

Original Paper

# Using Semiautomated WhatsApp Messages for Daily Stress Measurements: Integrated Usability and Feasibility Study

Janika Thielecke<sup>1,2</sup>, PhD; Maartje Bakuys Roozeboom<sup>1</sup>, PhD; Irene Niks<sup>1</sup>, PhD; Elsbeth de Korte<sup>1</sup>, PhD; Sadegh Shahmohammadi<sup>1</sup>, PhD

<sup>1</sup>Work and Health, The Netherlands Organization for Applied Scientific Research (TNO), Leiden, The Netherlands

<sup>2</sup>Department Clinical-, Neuro-, and Developmental Psychology, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

**Corresponding Author:**

Janika Thielecke, PhD

Work and Health, The Netherlands Organization for Applied Scientific Research (TNO)

Sylviusweg 71

Leiden 7415 NC

The Netherlands

Email: [janika.thielecke@tno.nl](mailto:janika.thielecke@tno.nl)

## Abstract

**Background:** Stress is a key determinant of health outcomes and may influence work performance. Questionnaire-based assessments of stress are typically broad and retrospective. Daily stress measurements via smartphones offer more granular, real-time data but have adherence issues. Using an already established communication medium (WhatsApp) and a more conversational style assessment might improve adherence and help collect more detailed insights into (work) stress, underlying stressors, and countering energy sources.

**Objective:** This study focuses on the usability and feasibility of semiautomated voice- and text-messages (with and without emojis) via WhatsApp as a method to collect daily data on experienced work stress, stressors, and energy sources.

**Methods:** A sample of 210 workers was recruited via social media and participated in a 10-workday diary study using semiautomated WhatsApp messages to rate daily stress, stressors, and energy sources. Questions (with and without emojis) were presented by a chatbot as text messages with clickable buttons (multiple-choice questions; MC) or with instructions to answer with either a voice or a text message. The study used an experimental design with 4 groups: (1) week 1 voice, week 2 text/MC with emojis; (2) week 1 voice, week 2 text/MC without emojis; (3) week 1 text/MC, week 2 voice with emojis; (4) week 1 text/MC, week 2 voice without emojis. Pre- and poststudy web-based questionnaires assessed demographics, familiarity with voice messages, and usability, including participants' preference for research studies. Open answers were coded using artificial intelligence, and the number of stressors or energy sources was compared across the 3 collection methods (MC, voice, and text messages) to determine if the amount and quality of information collected differ per method within participants.

**Results:** A total of 158 workers completed at least 80% of scheduled conversations. The sample was predominantly women (170/210, 81%), highly educated (173/210, 82%), and a slight majority worked part-time (109/210, 52%). Mean adherence to the daily schedule was very high (mean of 95%). The postquestionnaire revealed a strong preference for MC and text over voice messages, mostly due to ease and convenience in a variety of situations. The number of stressors per week was approximately 3 times higher in the MC-condition than in the voice condition, even though average stress levels per week did not differ significantly within participants. The number of energy sources was comparable between open answers in the voice and text conditions, but voice messages consisted of more words.

**Conclusions:** Collecting (work) stress data via semiautomatic WhatsApp messages is a feasible method with low effort for participants. Usability ratings indicated a strong preference among participants for MC and text messages over voice messages. Future research should explore usability in more diverse samples and in direct comparison to traditional assessment methods.

*JMIR Form Res* 2026;10:e84032; doi: [10.2196/84032](https://doi.org/10.2196/84032)

**Keywords:** work stress; whatsapp; chatbot; digital stress assessment; monitoring; artificial intelligence; AI; journalling; diary study; feasibility; usability

## Introduction

Stress is a fundamental aspect of human life and plays an important role across various domains, including health and occupational settings. Stress is not only a key determinant of mental and physical health outcomes [1-4] but also influences work performance [5], productivity [6,7], cognitive processes [8,9], and overall well-being of employees [10].

Especially in the workplace, understanding stress and contributing factors such as long-term stressors, daily hassles, and sources of energy is critical for developing and applying effective strategies that support the employee's well-being. To be able to act on work-related stress in a timely manner, an important precondition is that work-related stress and contributing factors can be measured close to real-time and with enough detail to inform actions [11,12]. In the next sections, we first briefly discuss the advantages and disadvantages of different stress measurement methods in this regard. We then suggest a novel way to assess (work-related) stress and influencing factors with a conversational style assessment. We then continue by explaining how this study tests the usability and feasibility of this new approach.

Measuring stress in general, including in the workplace, is a challenge, as (work-related) stress is a highly subjective experience, depending on an individual's perception and appraisal of (potential) stressors [13]. While physiological, presumably more objective stress measures in the workplace are on the rise [14,15], the most common way to assess (work-related) stress today remains through standardized self-report questionnaires consisting of multiple items [13, 16].

Commonly used questionnaires to assess (work-related) stress rely on one-time, retrospective questions covering a period of several days or weeks [13]. However, these approaches are less suitable for uncovering daily stress and hassles [13]. That is, retrospective recall is considered prone to biases, especially over longer periods of time, as individuals may misremember or reinterpret their experiences based on their current emotional state [17]. At the same time, one-time questionnaires are likely to capture a summary evaluation of stressful moments or could be impacted by a single, extreme event but fail to capture the context in which stress was experienced, making it hard to act on.

Alternatives to retrospective one-time questionnaires are diary studies and ecological momentary assessments (EMAs), which are both used to capture individuals' experiences in or near real-time with short questionnaires often consisting of single items [18]. This is tackling the problem of recall bias. With the spread of technology, these methods have become more widely applicable with specialized software and applications to facilitate web-based data collection. Traditional diary studies typically ask participants to record their thoughts, feelings, or behaviors at the end of the day or at scheduled intervals, either as open questions or short

questionnaires. Similarly, EMA is a more structured and intensive version of diary studies, often involving multiple prompts for short questionnaires throughout the day to capture momentary experiences. Repeated or daily measures in these studies explicitly aim to capture the different situations in which stress might be experienced. This gives much more granular and actionable insights compared with a once-a-week or monthly measure.

Nevertheless, both traditional questionnaires and EMA face some shared challenges. Accessing context factors is taxing, as closed multiple-choice (MC) questions, while quick and convenient, might not capture a situation completely. Open questions, on the other hand, are time-intensive for both participants and analysts, especially if asked multiple times per day. Additionally, both types require time and motivation from participants to answer the questions. While traditional questionnaires require time and motivation only once or occasionally, adherence to daily questionnaires likely requires greater motivation from participants. Adherence to EMA is one of its biggest challenges, as research has shown that many participants quickly uninstall (research-related) apps, reducing long-term engagement and adherence [19,20]. On the other hand, EMA can provide more direct insights and moments of reflection for the participants [21].

To address the above-mentioned challenges (ie, collecting real-time contextual information and engaging participants in daily measurements) while keeping the advantage of real-time EMA, we suggest focusing on a more conversational style stress measurement using familiar and established conversation platforms like WhatsApp.

WhatsApp is currently the most popular messaging application globally, with 3 billion active users [22] and in the Netherlands, with 13.3 million active users [23]. Given its integration into daily life, WhatsApp makes it easier to take part in research studies and answer questions. It therefore has an advantage compared with the use of special research applications. In a conversation-like assessment via WhatsApp, questions are asked by a chatbot and can be either written chats, voice chats, clickable buttons, or a mixture of those. Providing more detailed answers or situational information to a "conversation partner" (ie, a chatbot) could feel more natural and engaging for participants, as suggested by the trend to discuss mental health topics with large language models (LLMs) like ChatGPT [24]. Being able to record voice messages instead of having to type the answer might also reduce the burden of providing open answers and increase overall adherence.

Concerning the integration of open answers from participants, the emergence of LLMs also makes it easier to process potentially large amounts of open-ended responses, whether in text or spoken form. Advanced natural language processing (NLP) techniques allow researchers to efficiently analyze and categorize responses, identifying themes [25,26] and sentiment [27,28] associated with stress.

Semiautomated chatbots using emojis or GIFs for engagement are already used in marketing for customer satisfaction surveys and interactive support [29-31], showing the potential of using this medium for measurement and intervention. In addition, the low-threshold communication might reach individuals who are less frequently included in research studies, such as young people or people with a lower education [21].

The use of communication apps for work-related stress measurements also aligns with broader trends in digital health and workplace well-being initiatives. Many organizations are exploring digital tools to monitor employee stress and well-being, with the goal of developing targeted interventions and support systems [12,32]. If harnessing commonly used communication channels for work-related stress assessments proves to be feasible, this approach could be further developed into a personalized and low-effort monitoring and support system for employees.

