., 2021; von Klipstein et al., 2022) but also research about student stressors during a lockdown due to COVID-19 (Fried et al., 2022), substance-use (Bertz et al., 2018), or the linkage between heart conditions and stress (Fanning et al., 2020).

The widespread use of EMA directly relates to its advantages. These include reduced recall bias through real-time assessment, the possibility to assess dynamic and complex processes through repeated measures per day, and the identification of person-environment interactions and individual stressors (Bringmann et al., 2021; Ebner-Priemer & Trull, 2009; Hamaker & Wichers, 2017; Leertouwer et al., 2021). As a result of the aforementioned factors, EMA increases ecological validity, which is a measure of how generalizable findings are to real-life settings (Mestdagh & Dejonckheere, 2021).

Compliance in EMA Research

Participant compliance is a particular challenge for EMA, as the burden for participants is high due to the repeated assessments per day (Rintala et al., 2019). Missing assessments may bias the captured experiences and behaviors in the sample (Shiffman et al., 2008). There are good reasons to assume that many missing assessments are not random (Leertouwer et al., 2021) because many experiences may depend on context (Shiffman et al., 2008). For example, participants may be less likely to complete a survey when they are in a bar with friends compared to when they are home alone. As a result, certain types of thoughts, feelings, behaviors, or situations can be systematically underrepresented in the data.

Increasing participant compliance can be achieved via different methods. These can be broadly categorized in three areas: a) study design, b) extrinsic rewards and c) intrinsic rewards (Doherty et al., 2020). First, study design choices are often limited due to methodological considerations. An example is fixed prompt schedules vs random prompt schedules. Fixed schedules often have higher response rates, but participants tend to adjust their behavior to these schedules (Rintala et al., 2019). Second, extrinsic rewards such as monetary reimbursement have been shown to be effective but participants might again adjust their behavior to the study (Doherty et al., 2020). Thus, high reimbursement is of limited use since it may affect the nature of observations. Furthermore, financial reimbursement for long studies is often limited due to financial constraints for research projects. Third, Hsieh et al. (2008) and Bälter et al. (2012) showed that intrinsic rewards such as real-time feedback on collected data are an effective way to increase compliance. Another way to increase compliance via intrinsic reward could be to provide personalized data reports to participants after study completion.

Challenges for Personalized Data Reports

The current work is about the implementation and assessment of personalized data reports as incentive for participation in our study. We define personalized data reports as a method to give participants access to summaries and analyses on their own data. There are many ways to create feedback for participants, and several ethical, methodological, and technical decisions must be made, which we briefly touch on in the following.

The first ethical consideration is the setting. Will participants receive a report in a supervised and plannable setting, or will they not be guided? This directly leads to the selection of variables on which participants receive a report. In an online research setting such as our study where participants do not receive the report under supervision, special care must be taken to ensure that participants do not receive potentially harmful information. While there is a literature showing that assessing e.g., suicidality does not increase suicidal behavior (DeCou & Schumann, 2018), van Helvoort et al. (2020) could show that receiving false feedback can lead to an exaggeration of symptoms. Thus, variables assessing clinical constructs, such as suicidality, are difficult to include in a report in an unsupervised setting because information such as "You have been at high risk for suicide in the past three months" could lead to potential harm to individuals when misinterpreted. We used an iterative process for variable selection, which is further elaborated in the method section.

One methodological challenge is the study design. Is the study an observational or an intervention study? Providing feedback can not only be harmful but could also change participants' behavior and thus serve as an unwanted intervention. For example, if participants receive feedback about the association of negative mood with their social media use, they might change their usage behavior. We solved this through the selection of variables included in the reports.

Another decision is whether one wants to show raw data vs. already analyzed data. While analyzed data has the advantage of giving different insights, e.g., correlations are identified, knowledge is needed to be able to understand analyses (Bringmann et al., 2021). As not all participants may have this knowledge, analyses need to be explained thoroughly to ensure that participants gain the desired insights. The selection of analyses is an important consideration and can range from mere descriptive statistics of mood states to complex multivariate statistical methods. For a personalized report to be rewarding, it is important for participants to gain insights from these reports. One promising approach for participant insights are time-series analyses, since they can detect relationship patterns between variables or trends over time. In time-series analyses, longitudinal data is analyzed on how good observations can be predicted by preceding observations (Hamaker & Wichers, 2017). A particular statistical method to analyze time-series data is the vector autoregressive (VAR) model (Hamaker et al., 2015), which shows relationships between different variables. These can be conceptualized as networks. Relevant for the current work is the so called contemporaneous network in which relationships that occur at the same time are estimated (Epskamp et al., 2018). If the results are translated adequately back to participants, this method could yield personal insights for participants. The created reports entail descriptive statistics, time-series visualizations, and network analyses. Elaborations on the analyses can be found in the method section.

The last point to consider is of technical nature. One needs to decide the medium through which the feedback is delivered. Options include report files, such as HTML or pdf- files, or interactive apps (e.g., Bringmann et al., 2021). If one decides for report files one must consider how to distribute these files. Options include to make files downloadable from a website or to send the reports via email. In this case we opted for downloadable HTML files with encrypted file names to ensure data security.

Yet another challenge is that tools for feedback generation are available for clinical and supervised settings, e.g. Bringmann et al., (2021) and Bos et al., (2022), however tools for feedback generation in large samples are to our knowledge not freely available yet. While it is possible to manually create personalized feedback in small samples, it is neither feasible nor practical to manually create a large volume of individual reports. The reason is the high workload and potential errors associated with the manual generation of many reports.

To automatically generate many reports and conduct individualized analyses, the R programming environment (R Core Team, 2022) can be used. Creating a code to handle datasets from hundreds of participants to create individualized data reports meets two major challenges. First, missing data can cause the R code to break, since analyses on non-existent data cannot be performed. The second challenge is writing code that adequately returns valid analyses across a variety of different numbers of observations, since assumptions for analyses might depend on sample size and data distribution.

Current Study

The purpose of the current work is twofold. The first objective was to develop a tool which allows the communication of EMA data to study participants in a way that is understandable but also insightful, while navigating ethical, methodological, and technical challenges. The guiding question is "How to utilize collected EMA data as motivation and reward for a large sample of research participants?" This tool is developed in the R programming environment. An iterative process of report generation and discussion of these reports in our team was used to come up with the final reports.

The second objective is to assess participants' perception of the generated reports to further improve it for future use. To assess participants' perception of the personalized data report, a feedback survey on this matter was sent out to all participants who wanted to receive a personalized data report.

Methods

Sample

This work is part of the five-year WARN-D project. The goal of WARN-D is to follow 2000 students, divided in 4 cohorts of 500 participants, over a period of two years to gain a better understanding of stressors and experiences students face that could ultimately lead to mental health problems. The research is divided in three stages: a baseline survey, an EMA phase, and a follow-up phase. In this thesis, data from the EMA phase of the first cohort of WARN-D participants is used. In this phase participants were asked to complete four questionnaires per day over the span of 85 days and an additional questionnaire on Sundays. Furthermore, participants were provided with a Garmin vivosmart 4 smartwatch to track activity data; these data will not be analyzed for the present thesis.

To participate in the WARN-D study, students needed to be at least 18 years old, be enrolled at a Dutch university or Dutch university of applied science (“HBO”, “MBO”, “WO” degrees), read fluent English or Dutch, have a European bank account, and have a functional Android or iOS smartphone. Students were excluded if they failed to fulfill self-reported screening criteria for schizophrenia/psychosis/thought disorder, major depressive disorder, mania/bipolar disorder, substance use disorder, suffered from moderate to severe suicidal ideation, or stated that seeing an estimate of burned calories would stress them. The study protocol of WARN-D was approved by the ethics committee of the European Research Council and the Psychology Research Ethics Committee at Leiden University (No. 2021-09-06-E.I.Fried-V2-3406).

In total, 448 participants were recruited for the first cohort. Recruitment took place via posters, email-newsletter, social media (Facebook, Instagram, Twitter), and word-of-mouth. Participants were reimbursed for participation and could receive 7.50 € for completing the baseline and up to 45 € for completing the EMA phase. The reimbursement for the EMA phase depended on the number of completed surveys. Out of the 448 participants that completed the baseline, 428 started the EMA phase. In the first survey of the follow-up stage, participants were asked whether they want to receive a personalized data report and 398 participants indicated that they want to receive one.

Procedure

Participants used the Ethica data app for Android (Ethica Data Services Inc., 2021a) or iOS (Ethica Data Services Inc., 2021b) on their smartphone to receive four prompts per day at semi-random times. Each prompt was available for 20 minutes. On Sundays, an additional survey was prompted which included questions about the previous week and was available for 10 hours. This resulted in both daily and weekly survey patterns. Participants received survey prompts from December 6, 2021, through February 28, 2022. Meta-information about the different surveys can be found in Appendix A. Surveys were available in Dutch as well as in English and languages could be switched in the Ethica app. Participants received a Dutch or English report based on their indicated language preference.

Development of FRED - Feedback Reports on EMA Data

In the following the methods relevant for the development of FRED is described first. The methods regarding the evaluation survey to assess participants' perception of the personalized data reports is described second. The first part comprises a section about our approach a) to overcome technical- challenges, b) to overcome ethical- and methodological-challenges, c) to increase participants' understandability of analyses. The last section contains d) the methods of relevant analyses. The second part entails e) information about the evaluation survey and f) analyses of the evaluation survey.

Approaching Challenges Regarding Technical Issues

FRED is software that can be used by researchers to generate large numbers of personalized feedback reports for participants in an EMA study.

The personalized data reports were delivered as downloadable HTML files. To communicate various statistical results, we mainly used data visualization in the reports. Each figure presented in the data report is accompanied by explanatory text. The reports are organized in multiple segments, varying in content and complexity. Sections range from mere descriptive statistics through time series visualizations to network models. One challenge to write functionable code is the unique data availability per participant.

FRED is written in a way that if an analyses cannot be conducted for an individual participant it gets skipped for this participant. The same applies to respective text and structure elements. FRED was developed to provide a framework to generate a large amount of personalized data reports, not a standalone tool. This means the developed code needs to be adjusted for other projects.

FRED is based on two interacting R files. One file is an R script, which contains all the processing steps that need to be performed for everyone. Additionally, this script includes code that creates the individual reports. The code to create the actual content of the individual reports is in the second file, an R markdown file, which works on individualized data subsets. The R code is adjustable to meet the purpose of other feedback reports (e.g., incentive vs clinical report).

To properly function, there are two different data frames needed as input. Both data frames include all data of the full sample. One needs to be in long format (each item per participant has its own row), the other in semi-long format (each prompt has its own row, but multiple items per row). An overview of the used R packages can be found in Appendix A.