This study adds valuable knowledge on the usability and feasibility of communication apps like WhatsApp as a tool for daily data collection in the pursuit of low-threshold (work-related) stress monitoring and interventions. In terms of usability, we are interested in who would be willing to participate in a study on work-related stress using this new data collection method (sample description). Another open question concerns how individuals perceive this method and their preferences regarding question types (ie, text, voice message, or MC buttons). In terms of feasibility, we need a better understanding of how this method performs in terms of uptake and adherence and the quality of information collected. In this study, we therefore investigate adherence rates as well as the effect of emojis on adherence, as emojis are a typical and easy-to-implement feature of mainstream WhatsApp communication. To evaluate the quality of the data collected, we will look at (1) potential differences in the number of reported contextual factors (ie, stressors and energy sources) per data collection method, (2) the number of words used to describe energy sources using voice or text messages, and (3) the relationship between the number of reported stressors and participants' stress ratings.

We formulated the following research questions for our exploratory study on usability (RQ1) and feasibility (RQ2 and RQ3), focusing on employees who took part in a 10-day diary study:

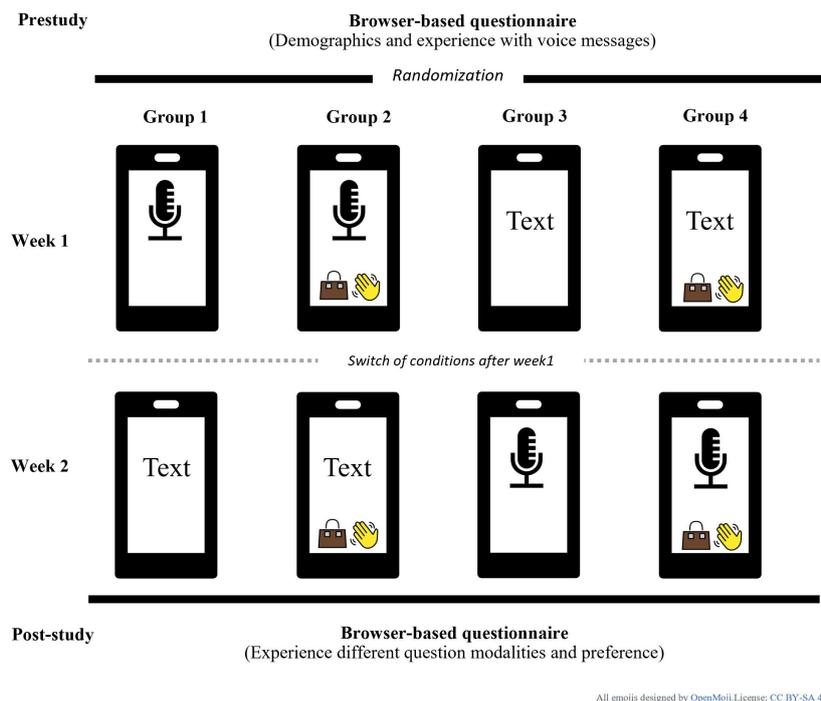
- RQ1: What is the user experience of participants over a 2-week period?
- RQ2: What is the level of adherence (percentage of daily measurements completed) in relation to the different collection methods and the use of emojis in WhatsApp?
- RQ3: What is the quality of the data in the different collection methods regarding the number of stressors and energy sources reported, the number of words used, and the relation of stressors to the stress rating?

## Methods

### Study Design

To answer the research questions, we adopted an experimental study design in a 2-week diary study on work-related stress and its contributing factors, using a semiautomated WhatsApp chatbot that asked 4 questions at the end of each workday. Each participant experienced both answer formats: voice messages or a combination of text messages and MC questions. Answer methods were switched after the first week: Half of the participants began answering questions using voice messages (voice condition) and would use MC messages (operationalized as template buttons) and text messages in the second week (text/MC-condition). The other half of the participants followed the reverse order. Additionally, both conditions were presented either with or without emojis, resulting in 4 experimental groups (Figure 1). For emojis, we chose more technical emojis, such as the battery symbol for the question on energy givers and the microphone for voice messages, to avoid adding emotional stimuli. Demographic data and user experience were collected via browser-based questionnaires before and after the daily WhatsApp interactions. Participants joined the study in 6 weekly waves between September and October 2024.

**Figure 1.** Study design (emojis adapted from [33], which is published under Creative Commons Attribution 4.0 International License [34]).



All daily messages were sent semiautomatically using the marketing tool [35], connected via the research organization's Meta account to a WhatsApp Business account created solely for this study. Eazy is an omnichannel marketing platform that is advertised to streamline and automate communication and support marketing campaigns; it is not routinely used in research contexts. Eazy programmed the 4 different experimental conditions as simple automatic, rule-based chatbots to interact with the participants. When participants answered a message, this triggered the next messages in the conversation script. The chatbot always instructed participants on how to respond, either by providing buttons with predefined answers or by requesting a text or voice message (Multimedia Appendix 1). Participants were informed at the beginning of the study that no human would monitor or engage in the WhatsApp conversation. All messages exchanged between the chatbot and participants were end-to-end encrypted.

### Ethical Considerations

The research proposal was approved by The Netherlands Organization for Applied Scientific Research (TNO) Internal Review Board (registration number: 2024-071) after reviewing the research design, privacy considerations, ethical aspects, and the potential burden and risks to participants. All participants provided informed consent via a web-based survey (Survalyzer) prior to enrolling in the WhatsApp conversation. Data from the browser-based questionnaires were linked to WhatsApp conversations using phone numbers, which were subsequently removed and replaced with pseudonyms. All analyses were conducted on pseudonymized data. Participants did not receive direct compensation but could enter a lottery for 50 Amazon gift vouchers (1 × €100, 3 × €50, 46 × €20; conversion rate on payment date: 1€ = US \$1.0580) if they completed at least 80% of daily measurements.

### Procedure

We recruited unsystematically on social media for 6 weeks. Recruitment started on LinkedIn because of its closeness to the work setting. There, we published and shared posts from the research team members and promoted one of the posts through a paid ThoughtLeader campaign through TNO as the responsible research organization. After a limited response, we collaborated with TNO's Marketing and Communication department, which professionally organized paid advertising campaigns on Meta platforms (Instagram and Facebook) using the official TNO account. Social media campaigns were not specifically targeted at specific groups, as we were interested in the broad working population. All social media posts included a picture of a phone, a short description of the study (duration, topic, inclusion criteria, lottery), highlighted the use of voice data, and included a direct link to the Survalyzer sign-up platform. People clicking on the link were shown detailed participant information, including contact information to reach the study team (via email). Participants who wanted to participate could digitally sign an informed consent. Participants were then asked to fill in a short prequestionnaire on demographics, use of voice messages, and the telephone number they wanted to use. At the start, participants were informed how to disable automatic messaging by sending a stop word in the WhatsApp chat. On the final page, participants were automatically randomized and either shown a QR code to initiate the WhatsApp conversation (waves 1-2) or informed that the study team would add them manually using their telephone number (waves 3-6). During the study, the procedure was changed from QR code initiation to manual addition due to technical difficulties in renewing QR codes promptly.

The semiautomated messages were scheduled separately for each wave by the study team, starting on Thursdays.

Messages were delivered Mondays through Fridays at 6 PM local time to the participants' own devices. After the first week, participants received a message informing them that they were halfway through the study and that the response method would now change. At the end of the study, participants were sent a link to an evaluation questionnaire and received a reminder one week later.

## Randomization and Blinding

Randomization was done fully automated in Survalyzer. Participants were randomly assigned to one of 4 groups. Data collectors were not blinded because telephone numbers from waves 3 to 4 had to be manually transferred from Survalyzer to the marketing tool by the study team. Participants were blinded to their assigned condition.

## Participants

Participants had to be 18 years or older, living in the Netherlands, and native Dutch speakers without a speech impairment. They were further required to be employed for at least 20 hours or 3 days per week, not to work night shifts, already have an existing WhatsApp account, and provide informed consent. Participants confirmed these criteria when providing informed consent.

## Measurements

All data was collected through the web-based questionnaire tool Survalyzer (pre- and postquestionnaire), and through WhatsApp messages managed via Eazy.

## Prequestionnaire

In the prequestionnaire, we assessed participants' gender (, woman, man, other), age ( $\leq 24$ , 25-34, 35-44, 45-54, 55+ years), number of weekly workdays (3 d, 4 d, 5 d), highest educational attainment (Dutch equivalents of low, medium, and high education), and frequency of voice message use on WhatsApp (never, occasionally, sometimes, often).

## Daily Measurements

Daily measurements always started with the question of whether the participant had worked that day. If not, no further questions followed, and participants were sent a closing message wishing them a pleasant day. On a workday, the same 3 questions about experienced stress, stressors, and energy sources were asked, but the response methods differed across experimental conditions.

Experienced stress was measured consistently across all days using a single item adapted from the SMS-based stress assessment by Arapovic-Johansson et al [36]. Participants were asked: "Stress is a state in which you feel tense, restless, nervous, or anxious. Have you experienced stress today?" Participants responded using a drop-down menu ranging from 1 ("not at all") to 10 ("very much").

The questions on experienced work-related stressors and energy sources were self-developed as there were no validated single-item measurements available. The question on work-related stressors was answered either in an

MC-format or with a voice message. For the MC-condition, we chose ten broad and commonly reported work-related stressors [37,38] to accommodate a broad target group (the amount of work; time pressure; emotional burden; constantly having to process information; negative atmosphere; work impediments/problems with systems or processes; uncertainty about tasks/task distribution or planning; conflicts or threats; lack of support or appreciation from supervisors; overlapping of work and private life; other [open answer]). Each stressor was presented separately with "yes" or "no" response options. In the voice condition, the question was phrased as follows, without examples: "What stressors did you experience today in your work? Share it with us in a voice message!." The question on energy sources was always an open question to be answered either by text or voice message ("Where did you draw energy from today or in which area did you have a good experience? Please describe this briefly in a [voice or text] message.>").

## Postquestionnaire

For the postquestionnaire, we measured participants' preferences and experiences with different communication methods using self-created items. We asked which method they found more pleasant ("Which method did you find more pleasant to use?"), easier ("Which method did you find easier to use?"), and safer ("Which method did you find safer to use?") to use, and which allowed them to express themselves best ("With which method could you best express your experiences or opinions?"). For each question, participants could choose between "voice messages," "text messages," "multiple-choice questions," and "no preference." We further asked if they experienced any technical problems ("Did you experience any technical problems?") with any of the methods. Three open-ended questions asked participants what they liked or disliked about each method ("What did you like or dislike about [text messages / voice messages / multiple-choice questions]? When or in what situation would you like to use [text messages / voice messages / multiple-choice questions]?").

Concluding, we asked what kind of methods participants would generally prefer in research ("If you were asked to participate in a study, which method would you prefer?") with the following options: "WhatsApp questionnaires (multiple-choice questions)," "WhatsApp text messages," "WhatsApp voice messages," "A (phone) conversation," "Online questionnaires in a browser (like this one)," "A questionnaire in a special research app," "Group discussion," "Other, namely: ...").

Participants who wished to enter the lottery as a token of appreciation for their effort were asked to enter their name and email address.

## Statistical Analysis

To answer research questions 1 on participants' experiences, we present descriptive statistics from the pre- and postquestionnaire as well as stress ratings from the daily measurements. For research question 2, we compared adherence rates between the different conditions. Adherence was defined

as the proportion of working days on which participants responded in the WhatsApp conversation, relative to the number of working days they reported in the prequestionnaire.

To test whether the use of emojis was related to the adherence, we conducted independent Wilcoxon rank tests between the emoji groups (groups 1 and 3) and the nonemoji groups (groups 2 and 4) for overall adherence and adherence per week.

To answer research question 3 concerning the differences in the number of named work-related stressors and energy sources in total, we compared groups using paired Wilcoxon signed-rank tests. The same method was used to compare the average number of words used in the text and voice answers on energy sources.

To compare the relation of number of named work-related stressors and the stress level, we report Spearman rank correlations between the number of work-related stressors reported per day/week and the corresponding stress ratings (same day and weekly average), stratified by condition. Before all analyses, we conducted a paired Wilcoxon signed-rank test to examine whether participants' average stress ratings differed significantly between Week 1 and Week 2. As no significant differences were found, all participants were included in subsequent analyses.

All statistical analyses were conducted in R (version 4.3.3; R Core Team) using nonparametric tests, as the assumption of normality was not met for most variables. A  $P < .05$  was set as a criterion for significance. Missing data were analyzed as observed and not imputed.

## Coding of Open Answers

Voice messages were transcribed using Whisper Model (large-v2) [39] from speech to text before content analysis as a well-tested process for transcription [40]. Transcription was performed locally to maintain data privacy. For quality assurance, we manually checked 10% of transcripts. All open answers in the daily measurements about work-related stressors and energy sources were coded into categorical variables using 2 runs with different degrees of supervision. First, a Chat-GPT-o1 model deployed in Microsoft Azure OpenAI [41] was used to generate potential labels for both work-related stressors and energy sources loosely informed by the Job Demands–Resources model [42] with instructions to also consider non-job-related factors. The prompts used to generate these labels are provided in [Multimedia Appendix 2](#). Coding runs were done separately for work-related stressors and energy sources and based on all open texts participants sent (ie, text and voice messages, plus additional text responses for extra stressors in the MC-condition). This resulted in a total of 311 different labels for work-related stressors and 582 labels for energy sources, as many were synonyms. We cross-checked results with 2 psychologists with expertise in the work health domain (JT and MBR).

Based on the review of 10% of the material, labels were combined and redefined into 29 labels for work-related stressors and 31 labels for energy sources. The translation of the redefined labels, their definition, and examples are provided in [Multimedia Appendix 3](#). In the second run, these newly defined labels were used, limiting output to the defined categories (classification task), including examples from the material. The prompts used for the second run are provided in [Multimedia Appendix 2](#). All answers were manually checked for coding accuracy by one of the researchers (JT). The final classification accuracy was 95%.

Open answers in the browser-based evaluation questionnaire were categorized manually by a social psychologist (MBR).

## Power

To answer our research questions on feasibility, we needed to include 102 participants to detect a large effect between groups ( $d=0.5$ ) with 80% power (based on a 2-sided independent  $t$  test). We set our recruitment goal to 250, assuming a conservative 50% dropout rate (defined as not starting or completing fewer than 80% of daily measurements) based on prior studies [19].

## Data Exclusion

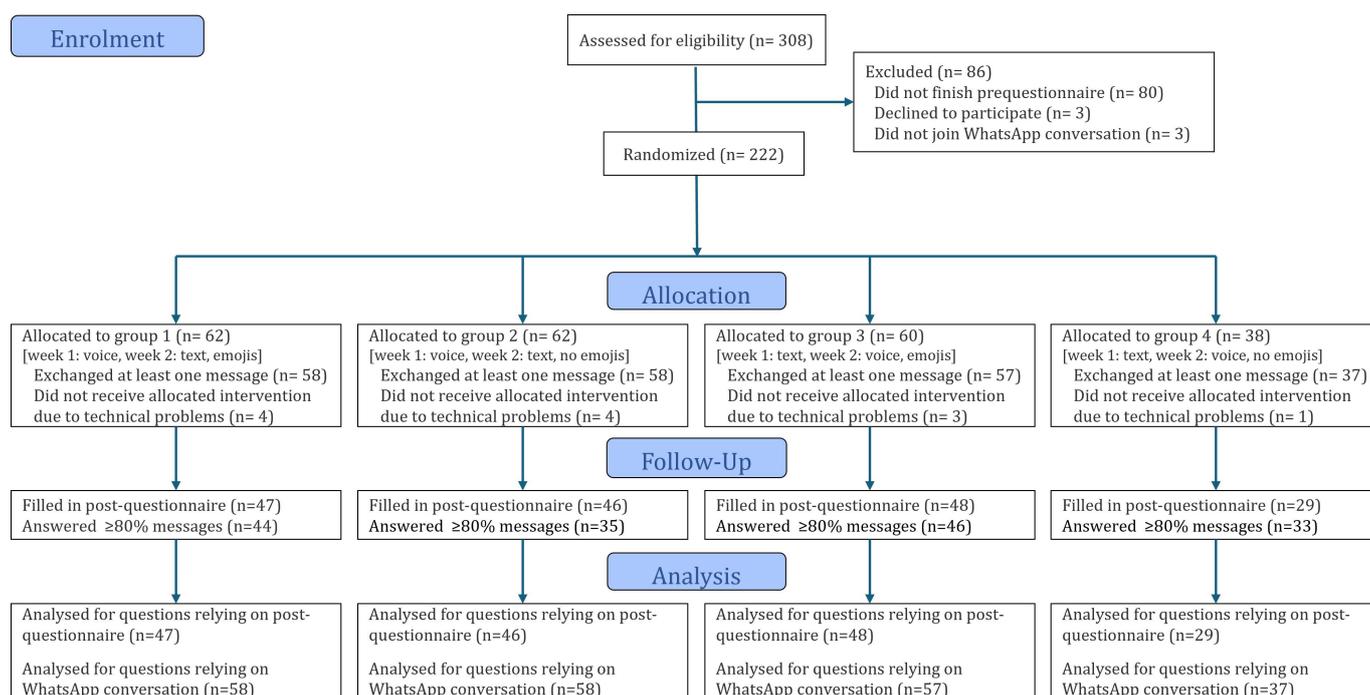
All messages sent by participants who did not answer a question were deleted (eg, wishing the chatbot a nice day or telling it that they would be free the next day) as were answers given after midnight, creating a 6-hour window for response. Duplicate questions and answers, as well as day entries exceeding 10 days due to technical issues, were removed; only the first valid response was retained. Conversations with 17 people from wave 1 and 2 had to be removed as their telephone numbers could not be matched to an informed consent, likely because the QR code for sign-up was shared or scanned using a different device.