Approaching Challenges Regarding Ethical and Methodological Issues

Researchers often collect a lot of data, but it is not possible to provide participants with all possible information because of the guiding principles mentioned above. The reports should cover enough information to be insightful for participants, but at the same time, the information provided should not be overwhelming. Therefore, a variable selection was required. An overview of all assessed variables can be found in the codebook following https://osf.io/frqdv/. We used an iterative process to determine the final selection of reported variables. First, we discussed potential variables, in the next step we started drafting reports, which concluded in discussing these reports and selected variables again. The variables that we selected to give feedback on can be found in Appendix B.

We used several guiding principles to make the reports as accessible and interpretable as possible. First, we tried to explain data visualizations as simple as possible, since the sample also includes participants that do not have a lot of prior knowledge of statistics and data interpretations. For the same reason, a second consideration was to present the data in a way that is least subject to false interpretation and difficult explanations. To achieve this goal, all reported bar graphs showing relative frequencies, had the same limits of 0 and 100%. Additionally, when visualizing time series, participants were not shown trend lines but only raw data points if they had completed less than 25% of the surveys. Lower reliability of these trendlines at the periphery was explained to the other participants. Only contemporaneous networks were included to reduce the misconception of causality and correlation. Furthermore, the explanation of the networks was simplified as much as possible by using common terms such as “relationships” instead of “partial correlations” and by omitting specific numerical information.

Analyses

To provide participants with valuable insights, a variety of decisions about analyses needed to be taken. The personalized data reports consist of summary statistics for a) continuous variables and b) categorical variables. Furthermore, they include c) time-series visualizations which include, depending on data availability, trend lines. Additionally, for some participants d) network visualizations on different items were shown. The few analyses that were not visualized but rather reported as text are single values such as averages of affect for a particular week, or days and prompts on which participants had the most positive affect.

For the visualization of trends in time-series, LOESS functions were used. LOESS is an acronym for locally estimated scatterplot smoothing and is a local regression. The main advantages of using LOESS are that the relationship between dependent and independent variables do not need to be specified beforehand and that it is able to depict different types of relationships (Jacoby, 2000). This makes LOESS functions useful for these idiographic data-reports. Trends were only estimated for participants having 25% or more surveys completed, since these trends are more reliable when more data is available. Smoothing was accomplished with ggplot2’s geom_smooth function. Important for smoothing is finding a parameter that does not over-smooth the data (trends are not visible anymore) but at the same time does not overshoot (showing trends that are not backed by the data). This is even more challenging since smoothing depends on data availability. However, setting individual parameters per participant possesses the challenge of generalizable explanations, which is why smoothing parameters were the same across participants, but different for different figures. Finding a suitable smoothing parameter was done by visual inspection of figures for participants with very little, medium, and very high data availability.

Two different contemporaneous networks based on partial correlations (Epskamp et al., 2018) were estimated for participants having completed more than 50% of the surveys. Mansueto et al. (2022) implied that the number of surveys are sufficient to estimate networks in a clinical context. We needed to select variables on which we wanted to compute the networks. A core feature of suitable variables is a normal distribution. The items for the networks were selected according to the normal distribution of these variables according to the Shapiro-Wilk test. We calculated the Shapiro-Wilk test for each variable across participants and chose the ones with the least deviation from normality. To decrease overfitting of the network- models, we used the least absolute shrinkage and selection (LASSO) parameter λ. Furthermore, we used the tuning parameter γ (Chen & Chen, 2008) to control the sparseness of the networks. We chose a γ of 0 to have fuller networks. For further information on these parameters see Appendix A. Since the study period of 3 months might have caused non-stationarity, we detrended a linear trend in the data by using residuals (Mansueto et al., 2022). Because the combination of a Kalman filter and LASSO has been shown to perform well with similar amounts of observations (Mansueto et al., 2022) this combination of methods was chosen for the network analyses.

Evaluation- Survey

To assess the perception of the personalized data reports, every participant who was provided with a personalized data report received a link to an evaluation survey in Qualtrics (Qualtrics, 2022) along with the report. This survey was designed to cover the most important aspects of data reports, without changing the character of an observational study to an intervention study. Items covered different answer types such as Likert-scales, open text answers, and a ranking. The items and answer scales can be found in Appendix C. The feedback survey was not reimbursed.

Analyses of Evaluation Survey

To get insights into the perception of reports generated with FRED and how to further improve it we analyzed the feedback survey. We included various descriptive statistics of the feedback survey as well as an analysis of the relationship between completed surveys and the insightfulness of the report. In the feedback survey we asked participants to rank the different sections of the report according to their interestingness. Additionally, the open text entries will be qualitatively described.

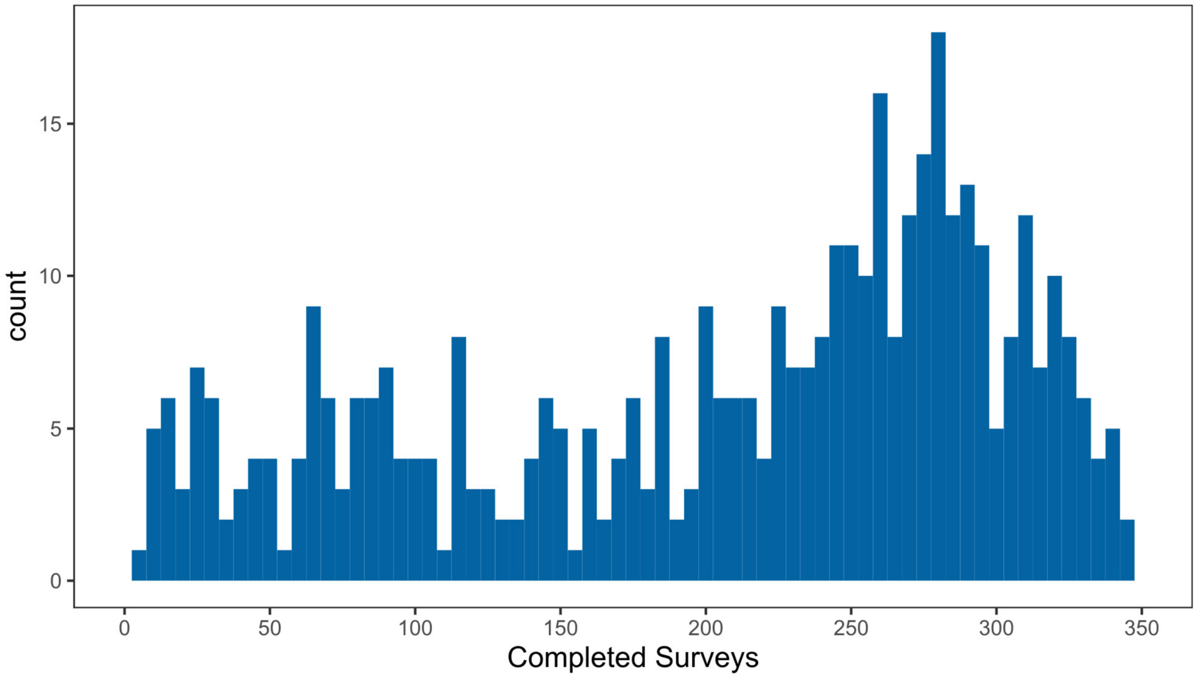
Results

Data Availability

Of the total 352 EMA timepoints, participants on average completed 58 % (N = 204) of the surveys, and the median of completed survey was 67 % (N = 236). The range was wide, with 1 % to 99 % of surveys completed across participants. Figure 1 shows the distribution of data availability per participant. The bimodality of the distribution likely stems from a sizeable number of participants dropping out around Christmas (after 72 surveys).

Figure 1

Distribution of Data Availability



Note. Distribution of data availability. This figure shows the distribution of number of completed surveys. The x-axis shows the number of completed surveys and the y-axis shows the counts in bins of width 5.

Sections Included in the Final Personalized Data Reports

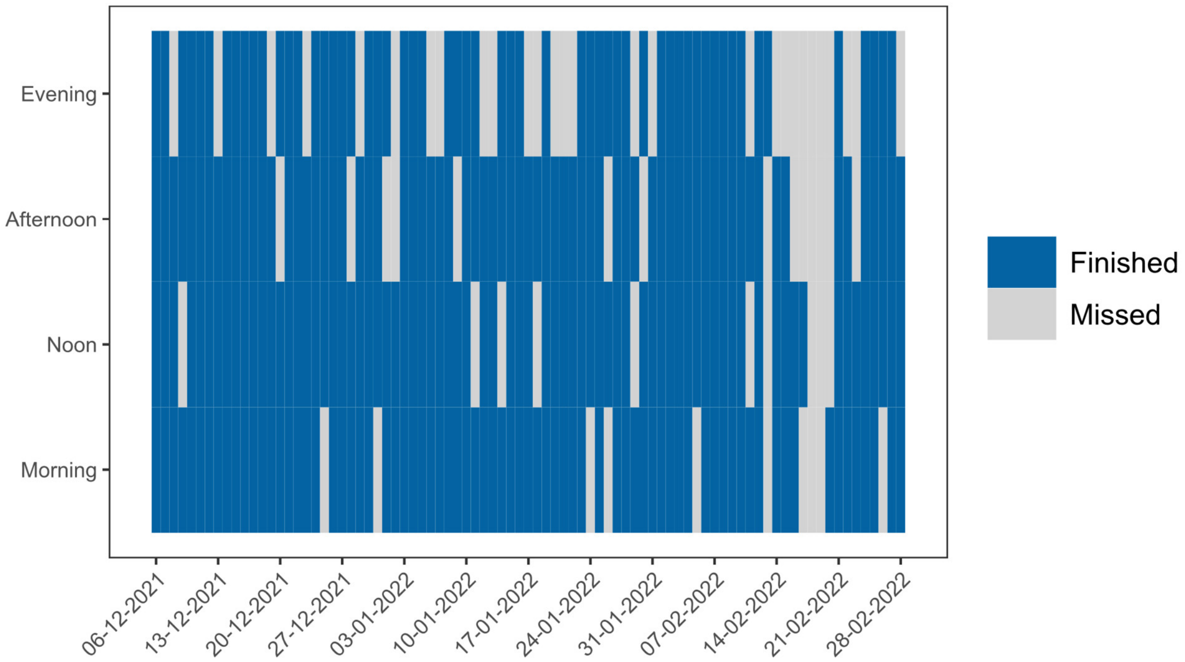
A full version of an example report can be found following this link https://osf.io/hgams/?view_only=0e60de0d8a6a4f6f9b5713870209ff4f. This example report is based on actual data, but the pseudonym was changed. The report starts with explanations what the report entails and how to read it. This is followed by the core content, which we describe in the following.