## Results

### Participants

In total, 222 participants gave informed consent and filled in the prequestionnaire. A total of 210 (95%) participants successfully joined the WhatsApp conversation and completed at least one daily measurement. Furthermore, 158 (71%) participants completed the daily measurements for at least 80% (adjusted for working days) and 170 (81%) participants completed the poststudy evaluation questionnaire ([Figure 2](#)). The sample was predominantly women (170/210, 81%), highly educated (173/210, 82%), and working part-time (109/210, 52%) ([Table 1](#) displays the total sample and [Multimedia Appendix 4](#) for group characteristics). The average stress level across the study period was 4.66 (SD 1.73) and did not differ between groups.

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) flowchart of participants (adapted from Hopewell et al [43], which is published under Creative Commons Attribution 4.0 International License [34]).



Citation: CONSORT Flowchart based on the template provided by: Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 Statement: updated guideline for reporting randomised trials. *BMJ*. 2025; 388:e081123.

**Table 1.** Sample description based on prequestionnaire (N=210).

Variable and category	Values, n (%)
<b>Gender</b>	
Woman	170 (81)
Man	38 (18.1)
Other	2 (1)
<b>Age (years)</b>	
18 to 24	4 (1.9)
25 to 34	44 (21)
35 to 44	47 (22.4)
45 to 54	69 (32.9)
55 or older	46 (21.9)
<b>Education</b>	
Low	4 (1.9)
Medium	33 (15.7)
High	173 (82.4)
<b>Workdays per week</b>	
3 days	17 (8.2)
4 days	92 (44.2)
5 days	99 (47.6)
<b>Use of voice messages in everyday life</b>	
Never	34 (16.3)
Once in a while	73 (34.9)
Sometimes	70 (33.5)
Often	32 (15.3)

## RQ1: User Experience in Post Questionnaire

The postquestionnaire revealed a strong favor for the MC, as they were most frequently rated as more pleasant (75/170, 44%), easier (113/170, 67%), and safer (71/170, 42%) to use than the other methods, followed by text messages (Table 2). The open formats of text (57%) and voice messages (28%)

were named more often regarding the method with which participants could better express themselves. As a general preference for research participation, WhatsApp questionnaires with MC questions (85/170, 52%) were rated highest, followed by traditional browser-based questionnaires (30/170, 18%). Technical problems were experienced by 15 participants (8.8%) with the voice messages and by 2 participants (1.2%) with both voice and MC questions.

**Table 2.** Usability rating in postquestionnaire (n=170).

Variable and modality	Total, n (%)
Nicer method	
Voice message	14 (8.2)
Text message	67 (39.4)
MC <sup>a</sup> questions	75 (44.1)
No preference	14 (8.2)
Easier method	
Voice message	11 (6.5)
Text message	36 (21.2)
MC questions	113 (66.5)
No preference	10 (5.9)
Safer method	
Voice message	1 (0.6)
Text message	32 (18.8)
MC questions	71 (41.8)
No preference	66 (38.8)
Best to express oneself	
Voice message	48 (28.2)
Text message	96 (56.5)
MC questions	11 (6.5)
No preference	15 (8.8)
General preference for research method	
WhatsApp questionnaire (MC-questions)	85 (51.5)
Online questionnaire in a browser (like this one)	30 (18.2)
WhatsApp text messages	25 (15.2)
Other namely	9 (5.5)
WhatsApp voice message	8 (4.8)
(Telephone) call	4 (2.4)
Questionnaire in a special research app	3 (1.8)
Group discussion	1 (0.6)

<sup>a</sup>MC: multiple-choice.

The manually coded open answers to the question of what participants particularly liked or disliked about the different methods were consistent with the closed-question responses (Table 3). MC-questions yielded about 2.5 times more

positive (n=138) than negative (n=55) aspects, with ease and convenience most frequently cited (n=102), although some participants noted the lack of nuance and context (n=30).

**Table 3.** Positive and negative aspects of the 3 question modalities as named in the postquestionnaire (n=170).

Modalities	Total
Voice messages	
Positive aspects	120
Ease of completion, speed, convenience	41
Ability to think aloud and organize thoughts	41
Ability to add nuance to answers, provide elaboration, offer context, suitable for long and complex answers	17
Ability to express emotions, describe feelings, share/convey feelings	12
Ability to use own words or articulate the story well	9
Negative aspects	167
Not suitable for every situation and time, uncomfortable with bystanders	74
Unpleasant or uncomfortable to send voice messages	29
Answers cannot be adjusted, must be correct in one go, little room/time to think about the answer, no room for reflection	21
Difficulty in articulating thoughts	19
Uncomfortable sharing voice data, uncomfortable sharing personal information in this way, does not feel anonymous	17
Feels impersonal, you do not know who you are talking to	4
Technical problems when recording/sending voice messages	3
Text messages	
Positive aspects	198
Space for elaboration, nuance, providing context	54
Easy, accustomed to text messages, convenient, quick	38
Time to think about the answer and possibly adjust it, offers a moment of reflection	38
Can be completed at any time, anywhere	34
Pleasant for concise answers (or unpleasant for long answers)	14
Feels anonymous/privacy feels ensured	10
Space for elaboration, nuance, providing context	10
Negative aspects	23
It takes time, it is difficult to formulate an answer	21
Other negative comments that do not fall into any of the other categories or are not further specified	2
Multiple-choice buttons	
Positive aspects	138
Easy, quick, requires little effort from the respondent, convenient	102
Providing options gives respondents guidance, opportunity to reflect, offers frameworks	28
Other positive comments not falling into any of the other categories or not specified	8
Negative aspects	55
No possibility for nuance, providing context, lack of depth	30
Answer options were missing, options were not comprehensive, cannot share personal experiences	24
Technical problems/objections	1

Text messages were predominantly evaluated positively, with 198 positive and 23 negative aspects reported. Participants highlighted the space for nuance and elaboration (n=54), ease of use and convenience (n=38), and time to think about the answer (n=38) as positive. A total of 21 participants mentioned that they found it difficult and time-intensive to formulate their answer in text messages.

Voice messages elicited nearly 1.5 times more negative aspects (n=167) than positive aspects (n=120). Positive aspects were similar to the ones mentioned about text messages and included ease of completion (n=41), time to reflect (n=41), and ability to provide nuance and context (n=17). Negative aspects particularly emphasized that voice messages were not suitable in all settings and that the timing of prompts was often inconvenient for recording voice

messages (n=74). Twenty-nine participants reported feeling uncomfortable responding via voice messages.

## **RQ2: Adherence**

Mean adherence to the daily schedule was high (mean 95%, SD 33%). This is partly because 78 participants responded on more days than their typical working days, resulting in adherence rates of 100% or higher, even though adherence remains high after capping adherence rates at 100% (mean<sub>capped</sub> 85%, SD<sub>capped</sub> 23%). A screening of open responses suggests these participants may have had flexible schedules, as they mentioned working half days or completing tasks in peace. Mean adherence in week one was higher (mean 101%; mean<sub>capped</sub> 90%) than in week 2 (mean 89%; mean<sub>capped</sub> 79%). Participants tended to respond quickly to WhatsApp

messages, with 50% of responses received within 13 minutes (mean 11.01 min, SD 427.79; 25th percentile: 0.35 min, Median: 13.33 min, 75th percentile: 69.38 min).

No statistically significant difference in adherence was detected between

the groups with (mean 96%, SD 36%) and without (mean 946%, SD 31%;  $W=5021, P=.43$ ) emojis groups with the available data.

In week 1, a statistically significant difference was observed between the groups starting with the text condition (mean 107%, SD 24%) and in groups starting with the voice condition (mean 97%, SD 36%;  $W=6195, P=.046$ ). No statistically significant difference was observed in week 2 (starting with text: mean 93%, SD 34%; starting with voice: mean 85%, SD 46%;  $W=5660, P=.48$ ). The statistical test resulted in comparable results with the adherence rates capped to 100% (Multimedia Appendix 4).