Report Section 1: Completed Surveys

In this section, three types of information are given to participants. First, we present the percentage of overall completed surveys. For our example participant, this sentence reads: "Looking into your data, you completed 82.4 % of the surveys we sent you 4 times per day." Second, we present the percentages of completed Sunday surveys, which reads for this participant "Of the 12 weekly surveys we sent you on Sundays, you completed 100 %." And third, we present a figure that indicates which particular daily surveys were completed. This figure is a heatmap with two colors, indicating which surveys were completed or missed. An example plot can be found in Figure 2. All example plots shown in the current work are from the previously mentioned report.

Figure 2

Example of a "Completed Surveys" Figure



Note. Example figure of Completed Surveys, indicating which of the daily surveys have been completed (blue) or missed (grey).

Report Section 2: General Summary

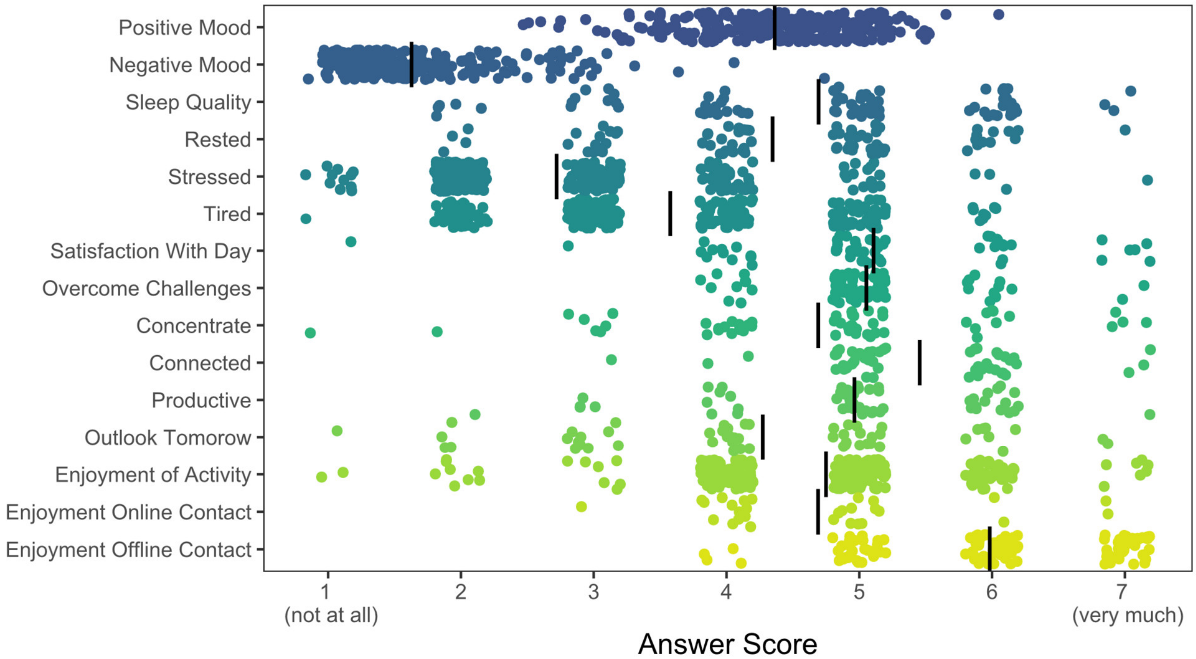
This section gives an overview of the raw data of selected, continuous variables, and additionally indicates the mean of these variables. Furthermore, participants receive information on which day, and on which prompts (morning, noon, afternoon, or evening) they had the most positive affect on average.

Individual raw data points are jittered, so that it is possible to see how often a participant selected a certain answer on this item rather than plotting the data points exactly over each other. Figure 3 shows an example plot for the "General Summary".

The information on which day and prompt participants had the most positive affect on average is presented as text, e.g., "On average, you had the most positive mood on *Sundays*. Out of the four surveys we asked you per day, on average, you had the most positive mood in the *morning* surveys." for a person having the highest average positive mood on Sundays and morning surveys.

Figure 3

Example of a "General Summary" Figure



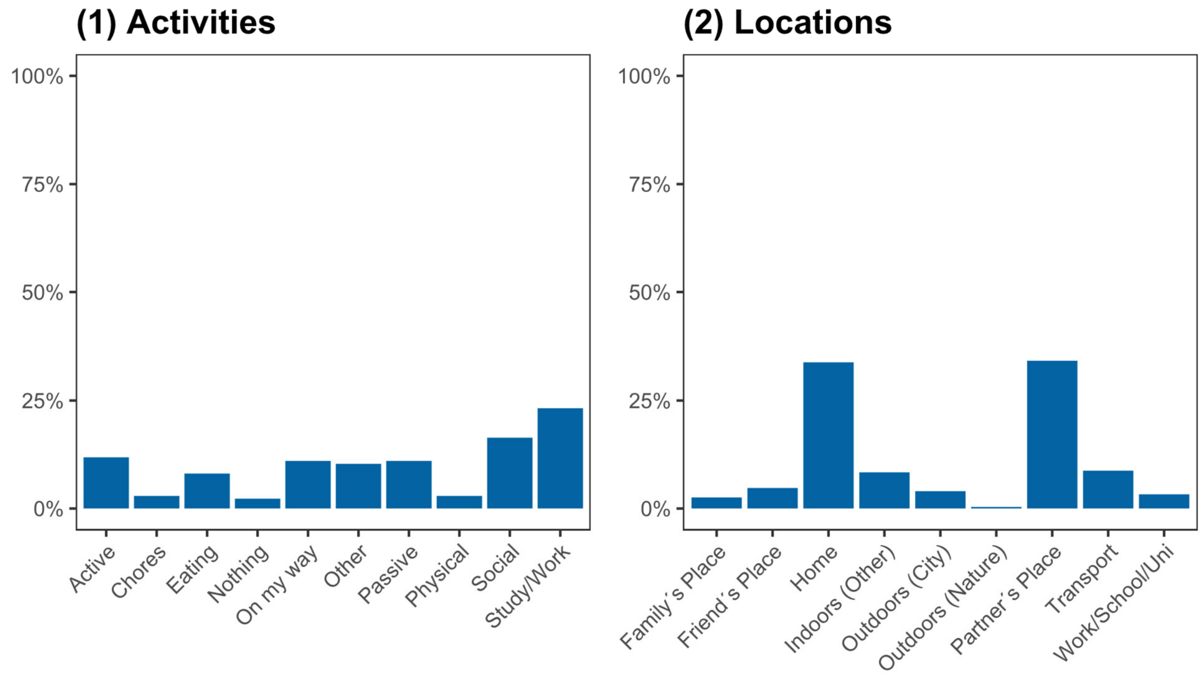
Note. Example figure of the General Summary. Selected variables are shown on the y-axis and answer scores on a Likert Scale 1 (not at all) to 7 (very much) are shown on the x-axis. This person had on average a higher positive mood than negative mood. The highest average score was obtained for the enjoyment of offline contacts. The mean on the respective item is indicated as a black vertical bar.

Report Section 3: Situations and Activities

Section 3 summarizes contextual items that refer to locations, online contacts, offline contacts, and participants' activities, see Figures 4 and 5. Items were categorical and multiple answers were allowed, besides for the item location

Figure 4

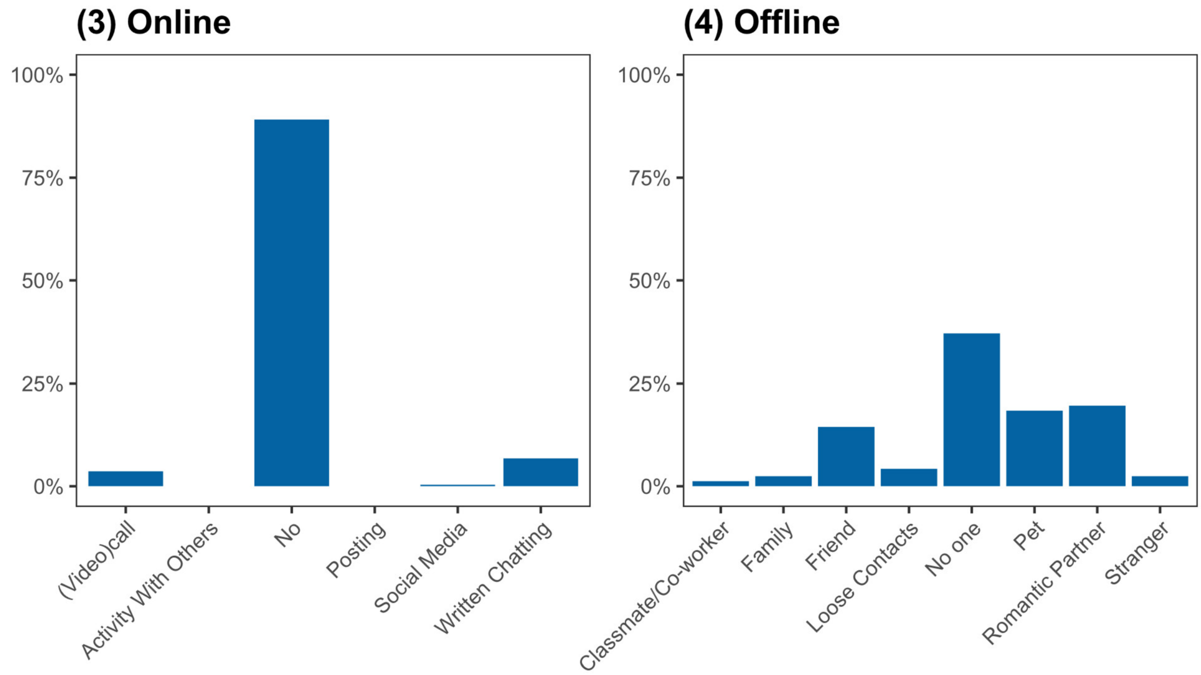
Example of "Activities and Locations" Figures



Note. Example figure for Activities (left panel) and locations (right panel). Items indicating the activity or location are shown on the x-axis, the y-axis indicates the relative frequency of how often participants indicated these.