### RQ3: Differences in Work-Related Stressors and Energy Sources Reported Per Modality

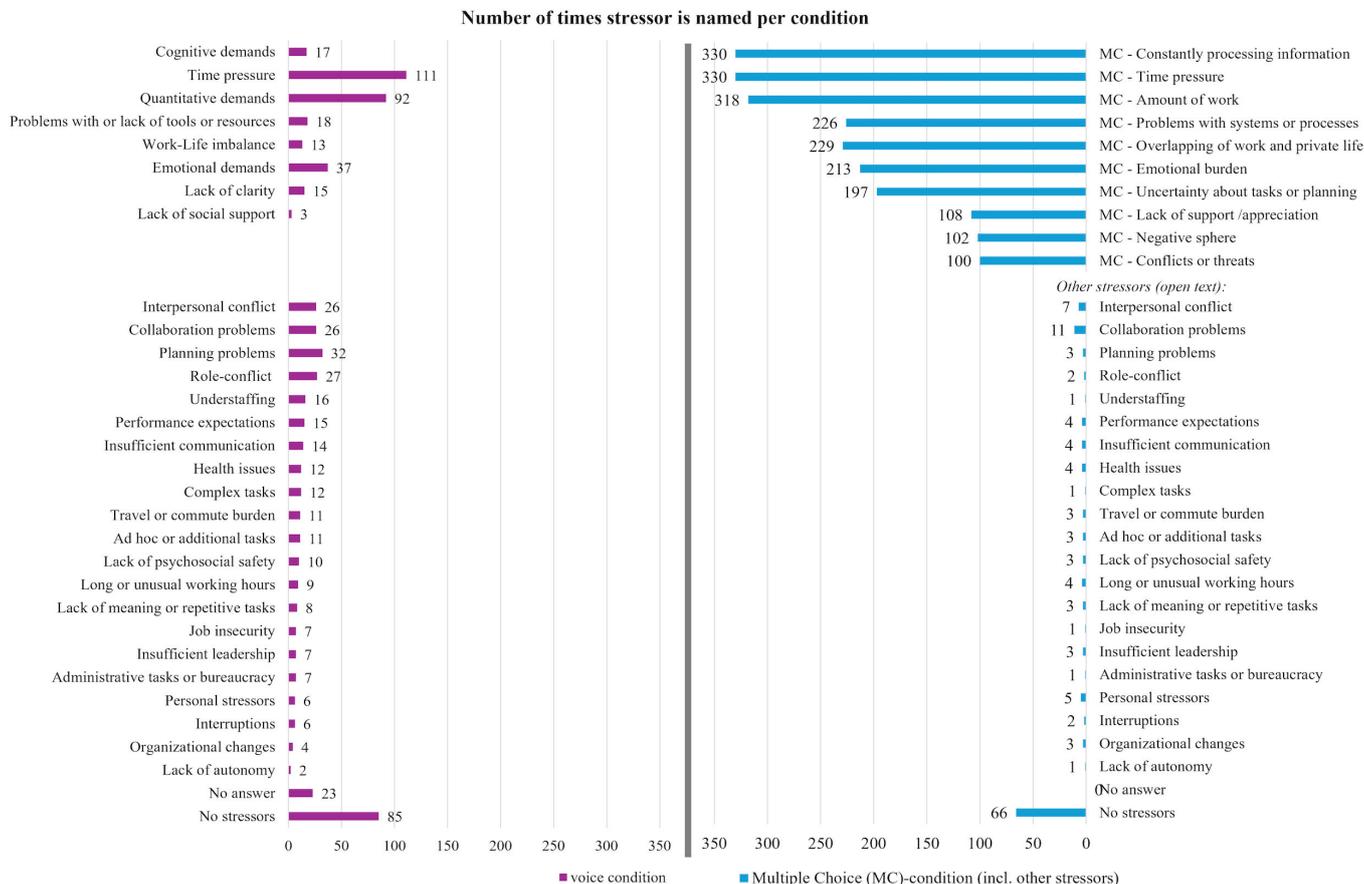
A total of 29 work-related stressors and 31 energy sources were identified in the open answers and used as categories

to analyze the content of open responses from voice and text messages, as well as open-text answers following selection of the MC category “Other stressors.” Not all MC options were directly represented in the categories for open responses; however, most work-related stressors from the MC options had equivalents in the open-text categories, such as “Amount of work” and “Quantitative demands.” The open-response categories also included unique items not present in the MC options, such as “Understaffing,” “Health,” and “Travel or commute burden.”

The average number of work-related stressors per day chosen in the MC-condition and coded from the open answers given as voice messages differed significantly ( $W=13,917, P<.001$ ). In the MC-condition participants reported an average of 3.32 (SD 2.15) work-related stressors per day compared with 1.3 (SD 0.96) work-related stressors named in voice messages.

The 2 conditions also differed in terms of which stressors were reported and their frequency. Figure 3 shows that MC options were selected between 100 times (“Conflicts or threats”) and 330 times (“Constantly having to process information”). In contrast, the most frequently reported stressor in the voice condition was “Time pressure,” mentioned 111 times.

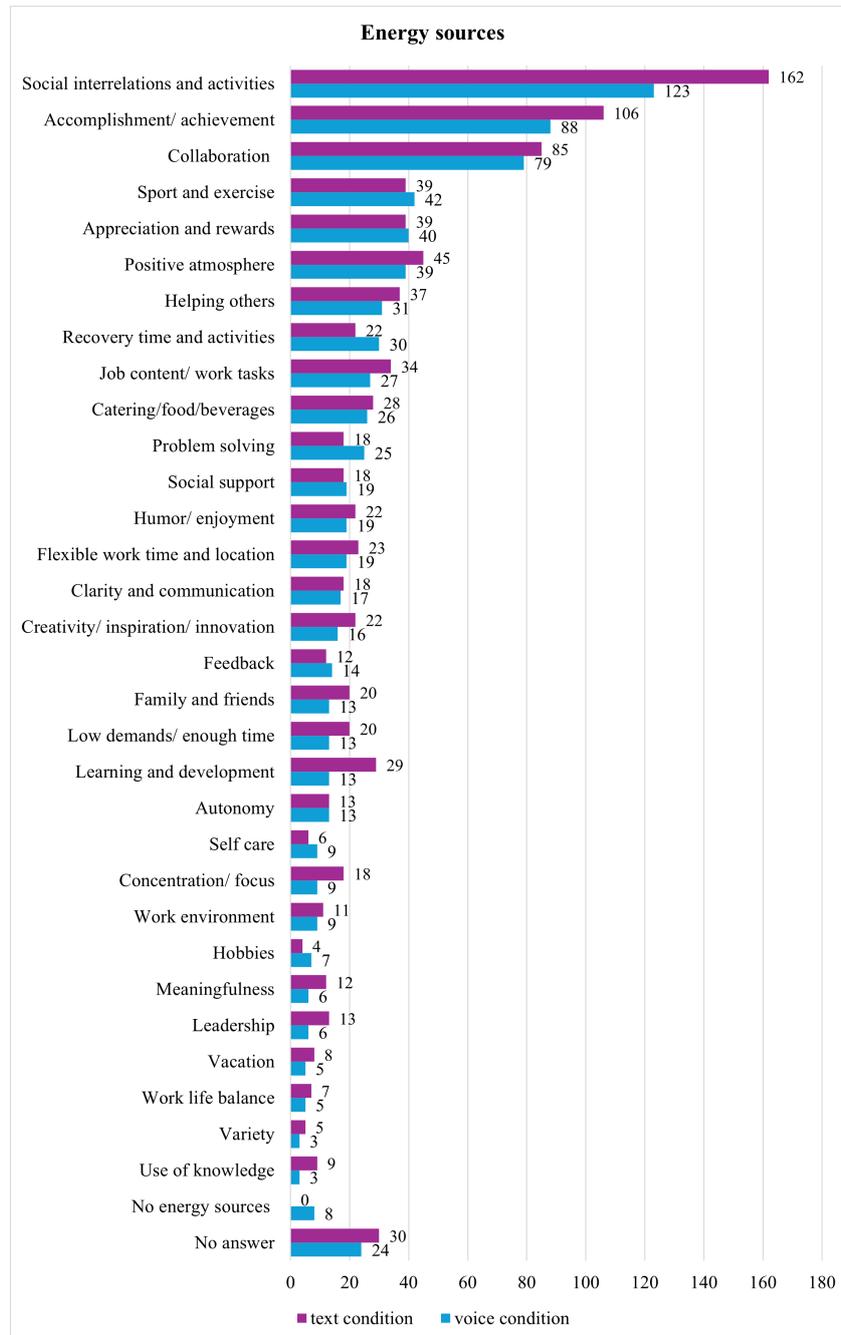
**Figure 3.** The number of total times work-related stressors were named in the voice or MC-condition, including open answers to the “other” category. MC-questions were not asked in the voice condition, and the closest category assigned to the open answers is presented alongside the MC-options. “Other stressors” described in open text were coded into the same categories as the voice answers. MC: multiple-choice.



The average number of energy sources per day reported in text messages and the number reported in voice messages did not differ significantly (voice: mean 1.65, SD 0.82 vs text: mean 1.55, SD 0.67;  $W=5231, P=.21$ ). On average, participants used more words over the whole week to describe their energy sources in voice messages (mean 28.15, SD 23.48) than in text messages (mean 16.67, SD 8.92;  $W=10,197, P<.001$ ).

The types of energy sources (Figure 4) were similar in both conditions, with the same top 3 energy sources (“Social interrelations and activities,” “Accomplishment/ achievement” and “Collaboration”) in both conditions.