Figure 5

Example of "Online- and Offline- Contacts" Figures



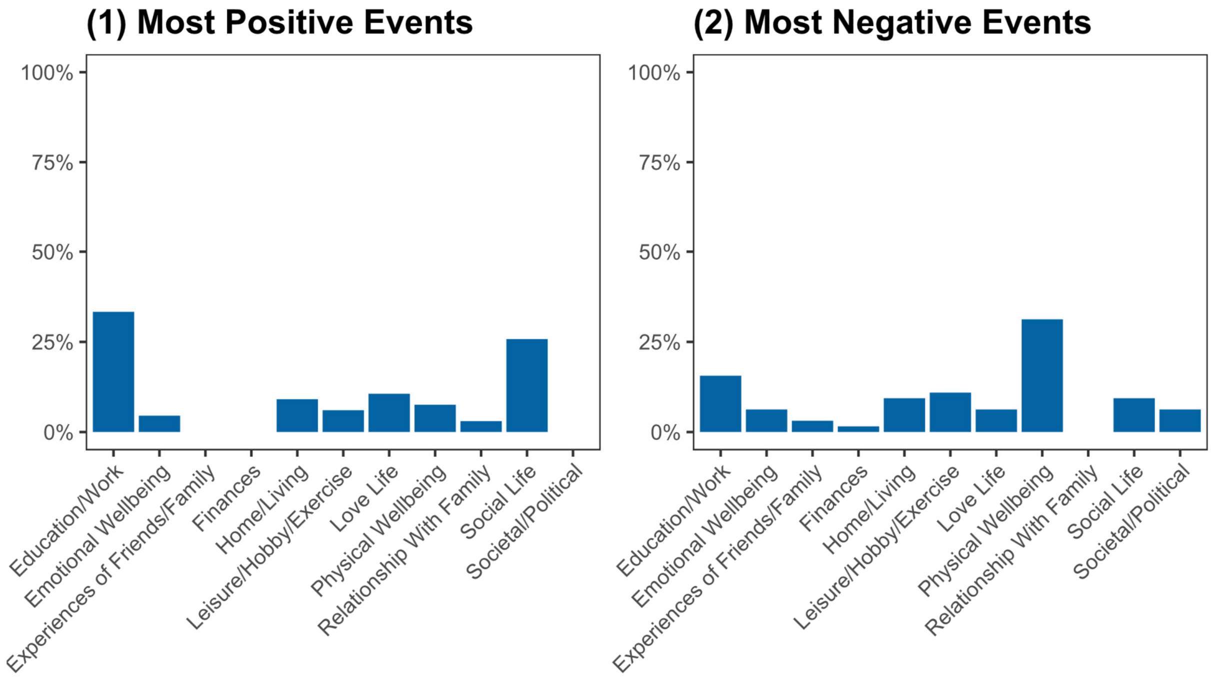
Report Section 4: Events

The fourth section is the first one that depends on data availability. Items asking for the most positive and most negative events per day were prompted in the evening surveys. Participants who never completed an evening survey did not receive this section.

The "Events" section shows relative frequencies of categories to which participants assigned their most positive and most negative event of the day. An example can be found in Figure 6.

Figure 6

Example of "Most Positive and Most Negative Events" Figures



Note. Example figure for most positive events (left panel) and most negative events (right panel). Items indicating the category to which an event belongs are shown on the x-axis, the y-axis indicates the relative frequency of how often participants indicated these.

Report Section 5: Development of Answers

The fifth section contains up to four time-series figures. These figures contain raw data points and a smoothed trendline if participants completed more than 25% of the surveys

The four different figures in this section are "Daily Mood and Feeling Tired" (Figure 7), "Morning and Evening Surveys" (Figure 8) and two figures about "Emotional and Physical Well-Being" (Figure 9 and Figure 10). All of them have Likert-scales as y-axis and timepoints as x-axis. Weekends in these graphs are greyed out to highlight potential differences between weekdays and weekends.

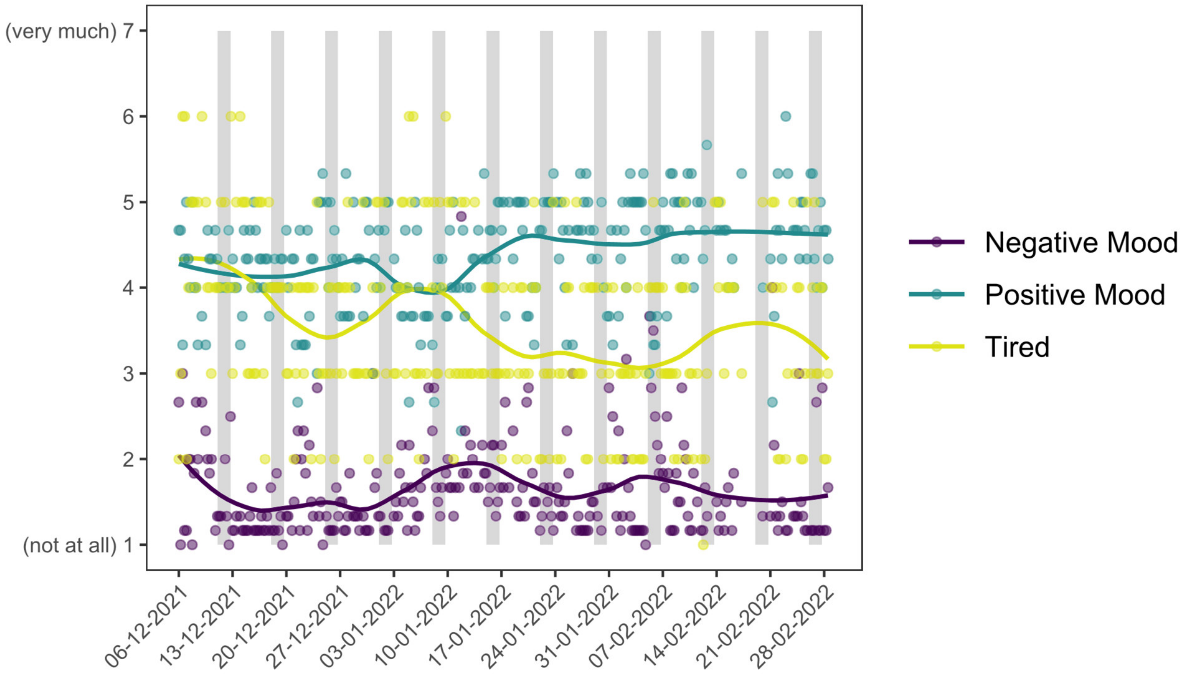
The “Daily Mood and Feeling Tired” figure is based on the four daily surveys and can have a maximum 340 timepoints. In this graph the data points for the items Sleep Quality, Satisfaction with Day and Outlook Tomorrow are shown. The smoothing parameter for this LOESS-curves was set to 0.35.

The “Morning and Evening Surveys” figure is based on the evening and morning surveys and can have a maximum of 85 timepoints. In this graph, the data points for the items Positive Mood, Negative Mood and Tired are shown. Again, the smoothing parameter for this figure was set to 0.35.

The “Emotional and Physical Well-Being” figures are based on the weekly Sunday surveys and can have a maximum of 12 timepoints. In figure 9, the data points for the items Emotional Wellbeing, Physical Wellbeing are shown. In figure 10, the data points for Life Satisfaction and Weekly stress are shown. The smoothing parameter for these two figures was set to 0.5, due to the lower number of observations.

Figure 7

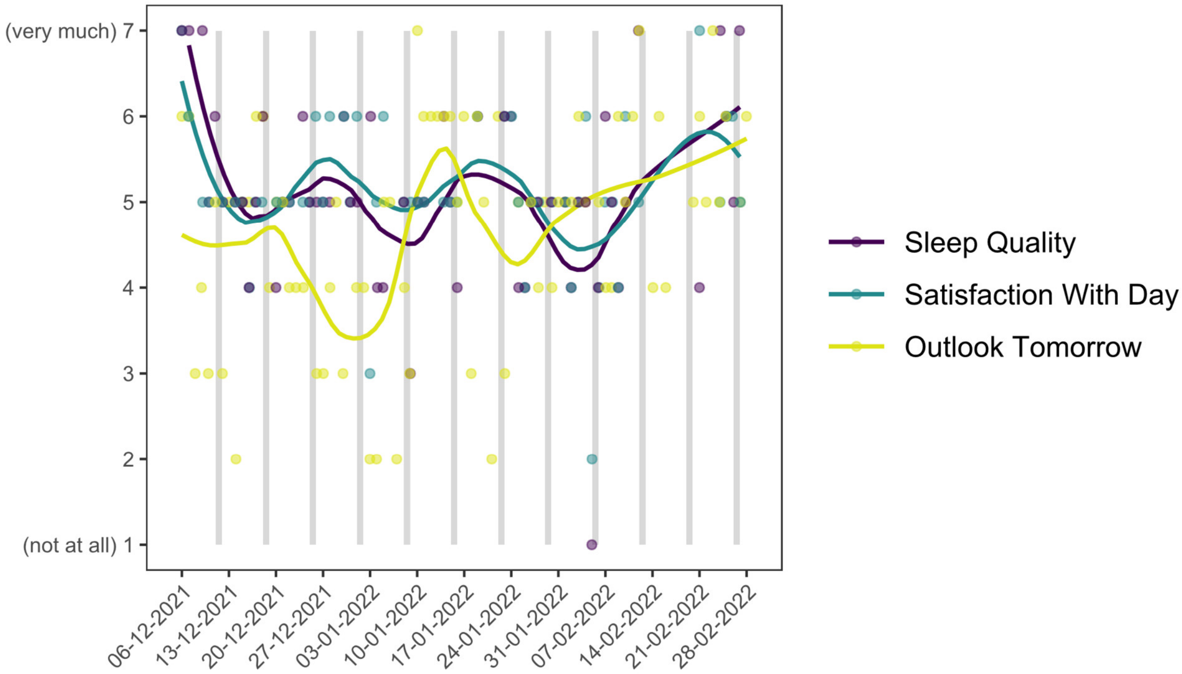
Example of "Daily Mood and Feeling Tired" Figure



Note. Example figure of a time series showing the development of mood and tiredness. Weekends are highlighted as grey bars.

Figure 8

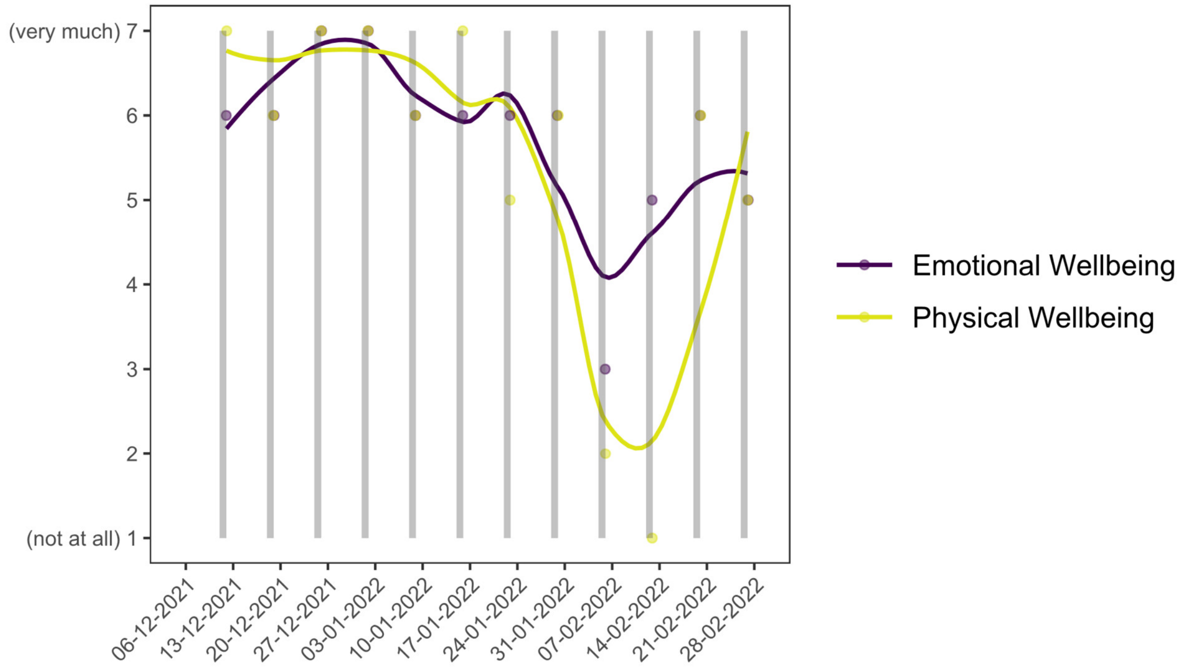
Example of "Morning and Evening" Figure.



Note. Here we can see considerable changes in reported sleep quality over time, which seems to be phasic. Furthermore, we can see the reported satisfaction with the day follows a similar pattern. Additionally, it is visible that the item outlook for tomorrow often follows an opposing trend.

Figure 9

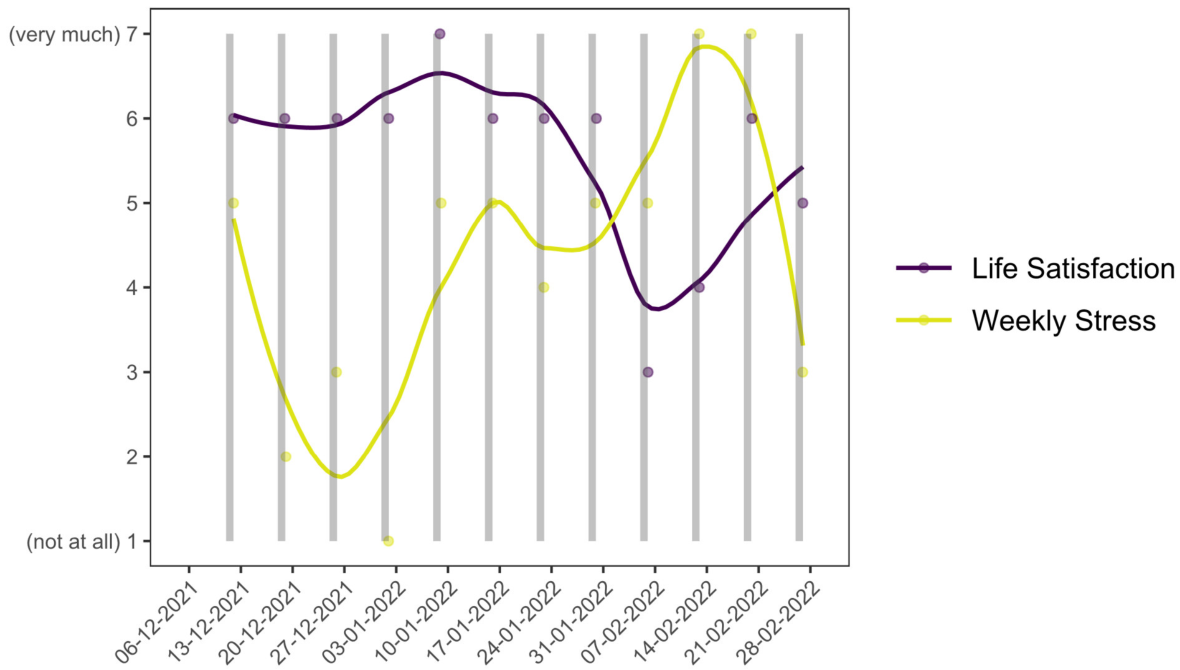
Example of "Emotional and Physical Well-Being" Figure.



Note. Here we can see that reported emotional wellbeing and physical wellbeing for the example participant follows the same trend with a higher amplitude for physical wellbeing.

Figure 10

Example of "Life Satisfaction and Weekly Stress" Figure



Note. In this figure it is visible that reported life satisfaction and weekly stress often follow opposing patterns. Roughly, the higher the stress the lower the reported life satisfaction.

Report Section 6: WARN-D Recap

The sixth section has two different sub-sections. The first subsection entails information about the week in which participants reported the most positive average mood. The second subsection entails information about the week in which the participant reports the most negative average mood. Both sections indicate the respective week, the average positive/negative mood in this week and the average positive/negative mood over the whole time. Furthermore, both sub-sections include figures about the context they were in (see Figure 4 and 5) and a figure about their mood development and tiredness (see Figure 7) for the respective week. An example text on the most positive week can read like "The week for which you reported the most positive mood was from 07-02-2022 to 13-02-2022. In this week you reported on average a positive mood of 4.8. In comparison, across the entire Stage 2, your positive mood was on average 4.4. As a reminder the scale was from 1 (not at all) to 7 (very much)."

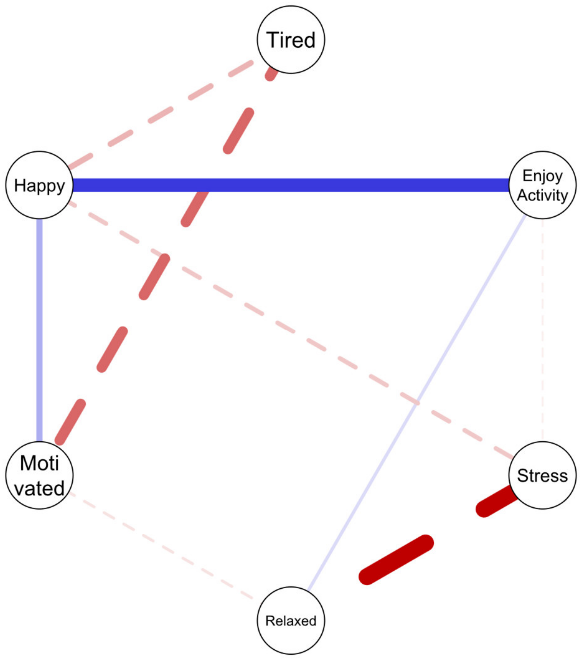
Report Section 7: Mood Network

The seventh section was only shown to participants if they completed more than 50% of the surveys. In this section a personalized network on mood variables ("Mood Network") was shown if the network was not empty. An empty network is the result of no nonzero partial correlations between variables. The mood network is a contemporaneous network indicating lag-1 controlled partial correlations of the six items tiredness, enjoyment of activity, stress, relaxation, motivation, and happiness. Autocorrelations of items were not shown to increase understandability.

In the networks positive partial correlations are indicated by blue edges. Negative partial correlations are indicated by red-dashed edges. The thickness of these edges indicates the strength of the partial correlation and was scaled to the largest partial correlation found in this network. If no relationship is found between nodes, this edge does not appear.

Figure 11

Example of "Mood Network".



Note. Here we see an example for someone who has the strongest positive correlation between happiness and the enjoyment of the current activity. The strongest negative correlation can be found between feeling relaxed and feeling stressed. Additionally, there is a weak positive correlation between enjoyment of activity and feeling relaxed. Other negative correlations are found between feeling tired and motivated, happy and stressed, happy and tired, and feeling motivated and relaxed.

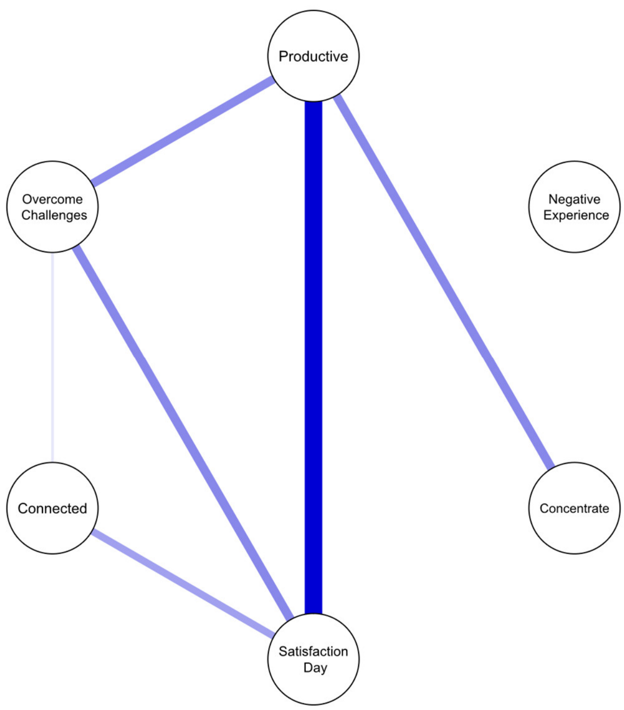
Report Section 8: Evening Network

The eight section entailed an "Evening Network" and is like the Mood Network in respect of computation, pre-processing, and criteria to whom it was shown. However, the Evening Network was not shown to four participants for whom the network analyses did not converge.

The Mood Network and Evening Network differ in the items shown to the participants, and the number of maximum observed surveys, since the Evening Network is based on items that were only asked once per day. The network includes items that cover participants' satisfaction with their day, overcoming of challenges, concentration, productivity, feeling of connectedness, and the rating of the most negative event of a day.

Figure 12

Example of "Evening Network".



Note. In this network we can only find positive edges between different nodes. These edges differ in the strength of the association.

Assessment of Reports Created With FRED

Sample Description

The personalized data reports were sent out to 398 participants. Every participant had the possibility to give feedback on the report they received, using the survey we sent out per email. 52 participants out of the 398 (13.1%) completed this survey. These participants completed on average 255.5 surveys (median = 272, min = 57, max = 349).

A summary of the descriptive statistics for the participants completing the baseline and participants completing the survey about FRED can be found in Appendix D. The subsample is composed very similarly to the full sample, with the difference that fewer English-speaking students are represented.

The reports were sent out with an error that happened during data cleaning. Due to this error, wrong relative frequencies were shown in the sections “Situations and Behaviors,” “Events,” and the “WARN-D Recap”. While it is difficult to quantify this objectively, the false information roughly affects 20% of the figures, and the respective explanations of these figures. The feedback survey was answered by 32 participants before a correction was sent out.