**Figure 4.** The number of total times energy sources were coded from the open answers in the voice and text conditions.



Considering the relation between the number of named work-related stressors on a daily as well as weekly basis with the responding stress level, the correlation was higher in the MC-condition. The number of MC stressors showed a stronger correlation at the daily level ( $r=0.58$ ) and weekly

level ( $r=0.49$ ) compared with the number of stressors reported in the voice condition (daily:  $r=0.43$ ; weekly:  $r=0.34$ ).

## Discussion

### *Principal Findings*

The present study set out to explore the usability and feasibility of different methods for collecting daily stress experiences in the workplace through a semiautomated WhatsApp conversation. In terms of usability, participants (predominantly highly educated women) had a clear preference for MC-questions and text messages over voice messages as answer formats. Feasibility, in terms of adherence and data quality, can be considered good. However, differences in the number of reported work-related stressors and energy sources warrant cautious reflection on the suitability of these methods.

### *User Experience*

The results from the evaluation questionnaire indicate a strong preference of our sample for the MC-style questions with clickable buttons and text messages over voice messages. The main reasons cited were the convenience and ease of use in any situation, which is in line with general assumptions about question formats [44]. The biggest critique of voice messages was that not every situation was suitable for recording a voice message. This can be especially true in our study, as we had fixed-scheduled messages that could have occurred when participants were still in a work or commute environment, and the questions concerned that same environment. Personalizing the message schedule could therefore be advantageous. Crozier and Cassell [45] reported that participants perceived the flexibility of an audio diary as an advantage compared with a scheduled interview.

### *Adherence*

Overall adherence to the daily message (without limiting to usual working days) was high, with 95% of messages answered. Even when limiting adherence to the number of days usually worked (ie, adherence rate capped at 100%), adherence was 85%. This is exceptionally high, given that traditional diary studies are considered prone to missing entries [21]. In a recent systematic review [19], half of the EMA studies (defined as assessing well-being at least twice daily) reviewed reported daily completion rates averaging 72% (range 43%-95%). The review noted that associations between adherence and age or gender were rarely reported, but higher completion rates were observed in shorter studies (1-2 wk) [19,46]. In addition, responses to the initial message were typically quick (median 13 min), which is highly desirable for EMA studies [47]. The single daily measurement, short study duration, and restriction to working days in our study may explain the high adherence.

The use of emojis did not yield observable differences in adherence rates between groups. This is noteworthy given the vast research on emoji use in social marketing, which claims to enhance engagement [48,49] and to make chatbots more human-like [50,51]. However, this primarily relates to engagement on social media, such as “likes” and comments on specific posts or one’s own postings. In these contexts, the

appeal of emojis is often attributed especially to women and younger generations [48,52].

Starting with the voice message method did, however, result in slightly lower response rates in the first week, in line with participants’ more negative evaluation of this method. Nevertheless, adherence in both groups was above 85%, even if capped at 100%. This may reflect both the selective sample and a possible novelty effect of the collection method. Our sample lacked sufficient diversity in age, gender, and adherence to examine subgroup differences, which is highly needed [19].

### *Differences Between Modalities*

Our results indicate a great difference in the average number of reported work-related stressors per day in the different conditions (MC-conditions: mean 3.32, SD 2.15; voice: mean 1.3, SD 0.96) as well as the different number of stressors named (10 MC-options vs 29 options derived from open answers). Literature on questionnaire methodology suggests that recall and familiarity bias can play a big role in how MC-questions are answered [44]. The absence of cues or examples in the open questions may partly explain this large difference. Thus, the question of whether to use open or closed questions persists, even within new technical solutions. This notion is long-standing, as Dohrenwend [53] in 1965 already suggested that in well-conducted questionnaires, open questions do not bear higher validity or information than closed ones, and the choice of question should always depend on the research question. Similarly, in a questionnaire on resilience and coping mechanisms, the authors found no added value of open questions over closed ones in predicting mental health outcomes [54]. Our finding that the number of MC stressors correlated more strongly with stress ratings than open-answer stressors points in the same direction within the context of diary-based stress measures.

All things considered, it is noteworthy that the open answers received contained a lot of information and were barely skipped, as is often observed in traditional questionnaires. This highlights the potential of open answers to collect more detailed insights on new topics [53].

The observable difference in the number of words used to report energy sources, which was higher in the voice condition, may reflect the unedited nature of spoken responses, as participants also suggested in the post-questionnaire. This aligns with prior research from Crozier and Cassell [45], which showed that in-depth audio diaries can yield additional insights not captured in “edited” written responses or closed questions, which can be interesting for studying thought and reasoning processes. As diary and EMA studies—besides data collection—are also often reported to give insights and offer reflection moments, greater attention to the reflection process appears warranted [21]. In the evaluation questionnaire, both open-ended formats were mentioned to aid reflection, and voice messages were particularly described as helpful for “organising one’s thoughts.” Indeed, a study on teacher reflections found that written reflections were more summative, whereas spoken

reflections provided greater detail and reasoning [55]. In the context of measuring stress, reflecting on it and potentially acting on it, these findings highlight the need for a deeper understanding of these differences in modalities.

## Implications

The differences in assessed work-related stressors when using different methods support Crozier and Cassell's [45] call for pluralism in stress research. In the following section, we discuss situations and research aims for which the methods explored in this study may provide added value. Starting with participants' preferred method, the MC question, this format most closely resembles a traditional questionnaire, albeit with a different technical implementation. The evaluation survey showed a preference for WhatsApp questionnaires with buttons (MC-condition) from half the sample, followed by a traditional (browser-based) questionnaire and open-text questions in WhatsApp.

As the comparison to the open-answer question showed, participants reported on way more different stressors than we suggested with the MC options. MC questions are therefore most suitable when the symptoms or behaviors of interest are already well defined. To ensure relevant answer options, we recommend a participatory approach in which MC options are derived from the target group (eg, the most common stressor for a certain profession or type of work). Research questions then could center around incidence or accumulating effects of those stressors. To leave some more room for personalization and nuance, an alternative could be adaptive MC-questions, where next to predefined answers, participants can create their own answers, which are stored and shown to them in the following measurements as possible in specialized EMA apps [56]. However, this and other question types (eg, ranking or matrix questions) require technology beyond the consumer-ready marketing platforms currently available for social media.

The open text messages were also still conceived as convenient and quick for participants in the evaluation questionnaire. Interestingly, participants still occasionally ( $n=21$ ) described them as time-intensive in the postquestionnaire, a negative aspect not raised with the voice messages. Participants acknowledged that it was easier to describe, in both open answer formats, nuanced and in context, what they experienced that day. While closed MC-questions can create a quick picture of the frequency of, for example, different work-related stressors and energy sources, open questions may play a crucial role in informing answer options, particularly for novel topics [53] and in revealing contextual information that influences measurement. If we look at the amount of different work-related stressors and energy sources, we find a large variety, including work-related energy sources such as autonomy, social support, and feedback [57] and energy sources outside of work such as hobbies, time spent with family, and recovery, indicating the open questions to be a rich information source. Insights from open responses illustrate concrete situations, offering opportunities to develop personalized advice for individuals and employers.

The general willingness to respond in written form within a conversational-style assessment suggests promise for developing stress assessments based on conversational agents rather than static chatbots, which is a rapidly emerging field. As participants increasingly recognize the conversational benefits of LLMs such as ChatGPT, the ability to systematically assess information comparable to questionnaires presents an interesting avenue for future research [58-60]. Additionally, written electronic diary entries combined with LLMs open up new therapeutic approaches that were not feasible before the advent of these models [61]. In this regard, our study demonstrates how LLMs can be used to analyze written material efficiently and semiautomatically, based on scientific models, thereby reducing barriers to incorporating open responses in stress research.

Our study does not support the idea that giving the option to answer with voice messages increased participation; it merely did not seem to hinder it in the current sample. As a standalone collection method, voice messages cannot be recommended unless explicitly designed with and for a specific target group (eg, younger individuals, people with visual impairments, or those with language difficulties [21]); . Given the comparable content, one could give participants the option to choose whether they answer per text or voice message. However, this option is already available on most smartphones via a dictation feature. Specific research questions may still require voice recordings, in which case WhatsApp or other messaging services could provide a feasible collection method. For example, research using voice recordings to diagnose and monitor health and disease [62] could facilitate audio recordings or combine them with questionnaire items or other biomarkers, thereby minimizing patient burden.

## Limitations

Several limitations of this exploratory usability and feasibility study should be considered when interpreting the results. First, despite the envisioned low-threshold and inclusivity of WhatsApp as a research tool, our sample predominantly consisted of highly educated women, a demographic overrepresentation that is common in this type of research, as a systematic review revealed that more than half of EMA studies reported an overrepresentation of women and highly educated persons [19]. This imbalance was likely amplified by our recruitment strategy, which included outreach through scientific networks and opaque advertisement algorithms on social media platforms. Future studies should aim to diversify sampling strategies to ensure broader generalizability of the findings.