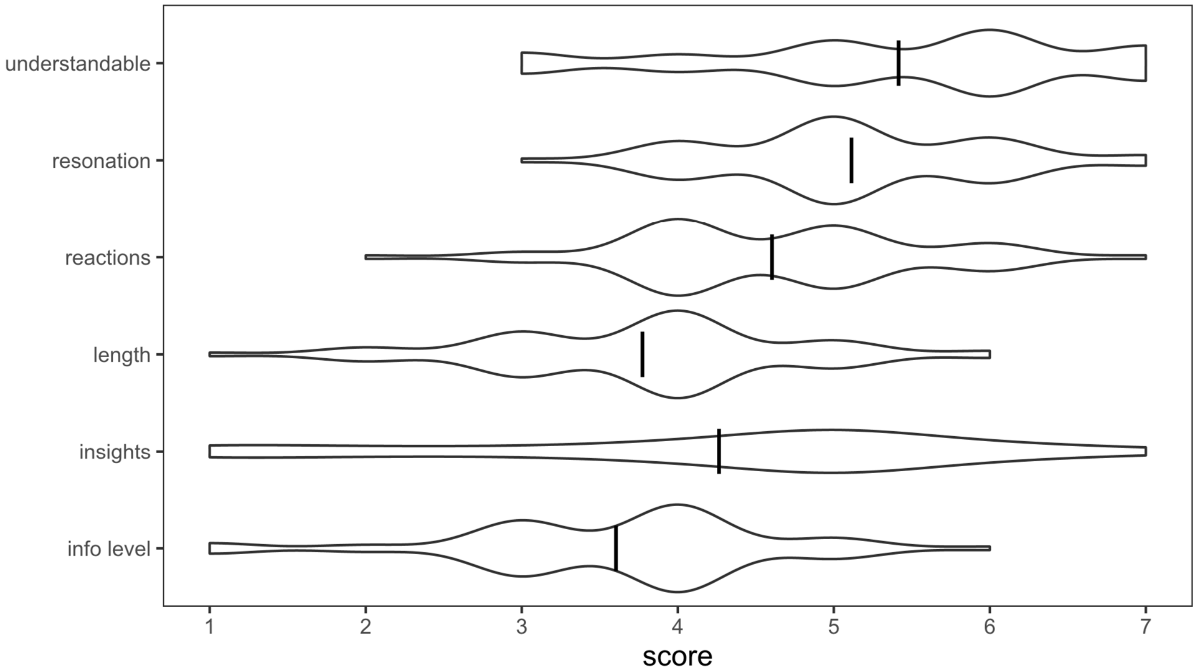
Data Description

All the following items were answered on a scale from 1 to 7. For the items ‘understandable’, ‘resonation’ and ‘reactions’ the anchor points were 1: Not at all and 7: Very much. The item insights was answered on a scale 1: Very negative and 7: Very positive. The items assessing information level and detail level were answered on a scale 1: Not enough, 4: Exactly right and 7: Too much.

Most participants found that the reports describe them well (*M* = 5.0, *SD* = 1.0) and indicated to understand the report (*M* = 5.3, *SD* = 1.3). On average, participants reported that the personalized data report is moderately insightful (*M* = 4.2, *SD* = 1.7). Reactions to the report were on average more positive than negative (*M* = 4.6, *SD* = 1). Furthermore, participants assessed the level of detail in the report as close to exactly right for a question ranging from too little detail to too much detail (M = 3.6, SD = 1). The same pattern is visible for the length of the report (M = 3.7, SD = 1). A more detailed distribution of these items can be found in figure 13.

Figure 13

Distribution of Likert-Scale Items

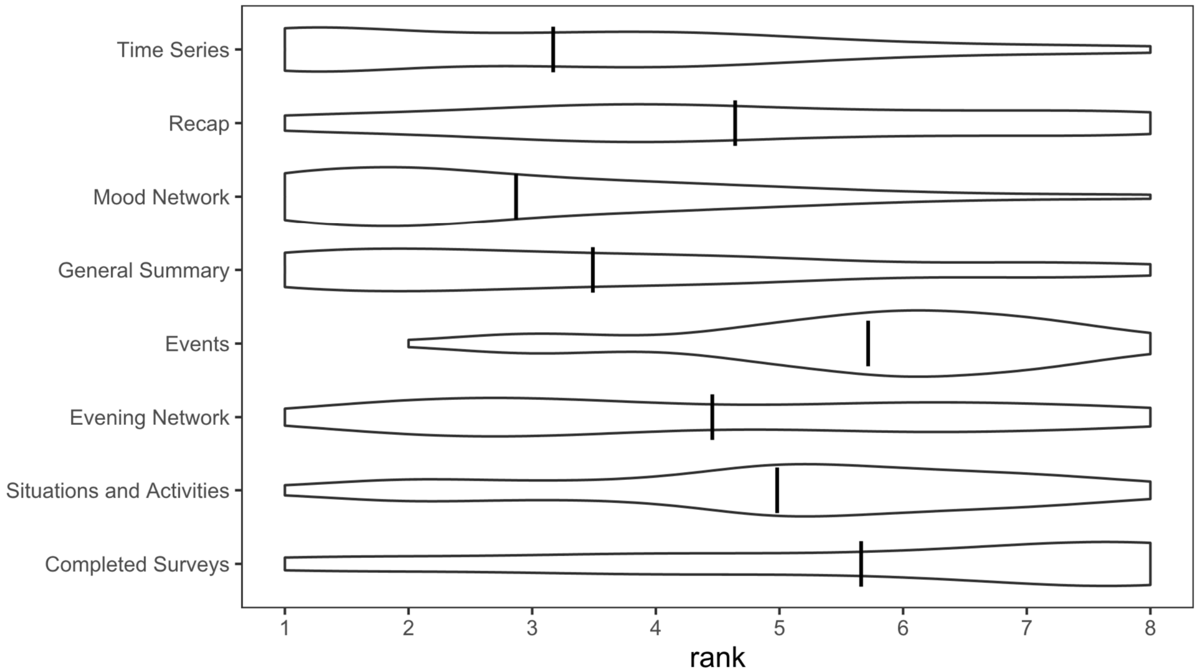


Note. This figure shows the distribution of the assessed items in the feedback survey. All these items were answered on 7-point Likert-Scale, however they differed in anchor points. The black vertical bar indicates the mean on the respective item.

Participants were asked to rank the different sections according to how interesting they find them, with 1 being the most interesting and 8 being the least interesting section to them. The sections sorted by interestingness, listed in decreasing order are: Mood Network (M = 2.8, SD = 1.7), Development of Answers/Time Series (M = 3.2, SD = 2), General Summary (M = 3.5, *SD* = 2.1), Evening Network (*M* = 4.4, *SD* = 2.2), Recap (*M* = 4.7, *SD* = 2.2), Situations and Activities (*M* = 4.9, *SD* = 1.9), Completed Surveys (*M* = 5.5, *SD* = 2.5), and Events (*M* = 5.7, *SD* = 1.7). An overview of the distribution of the rankings of the individual sections can be found in Figure 14.

Figure 14

Distribution of Ranking of Interest in Sections



Note. In this ranking 1 means the most interesting section and 8 the least interesting section of the personalized data reports. The black vertical bar indicates the mean ranking of the respective section. However, the median is also deductible, since the median is, where the distributions are the broadest.

Relationship Between Insightfulness and Completed Surveys

Additionally, the relationship between insightfulness and the number of completed surveys was assessed. We used a linear and quadratic model regressing number of completed surveys on insightfulness and a dummy-coded variable indicating if participants completed the survey only after the correction of reports.

Neither the linear model, $F\ (2, 49) = 1.75$, $p = 0.18$, $R^2 = 0.07$, nor the quadratic model, $F\ (3, 48) = 1.80$, $p = 0.16$, $R^2 = 0.10$, yielded a significant result (see Appendix C for detailed model estimates). This means that in this sample the insightfulness does neither linearly nor quadratically relate to the number of completed surveys.

Qualitative Data

Two different open text fields were available in the feedback survey. The first one inquired about reactions participants experienced when reading the report. 38 participants answered this item. A summary of a thematic analyses of these text entries can be found in Table 3.

Table 3

Summary of Open Responses to Participants' Reaction to the Report

Topic	Reactions	Frequency
Reported associations between variables	Positive, Interesting, not surprising, felt understood	7
Expected more	disappointment	2
Results were expected		4
High resonation	happy	2
New insights	Surprise, sad, curious, excited, interested	5
Missing smartwatch data	disappointment	2
No standard/reference		2
Not understanding report		2
Too little relationships	disappointment	2
Reflection moment	Surprised, positive, reassuring, Relieve, hopeful	6
Trajectory of mood	Sad*, interestedness	2

Note. The column "topic" summarizes topics participants reported to the item "Could you explain your answer, e.g., what reactions you had? (optional)". The column "Reactions" summarizes explicitly mentioned emotional reactions. *One person reported about the experience of seeing the mood deteriorate over time and felt it had caused them a low point.

The second open text field was about any additional feedback. This item was answered by 27 participants. A summary of common topics can be found in Table 4.

Table 4

Themes of Additional Comments

Theme	Comments	Frequency
Request for additional information on:	Substances, Relation between context and emotions, Numeric summaries/correlations, Open text fields, Reference scores	9
Visualizations	General summary overwhelming, Add legend to mood network, Raw data points in time series hard to distinguish	3
Feedback reports	Good explanations and visualizations, Thanks for effort	2
General feedback on EMA surveys	Phrasing of one item, Option to skip questions	2
Technical issues	Sending of smartwatches was not smooth, Additional option to complete surveys if missed	2
Missing smartwatch data	Link to other variables (emotions)	3
More raw data		2
WARN-D interpretation of results	Warnings for psychopathology, qualitative meanings of results	7

In summary, we learn from Tables 3 and 4 that participants' negative experiences caused by the reports are mainly because participants expected further information on additional variables and relationships between these. Furthermore, many participants asked for qualitative interpretation of results or warnings for psychopathology.

Discussion

We followed 428 students over 85 days during an EMA study. After completion of the EMA data collection, we wanted to provide participants with a personalized data report as a reward. There were three major challenges; a) technical: develop a tool that allows the generation of personalized data reports after completing an EMA study b) methodological: create insightful and understandable reports, c) ethical: minimizing potential harm. Furthermore, we assessed participants perception of the resulting reports, to see whether we could overcome the prior mentioned challenges. We successfully developed FRED as a tool to generate large numbers of reports on EMA data (technical challenge)

The feedback survey showed that most participants found the personalized data reports well understandable and moderately insightful (methodological challenge), with few participants reporting very little insightfulness and understandability. Furthermore, participants reported mostly neutral to positive reactions to the report (ethical challenge). The open text fields of participants who reported negative reactions, were mostly themed around disappointment due to too little information. Therefore, there is no evidence that these reports inflicted harm on participants. Overall, participants perceived the length and information level of the report as close to optimal.

In the feedback survey we additionally assessed how interesting participants find the different sections compared to the other sections in the reports. We can see that participants are interested in more complex analyses, but not exclusively. The top three categories are time-series visualizations, the Mood Network and a general summary of different variables. The additional Evening Network, as another example of a complex analyses, was ranked forth. Interestingly, participants were least interested in the domains of their most positive and most negative experiences per day. On average, participants ranked this category even lower than the information on how many surveys they have completed. The lower ranking of the sections reporting on relative frequencies on categorical data compared to more complex analyses could also be due to the lack of reference scores from other participants, as suggested by open answers from participants. Additionally, these sections were affected by the erroneous report which could have affected the insightfulness and interestingness of these sections.

Furthermore, we could not find a relationship between compliance (completed surveys) and the insights participants gained. The reason we expected such a relationship is because the more data we have per participant, the more analyses we have provided to participants, and the more reliable the analyses are. The compliance rate of participants across the EMA phase who completed the evaluation survey ranged from low to very high. Not finding this relationship can have many reasons such as a) no relationship between constructs, b) an underpowered regression, c) the operationalization of insightfulness, d) unmeasured confounders. The issue is that the result can be interpreted negative as well as positive. A negative interpretation could be the additional analyses we provided (trendlines for time-series, network-analyses) do not add any additional information. A positive interpretation could be even when compliance is low, providing participants with a report has positive outcomes.

A particular strength of the current work is that it provides a framework for a different setting than other feedback tools available right now. One difference is that FRED is designed to be used for large scale projects. Using FRED as a framework makes it easy to develop personalized data reports. Since it is written in R it is possible to easily adjust it to the individual needs of different studies. Furthermore, it is possible to use it in a wide variety of different contexts, such as observational as well as interventional studies

Limitations of the current work fall in two different categories: limitations regarding FRED as a tool itself and limitations regarding the feedback survey. The main limitation of FRED is that coding skills are required to be able to use it, although we are working on a shiny app implementation that will make report generation easier. Another limitation is, as visible in the qualitative data, that some participants expect information in the reports, that is subject to iatrogenic effects, such as warnings about a possible depressive episode (cf. misdiagnosis Mc Glanaghy et al., 2022).

There are bigger limitations affecting the feedback survey. The first limitation is that some participants received reports that included an error. This error could have affected participants resonation with the report and insights from the report. Furthermore, it could have affected the ranking of the different sections. A second limitation is the low response rate to the feedback survey. Only 13.6% of the invited participants completed this survey. Even though this sample seems representative for the full sample there might be selection bias. It is likely that only participants who were invested in the study completed the survey. This could have resulted in a participation bias (Elston, 2021) and a sample that is more extreme in the sense of more satisfied and unsatisfied participants filling in this survey.

Furthermore, to have an actual estimation of level insights participants gained through the reports other research designs need to be used. It is possible that participants gained insights just by participating in the EMA study, but not due to the personalized data reports. Thus, a pre-test could be conducted and/or a control-group could be used. To investigate this in more detail future studies should use these research designs to investigate the effects of feedback as suggested by Leertouwer et al. (2021). Additionally, it would be good to use control variables such as response style, agreeableness, and an item about participants' expectation of the report.

Future Directions

Future development of FRED can be categorized in two main domains - technical and content. Content development entails the implementation of different analysis as well as making different types of data, such as qualitative data, includable. An example of how feedback on qualitative data could be provided is wordclouds (Fellows, 2018). Wordclouds are essentially a way of indicating the frequency of particular words used in open text. There are already lexica that exclude common words such as “and”, “I”, etc. since feedback on these words is little insightful. A possibility could be to create a new lexicon that enables the exclusion of potential harmful topics such as death, suicide, and abuse. However, this might be challenging if reports are generated in multiple different languages. Another issue is that words can have a private meaning. For some people the word car might be a neutral word, for others it could be a reminder of a traumatic car-accident.

Future reports could also include visualizations and analyses of collected smartwatch data. Regarding smartwatch data it is important to consider how reliable the data is. Other analyses that could be included are temporal networks (Epskamp et al., 2018). We decided to not include these networks to reduce the possible misunderstanding of causality vs correlation. However, this could be a particular insightful analysis for participants, since it gives insights into one’s own working mechanism. However, it is important to consider the time frame between surveys and the nature of the relationship between constructs (linear, quadratic, exponential,…) to determine whether it is even possible to capture causal relationships between the different constructs (Robinaugh et al., 2021).

In the current version of FRED, we gave a recap of the on average most positive and least positive week participants had during the EMA phase. We operationalized the most positive week as the week with the highest average positive affect, and the least negative week as the week with the highest negative affect. We decided to use the mean and not the median, since the mean is a more known measure. However, the median might give a better representation. In the future, an operationalization that takes positive and negative affect simultaneously into account could be useful to increase the resonation as well as insights for participants.

For visualizations it is also possible to include animated time-series graphs, see e.g. Bringmann et al. (2021) and Klipstein et al. (2022). While animations stress dynamical changes over time (Heer & Robertson, 2007), they do not include additional information. Hence, the use of animations can make the information less accessible depending on the cognitive load of given visualizations (Kriglstein et al., 2012). We decided against animated graphs to keep the cognitive load low in the reports. Furthermore, automatic interpretations of some analysis could be implemented. An example for a simple automated interpretation could be the strongest/weakest relationship found in a network analysis.

Technical improvements include the implementation of parallel processing to increase rendering speed. This is in particular relevant for even larger datasets. Another consideration is to change the delivery medium from an HTML file to a shiny app (Chang et al., 2021). A shiny app has the advantage that reports can be more interactive, for example participants could choose on which variables they want to see a time-series or a network analysis, whether the mean or median should be chosen to calculate the most or least positive week. Implementing this would increase possible insights for participants. However, the amount of accessible information could also be too large to process. Using a shiny app comes with the same ethical and methodological considerations, and additionally data security gets even more relevant.

To summarize, we successfully developed a tool that allows the large-scale generation of personalized data reports for participants in an EMA study. Participants perceived the personalized data reports reasonably well. While there are still many things that could improve participants perception of the reports, FRED provides a promising framework.

References

Bälter, O., Fondell, E., & Bälter, K. (2012). Feedback in web-based questionnaires as incentive to increase compliance in studies on lifestyle factors. *Public Health Nutrition*, *15*(6), 982–988. https://doi.org/10.1017/S1368980011003041

Bertz, J. W., Epstein, D. H., & Preston, K. L. (2018). Combining ecological momentary assessment with objective, ambulatory measures of behavior and physiology in substance-use research. *Addictive Behaviors*, *83*, 5–17. https://doi.org/10.1016/j.addbeh.2017.11.027

Bos, F., von Klipstein, L., Emerencia, A. C., Veermans, E., Verhage, T., Snippe, E., Hadders-Prins, G., Wichers, M., & Riese, H. (2022). *PErsonalized Treatment by Real-time assessment (PETRA): A web-application for personalized diaries in psychiatric care* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/jh49z

Bringmann, L. F., van der Veen, D. C., Wichers, M., Riese, H., & Stulp, G. (2021). ESMvis: A tool for visualizing individual Experience Sampling Method (ESM) data. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *30*(11), 3179–3188. https://doi.org/10.1007/s11136-020-02701-4

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *shiny: Web Application Framework for R*. https://CRAN.R-project.org/package=shiny

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Comtois, D. (2021). *summarytools: Tools to Quickly and Neatly Summarize Data*. https://CRAN.R-project.org/package=summarytools

DeCou, C. R., & Schumann, M. E. (2018). On the Iatrogenic Risk of Assessing Suicidality: A Meta-Analysis. *Suicide and Life-Threatening Behavior*, *48*(5), 531–543. https://doi.org/10.1111/sltb.12368

Di Cook, T. N., McBain, M., & Fay, C. (2021). *naniar: Data Structures, Summaries, and Visualisations for Missing Data*. https://CRAN.R-project.org/package=naniar

Doherty, K., Balaskas, A., & Doherty, G. (2020). The Design of Ecological Momentary Assessment Technologies. *Interacting with Computers*, *32*(3), 257–278. https://doi.org/10.1093/iwcomp/iwaa019

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*(4), 463–475. https://doi.org/10.1037/a0017075

Elston, D. M. (2021). Participation bias, self-selection bias, and response bias. *Journal of the American Academy of Dermatology*, S0190962221011294. https://doi.org/10.1016/j.jaad.2021.06.025

Epskamp, S. (2021). *GraphicalVAR: Graphical VAR for Experience Sampling Data*. https://CRAN.R-project.org/package=graphicalVAR

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological*

Science : A Journal of the Association for Psychological Science, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Ethica (Version 550) (2021a). [IOS]. Ethica Data Services Inc. https://apps.apple.com/nl/app/ethica/id1137173052

Ethica (Version 550). (2021b). [Android]. Ethica Data Services Inc. https://play.google.com/store/apps/details?id=com.ethica.logger

Fanning, J., Silfer, J. L., Liu, H., Gauvin, L., Heilman, K. J., Porges, S. W., & Rejeski, W. J. (2020). Relationships between respiratory sinus arrhythmia and stress in college students. *Journal of Behavioral Medicine*, *43*(2), 308–317. https://doi.org/10.1007/s10865-019-00103-7

Fellows, I. (2018). *wordcloud: Word Clouds*. https://CRAN.R-project.org/package=wordcloud

Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental Health and Social Contact During the COVID-19 Pandemic: An Ecological Momentary Assessment Study. *Clinical Psychological Science*, *10*(2), 340–354. https://doi.org/10.1177/21677026211017839

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Antônio Pedro, Sciaini, Marco, Scherer, & Cédric. (2021). *viridis—Colorblind-Friendly Color Maps for R*. https://doi.org/10.5281/zenodo.4679424

Grolemund, G., Wickham, H., & Epskamp, S. (2021). Dates and Times Made Easy with lubridate: GraphicalVAR: Graphical VAR for Experience Sampling Data. *Journal of Statistical Software*, *40*(3), 1–25.

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling Affect Dynamics: State of the Art and Future Challenges. *Emotion Review*, *7*(4), 316–322. https://doi.org/10.1177/1754073915590619

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Heer, J., & Robertson, G. (2007). Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1240–1247. https://doi.org/10.1109/TVCG.2007.70539

Hsieh, G., Li, I., Dey, A., Forlizzi, J., & Hudson, S. E. (2008). Using visualizations to increase compliance in experience sampling. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 164–167.

Jacoby, W. G. (2000). Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, *19*(4), 577–613.