Second, while our study used a novel method for assessing stress, we did not compare this approach with established questionnaire-based assessments, nor did we validate the EMA items. This lack of methodological triangulation limits the evaluation of convergent validity and may affect interpretation of the measured constructs. This is a logical next step after proving the general feasibility of this collection method with this study.

Third, we used a single daily measurement, aligning our design more closely with traditional diary studies than with EMA protocols involving multiple daily prompts. End-of-day diary entries and EMA are not fully interchangeable, although they tend to yield more comparable results than weekly stress ratings [63]. This nuance may influence the granularity and ecological validity of the findings in our study. However, given that the overall burden was perceived as low in the evaluation questionnaire, this approach is likely also feasible with multiple prompts per day.

Fourth, logistical limitations in the technical setup may have impacted participant engagement. Participants could not begin automated data collection immediately after enrollment, as group assignment and message scheduling were handled manually. This could result in delays of up to one week, potentially reducing initial motivation. Additionally, we used 2 different data collection platforms—Survalyzer for informed consent and WhatsApp for the actual daily measures—due to ethical concerns of asking for consent via WhatsApp. Although no direct impact on participation was observed, this switch may have introduced variability in user experience and warrants refinement in future implementations. As assessments and feedback mechanisms become more complex, new technological challenges are likely to emerge.

Last, our approach to coding open-ended responses using LLMs is innovative but lacks established validation within psychological research. To mitigate this, we monitored outputs carefully, always manually double-checked 10% of responses at each stage with experienced psychologists and fully verified final categories before analysis, achieving high agreement between LLM coding and human coders. Nevertheless, the use of LLMs remains an emerging practice that requires further validation.

## Conclusions

Conducting diary studies via WhatsApp as a convenient, low-threshold method was perceived positively by participants, with a preference for MC questions and text messages over voice messages, due to their convenience and suitability for responding in any situation. However, our relatively homogeneous sample of mostly highly educated women leaves open the question of whether results would differ in more diverse samples. The high adherence, richness, and completeness of information gained through the study render this method feasible and promising for future-proof (work) stress research, provided it is carefully aligned with study goals and validated through direct comparison with traditional methods.

---

## Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. In accordance with the GAIDeT taxonomy (2025), GAI tools were used under full human supervision for proofreading and editing, as well as translation. The GAI tool used was the Copilot GPT-5 model on September 12 and December 4, 2025.

Responsibility for the content and integrity of the final manuscript rests entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. This declaration was submitted by JT.

Copilot was used to generate a list of suggested corrections for grammar and spelling in the manuscript, and all suggested changes were critically reviewed before being incorporated. The prompts in [Multimedia Appendix 2](#) were translated using Copilot and carefully reviewed by the first author. All Copilot conversations used in the preparation of the manuscript can be read in [Multimedia Appendix 5](#). The use of GAI as an analysis method is described in detail in the Method section.

---

## Funding

This research was supported by the Netherlands Organization for Applied Scientific Research Health & Work Research Program, which is funded by the Dutch Ministry of Economic Affairs and supported by the Dutch Ministry of Social Affairs and Employment, program number VP25.204.1-7. The funding body was not involved in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

---

## Data Availability

The datasets generated or analyzed during this study are not publicly available as they contain audio recordings that potentially include identifiable voice characteristics. Other nonsensitive data may be available from the corresponding author on reasonable request and subject to appropriate legal and ethical agreements.

---

## Authors' Contributions

EK, IN, JT, MBR, and SS contributed to the conceptualization and methodology of the study. JT and SS were responsible for data curation. Formal analysis was conducted by JT (statistics) and SS (large language models). JT and MBR carried out the investigation and validation. Project administration was managed by EK and SS. Visualization was performed by JT. JT prepared the original draft of the manuscript, and EK, IN, JT, MBR, and SS contributed to the review and editing of the manuscript.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Screenshots of conversation with the chatbot in both conditions (with emojis).

[\[PDF File \(Adobe File\), 804 KB-Multimedia Appendix 1\]](#)

### Multimedia Appendix 2

Prompts used for unsupervised coding of open answers about work-related stressors and energy sources.

[\[PDF File \(Adobe File\), 149 KB-Multimedia Appendix 2\]](#)

### Multimedia Appendix 3

Final categories used to code open answers on energy sources and stressors.

[\[PDF File \(Adobe File\), 121 KB-Multimedia Appendix 3\]](#)

### Multimedia Appendix 4

Sample description based split by groups and adherence rates when capped at 100%.

[\[PDF File \(Adobe File\), 242 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

Copilot conversations.

[\[PDF File \(Adobe File\), 4905 KB-Multimedia Appendix 5\]](#)

## References

1. Amiri S, Behnezhad S. Job strain and mortality ratio: a systematic review and meta-analysis of cohort studies. *Public Health*. Apr 2020;181:24-33. [doi: [10.1016/j.puhe.2019.10.030](https://doi.org/10.1016/j.puhe.2019.10.030)] [Medline: [31927337](https://pubmed.ncbi.nlm.nih.gov/31927337/)]
2. Harvey SB, Modini M, Joyce S, et al. Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occup Environ Med*. Mar 2017;74(4):301-310. [doi: [10.1136/oemed-2016-104015](https://doi.org/10.1136/oemed-2016-104015)] [Medline: [28108676](https://pubmed.ncbi.nlm.nih.gov/28108676/)]
3. Kivimäki M, Kawachi I. Work stress as a risk factor for cardiovascular disease. *Curr Cardiol Rep*. Sep 2015;17(9):630. [doi: [10.1007/s11886-015-0630-8](https://doi.org/10.1007/s11886-015-0630-8)] [Medline: [26238744](https://pubmed.ncbi.nlm.nih.gov/26238744/)]
4. O'Connor DB, Thayer JF, Vedhara K. Stress and health: a review of psychobiological processes. *Annu Rev Psychol*. Jan 4, 2021;72(1):663-688. [doi: [10.1146/annurev-psych-062520-122331](https://doi.org/10.1146/annurev-psych-062520-122331)] [Medline: [32886587](https://pubmed.ncbi.nlm.nih.gov/32886587/)]
5. Jackson AT, Frame MC. Stress, health, and job performance: what do we know? *J Appl Biobehavioral Res*. Dec 2018;23(4):e12147. [doi: [10.1111/jabr.12147](https://doi.org/10.1111/jabr.12147)]
6. Hassard J, Teoh KRH, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. *J Occup Health Psychol*. Jan 2018;23(1):1-17. [doi: [10.1037/ocp0000069](https://doi.org/10.1037/ocp0000069)] [Medline: [28358567](https://pubmed.ncbi.nlm.nih.gov/28358567/)]
7. Isham A, Mair S, Jackson T. Wellbeing and Productivity: A Review of the Literature. University of Surrey: Centre for the Understanding of Sustainable Prosperity; 2020.
8. Porcelli AJ, Delgado MR. Stress and decision making: effects on valuation, learning, and risk-taking. *Curr Opin Behav Sci*. Apr 2017;14:33-39. [doi: [10.1016/j.cobeha.2016.11.015](https://doi.org/10.1016/j.cobeha.2016.11.015)] [Medline: [28044144](https://pubmed.ncbi.nlm.nih.gov/28044144/)]
9. Shields GS, Sazma MA, Yonelinas AP. The effects of acute stress on core executive functions: a meta-analysis and comparison with cortisol. *Neurosci Biobehav Rev*. Sep 2016;68:651-668. [doi: [10.1016/j.neubiorev.2016.06.038](https://doi.org/10.1016/j.neubiorev.2016.06.038)] [Medline: [27371161](https://pubmed.ncbi.nlm.nih.gov/27371161/)]
10. Hirschle ALT, Gondim SMG. Stress and well-being at work: a literature review. *Cien Saude Colet*. Jul 8, 2020;25(7):2721-2736. [doi: [10.1590/1413-81232020257.27902017](https://doi.org/10.1590/1413-81232020257.27902017)] [Medline: [32667554](https://pubmed.ncbi.nlm.nih.gov/32667554/)]
11. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. *PLoS ONE*. 2021;16(3):e0248152. [doi: [10.1371/journal.pone.0248152](https://doi.org/10.1371/journal.pone.0248152)] [Medline: [33705457](https://pubmed.ncbi.nlm.nih.gov/33705457/)]
12. Zhu Z, Aitken JA, Dalal RS, Kaplan SA. The promise of just-in-time adaptive interventions for organizational scholarship and practice: conceptual development and research agenda. *Organ Res Methods*. Oct 2024;27(4):543-587. [doi: [10.1177/10944281231202740](https://doi.org/10.1177/10944281231202740)]
13. Crosswell AD, Lockwood KG. Best practices for stress measurement: How to measure psychological stress in health research. *Health Psychol Open*. 2020;7(2):2055102920933072. [doi: [10.1177/2055102920933072](https://doi.org/10.1177/2055102920933072)] [Medline: [32704379](https://pubmed.ncbi.nlm.nih.gov/32704379/)]
14. Alberdi A, Aztiria A, Basarab A. Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. *J Biomed Inform*. Feb 2016;59:49-75. [doi: [10.1016/j.jbi.2015.11.007](https://doi.org/10.1016/j.jbi.2015.11.007)] [Medline: [26621099](https://pubmed.ncbi.nlm.nih.gov/26621099/)]
15. Masri G, Al-Shargie F, Tariq U, Almughairbi F, Babiloni F, Al-Nashash H. Mental stress assessment in the workplace: a review. *IEEE Trans Affective Comput*. Jul 2024;15(3):958-976. [doi: [10.1109/TAFFC.2023.3312762](https://doi.org/10.1109/TAFFC.2023.3312762)]
16. Frisone F, Sicari F, Settineri S, Merlo EM. Clinical psychological assessment of stress: a narrative review of the last 5 years. *Clin Neuropsychiatry*. Apr 2021;18(2):91-100. [doi: [10.36131/cnfioritieditore20210203](https://doi.org/10.36131/cnfioritieditore20210203)] [Medline: [34909024](https://pubmed.ncbi.nlm.nih.gov/34909024/)]