Kriglstein, S., Pohl, M., & Stachl, C. (2012). Animation for Time-oriented Data: An Overview of Empirical Research. *2012 16th International Conference on Information Visualisation*, 30–35. https://doi.org/10.1109/IV.2012.16

Leertouwer, Ij., Cramer, A. O. J., Vermunt, J. K., & Schuurman, N. K. (2021). A Review of Explicit and Implicit Assumptions When Providing Personalized Feedback Based on Self-Report EMA Data. *Frontiers in Psychology*, *12*, 764526. https://doi.org/10.3389/fpsyg.2021.764526

Long, J. A. (2020). *jtools: Analysis and Presentation of Social Scientific Data*. https://cran.r-project.org/package=jtools

Mansueto, A. C., Wiers, R. W., van Weert, J. C. M., Schouten, B. C., & Epskamp, S. (2022). Investigating the feasibility of idiographic network models. *Psychological Methods*. https://doi.org/10.1037/met0000466

Mc Glanaghy, E., Jackson, J., Morris, P., Prentice, W., Dougall, N., & Hutton, P. (2022). Discerning the adverse effects of psychological therapy: Consensus between experts by experience and therapists. *Clinical Psychology & Psychotherapy*, *29*(2), 579–589. https://doi.org/10.1002/cpp.2648

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*, *41*, 1–8. https://doi.org/10.1016/j.copsyc.2021.01.004

Moritz, S., & Bartz-Beielstein, T. (2017). ImputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207–218. https://doi.org/10.32614/RJ-2017-009

Müller, K. (2020). *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here

Qualtrics. (2022). Qualtrics. https://www.qualtrics.com

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. 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

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

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

van Helvoort, D., Otgaar, H., & Merckelbach, H. (2020). Worsening of Self-Reported Symptoms Through Suggestive Feedback. *Clinical Psychological Science*, *8*(2), 359–365. https://doi.org/10.1177/2167702619869184

von Klipstein, L., Servaas, M., Schoevers, R. A., van der Veen, D. C., & Riese, H. (2022). *Integrating personalized experience sampling in psychotherapy: A case study* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/2srxq

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."* https://CRAN.R-project.org/package=cowplot

Appendices

Table of Contents

Appendix A 49

Used R Packages 49

Parameters for Network Analyses 49

Survey Schedule 50

Appendix B 51

Items Included in Reports 51

Appendix C 54

Items in Feedback Survey 54

Appendix D 55

Sample Description 55

Model Estimates of Relationship Between Insightfulness and Completed Surveys .. 55

Appendix A

Used R Packages

Multiple R packages are used for FRED to simplify data wrangling, processing of timestamps, analyses, visualizations, and file management. For data wrangling the package dplyr (Wickham et al., 2021) was used. For the processing of timestamps we used the package lubridate (Grolemund et al., 2021). Relative frequencies were calculated with the package summarytools (Comtois, 2021). All figures but the networks in the reports are generated using ggplot2 (Wickham, 2016) and jtools (Long, 2020). We used a color selection for figures that is colorblind friendly using the package viridis (Garnier et al., 2021). For figure alignment the package cowplot (Wilke, 2020) was used. We estimated networks for participants having completed more than 50% of the surveys. To do so we imputed missing data for the selected variables using a Kalman filter (Moritz & Bartz-Beielstein, 2017). Missingness of data was assessed using the naniar package (Di Cook et al., 2021) and the Kalman filter imputation was accomplished using imputeTS (Moritz & Bartz-Beielstein, 2017). For the network estimation the package graphicalVar (Epskamp, 2021) was used. The estimated networks were displayed using qgraph (Epskamp et al., 2012). To simplify file management and to be independent of absolute working directories, the here package (Müller, 2020) was used.

Parameters for Network Analyses

To decrease overfitting of the network- models, we used the LASSO parameter λ. LASSO puts a penalty on the inclusion of additional parameters, this results in a model in which edges are only included if the fit of this model improves, while balancing parsimony of the model. Shrinkage in LASSO refers to the fact that regression weights are shrunk, and selection refers to the fact that small partial correlations are set to 0. This is achieved by using the tuning parameter λ (Chen &

Chen, 2008). Using a very small λ results in higher variance and less bias of this particular model. In network analysis the tuning parameter λ controls the sparseness of the resulting network since higher values of λ lead to less removal of edges. The selection of λ was achieved using the Extended Bayesian Information Criterion (EBIC), which uses the parameter γ to control the preferred sparsity. A typical range for γ is between 0 and 0.5, with higher values of γ leading to sparser models. We chose a γ of 0, which makes the EBIC like the regular Bayesian Information Criterion, to estimate fuller networks per participant.

Survey Schedule

Table 1

Overview of Surveys

Survey Name	Time Schedule	Availability	Prompt Pattern	Number of Items
Morning Survey	9:49 - 10:19	20 min	Daily	21 (19 +2 cond.)
Noon Survey	13:34 - 14:04	20 min	Daily	18 (16 +2 cond.)
Afternoon Survey	17:19 - 17:49	20 min	Daily	18 (16 +2 cond.)
Evening Survey	21:04 - 21:34	20 min	Daily	36 (33 + 2 cond. + 1 opt.)
Sunday Survey	11:45 - 12:15	10 h	Weekly	46 (44 + 1 cond. + 1 opt.)

Note. Surveys were sent out at a time in the indicated time frames following a normally distributed jitter; cond.= conditional item (depending on the answer to another item), opt.= optional item.

Appendix B

Items Included in Reports

Table 1

Selected Items for Personalized Data Reports

Variable Name	Item	Scale
Concentrate	Today, I was able to concentrate and focus well.	*
Connected	Today, I felt connected to other people.	*
Emotional Wellbeing	This week, my overall mental health and emotional wellbeing were:	**
Enjoyment of Activity	I am enjoying what I am doing right now.	*
Enjoyment Offline Contact	I am enjoying my company.	*
Enjoyment Online Activity	I am enjoying this online activity.	*
Life Satisfaction	All things considered, I am satisfied with my life as a whole.	*
Negative Mood	I feel sad right now. I feel stressed right now. I feel overwhelmed right now. I feel nervous/anxious right now. I am experiencing negative thoughts right now. I feel annoyed/irritated right now.	Composite Score of individual items on * scale
Negative Experience	This event/experience was … Follow up on: 'My most negative event/experience today was:'	*
Outlook Tomorrow	I am looking forward to tomorrow.	*
Overcome Challenges	I was able to handle today's challenges well.	*
Physical Wellbeing	My overall physical health this week was:	**
Positive Mood	I feel relaxed right now. I feel motivated right now. I feel happy/cheerful right now.	Composite Score of individual items on * scale
Productive	Today, I felt productive/useful.	*
Rested	When I woke up, I felt well rested.	*
Satisfaction With Day	Overall, I'm content with how my day went.	*
Sleep Quality	Last night, I slept well.	*
Tired	I feel tired right now.	*
Weekly Stress	This week was stressful for me.	*

Variable Name	Item	Scale
Location	Right now, I am at:	home friend's place my family's place work/school transport/public transport other indoors outdoors city outdoors nature
Activity	Right now, my activity is (choose all that apply):	social (offline or online) physical (e.g. cycling, gym) active leisure (hobby, board game) passive leisure (e.g. watching TV, scrolling Instagram) studying/working chores (e.g. cleaning house) on my way to somewhere Eating Other nothing
Online Contact	social online contact right now, I am interacting with others online (choose all that apply):	No Yes: reading/ scrolling/liking Yes: posting Yes: written chatting Yes: call/videocall Yes: doing something with others (e.g.online gaming)
Offline Contact	Social offline contacts Right now, I am with (choose all that apply):	friend(s) acquaintance(s)/loose contact(s) family romantic partner classmates/co-workers strangers a pet no one

Variable Name	Item	Scale
Positive Events	This event/experience belongs to the category (choose all that apply):	social love life personal life education/work home leisure/hobby/pleasure relationship with family Experiences of friends/family Societal/political other
Negative Events	This event/experience belongs to the category (choose all that apply):	social love life personal life education/work home leisure/hobby/pleasure relationship with family Experiences of friends/family Societal/political other

Note. Table 2 contains all items on which feedback was given to the participants. An asterisk (*) indicates that this item was answered on a 7-point Likert scale from '1: Not at all' to '7: Very much'. Two asterisk (**) indicate a scale from '-3: very negative' to '3: very positive".

Appendix C

Items in Feedback Survey

Table 1

Items in Feedback Survey

Item	Scale
Overall, the report describes me well.	1: Not at all, 7: Very Much
I learned something new based on the report.	1: Not at all, 7: Very Much
The report is understandable.	1: Not at all, 7: Very Much
The amount of information was…	1: Not enough, 4: Exactly right, 7: Too much
The length of the report was…	1: Not enough, 4: Exactly right, 7: Too much
The reactions I had when reading the report were…	1: Very negative, 7: very positive
Could you explain your answer, e.g. what reactions you had? (optional)	Open text
How interesting did you find the different section in your report?	Ranking with most interesting section at top and least interesting at bottom; Only sections the participant received were shown
Do you have any additional feedback for us? Any feedback is welcome, e.g. what you liked, didn't like, or if there were things you would have wanted us to report.	Open text

Appendix D

Sample Description

Table 1

Sample Description

Variable	Participants Completing Feedback Survey	All Participants
Gender	F = 47 (90.4%) M = 3 (5.8%) Non-binary = 2 (3.8%)	F = 360 (80.4%) M = 71 (15.8%) Non-binary = 16 (3.6%) Not shared = 1 (0.2%)
Age	22.9 (SD = 4.3, range = 18-42)	22.7 (SD = 4, range = 18-53)
Language	EN = 17 (32.1%) NL = 36 (67.9%)	EN = 220 (48.6%) NL = 233 (51.4%)
Education	MBO = 2 (3.8%) HBO = 4 (7.7%) University = 46 (88.5%)	MBO = 14 (3.1%) HBO = 47 (10.5%) University = 387 (86.4%)

Model Estimates of Relationship Between Insightfulness and Completed Surveys

Table 2

Results for Relationship Between Insightfulness and Completed Surveys

Parameter	Linear Model Estimate (SD)	p-Value	Quadratic Model Estimate (SD)	p-Value
Intercept	3.34 (0.91)	<.001	0.12 (0.18)	.53
Completed Surveys	0.004 (0.003	.20	0.03 (0.02)	.12
Square Completed Surveys			< -0.001 (< -0.001)	.18
Report Version	-0.7 (-0.54)	.20	- 0.7 (-0.54)	.22

Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen

Generating Feedback Reports for Ecological Momentary Assessment Data

Aljoscha Rimpler

Research Master Thesis *Clinical and Health Psychology*
Date: 05.08.2022
Supervisor: Eiko I. Fried
Second reader: Niki Antypa