17. Weckesser LJ, Dietz F, Schmidt K, Grass J, Kirschbaum C, Miller R. The psychometric properties and temporal dynamics of subjective stress, retrospectively assessed by different informants and questionnaires, and hair cortisol concentrations. *Sci Rep*. Jan 31, 2019;9(1):1098. [doi: [10.1038/s41598-018-37526-2](https://doi.org/10.1038/s41598-018-37526-2)] [Medline: [30705360](https://pubmed.ncbi.nlm.nih.gov/30705360/)]
18. Augustine A. *Experience Sampling Encyclopedia of Quality of Life and Well-Being Research*. Springer International Publishing; 2023. [doi: [10.1007/978-3-031-17299-1](https://doi.org/10.1007/978-3-031-17299-1)] ISBN: 978-3-031-17298-4
19. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. *J Happiness Stud*. 2021;22(5):2361-2408. [doi: [10.1007/s10902-020-00324-7](https://doi.org/10.1007/s10902-020-00324-7)] [Medline: [34720691](https://pubmed.ncbi.nlm.nih.gov/34720691/)]
20. Kidman PG, Curtis RG, Watson A, Maher CA. When and why adults abandon lifestyle behavior and mental health mobile apps: scoping review. *J Med Internet Res*. Dec 18, 2024;26:e56897. [doi: [10.2196/56897](https://doi.org/10.2196/56897)] [Medline: [39693620](https://pubmed.ncbi.nlm.nih.gov/39693620/)]
21. McCombie C, Esponda GM, Schmidt U, Lawrence V. Qualitative diary methods in mental health research: a scoping review and recommendations for research and reporting. *Eur Psychol*. Jan 2024;29(1):3-16. [doi: [10.1027/1016-9040/a000511](https://doi.org/10.1027/1016-9040/a000511)]
22. WhatsApp: number of monthly active users 2025. Statista. 2025. URL: <https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/> [Accessed 2025-05-15]
23. Number of WhatsApp users in the Netherlands. Statista. 2023. URL: <https://www.statista.com/statistics/880842/number-of-whatsapp-users-in-the-netherlands/> [Accessed 2025-05-15]
24. Jung K, Lee G, Huang Y, Chen Y. "I've talked to ChatGPT about my issues last night.": Examining Mental Health Conversations with Large Language Models through Reddit Analysis. arXiv. Preprint posted online on Apr 29, 2025. [doi: [10.48550/arXiv.2504.20320](https://doi.org/10.48550/arXiv.2504.20320)]
25. Bunt HL, Goddard A, Reader TW, Gillespie A. Validating the use of large language models for psychological text classification. *Front Soc Psychol*. 2025;3:1460277. [doi: [10.3389/frsps.2025.1460277](https://doi.org/10.3389/frsps.2025.1460277)]
26. Qiao T, Walker C, Cunningham C, Koh YS. Thematic-LM: a LLM-based multi-agent system for large-scale thematic analysis. Presented at: WWW '25: Proceedings of the ACM Web Conference 2025; Apr 28 to May 2, 2025:649-658; Sydney NSW Australia. URL: <https://dl.acm.org/doi/proceedings/10.1145/3696410> [doi: [10.1145/3696410.3714595](https://doi.org/10.1145/3696410.3714595)]
27. Doan S, Ritchart A, Perry N, Chaparro JD, Conway M. How do you #relax when you're #stressed? a content analysis and infodemiology study of stress-related tweets. *JMIR Public Health Surveill*. Jun 13, 2017;3(2):e35. [doi: [10.2196/publichealth.5939](https://doi.org/10.2196/publichealth.5939)] [Medline: [28611016](https://pubmed.ncbi.nlm.nih.gov/28611016/)]
28. Murase S, Andritsch J. Integrating sentiment analysis to enhance mental health support chatbots. Presented at: 2nd International Conference on Computing and Data Analytics (ICCDa); Nov 12-13, 2024:1-6; Shinas, Oman. [doi: [10.1109/ICCDa64887.2024.10867341](https://doi.org/10.1109/ICCDa64887.2024.10867341)]
29. Leung CH, Yan Chan WT. Retail chatbots: the challenges and opportunities of conversational commerce. *JDSMM*. 2020;8(1):68. [doi: [10.69554/APSB6546](https://doi.org/10.69554/APSB6546)]
30. Lo Presti L, Maggiore G, Marino V, Resciniti R. Mobile instant messaging apps as an opportunity for a conversational approach to marketing: a segmentation study. *JBIM*. May 23, 2022;37(7):1432-1448. [doi: [10.1108/JBIM-02-2020-0121](https://doi.org/10.1108/JBIM-02-2020-0121)]
31. Mathews S, Lee SE. Use of emoji as a marketing tool: an exploratory content analysis. *Fashion Ind Educ*. Jun 30, 2018;16(1):46-55. [doi: [10.7741/FIE.2018.16.1.046](https://doi.org/10.7741/FIE.2018.16.1.046)]
32. Devaux M, Cecchini M, Vazquez-Venegas P, Aihara H. Digital and innovative tools for better health and productivity at the workplace. OECD; Aug 14, 2024. 169. [doi: [10.1787/2ce05eb6-en](https://doi.org/10.1787/2ce05eb6-en)]
33. OpenMoji. URL: <https://openmoji.org/> [Accessed 2026-03-02]
34. Attribution 4.0 international (CC BY 4.0). Creative Commons. URL: <https://creativecommons.org/licenses/by/4.0/> [Accessed 2026-03-02]
35. Spotler message. Spotler. URL: <https://spotler.com/message> [Accessed 2025-12-01]
36. Arapovic-Johansson B, Wåhlin C, Kwak L, Björklund C, Jensen I. Work-related stress assessed by a text message single-item stress question. *Occup Med (Lond)*. Dec 2, 2017;67(8):601-608. [doi: [10.1093/occmed/kqx111](https://doi.org/10.1093/occmed/kqx111)] [Medline: [29016877](https://pubmed.ncbi.nlm.nih.gov/29016877/)]
37. Cooper CL, Marshall J. Occupational sources of stress: a review of the literature relating to coronary heart disease and mental ill health. In: *From Stress to Wellbeing*. 2013:3-23. URL: <http://link.springer.com/10.1057/9781137310651> ISBN: 978-1-349-33630-2
38. Pindek S, Spector PE. Organizational constraints: a meta-analysis of a major stressor. *Work & Stress*. Jan 2, 2016;30(1):7-25. [doi: [10.1080/02678373.2015.1137376](https://doi.org/10.1080/02678373.2015.1137376)]
39. Whisper: open-source speech recognition model. GitHub. URL: <https://github.com/openai/whisper> [Accessed 2025-06-23]

40. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. arXiv. Preprint posted online on Dec 6, 2022. [doi: [10.48550/arXiv.2212.04356](https://doi.org/10.48550/arXiv.2212.04356)]
41. Azure AI foundry model. Microsoft. URL: <https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/models> [Accessed 2025-06-23]
42. Bakker AB, Demerouti E, Sanz-Vergel AI. Burnout and work engagement: the JD–R approach. *Annu Rev Organ Psychol Organ Behav.* Mar 21, 2014;1(1):389-411. URL: <https://www.annualreviews.org/toc/orgpsych/1/1> [doi: [10.1146/annurev-orgpsych-031413-091235](https://doi.org/10.1146/annurev-orgpsych-031413-091235)]
43. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ.* Apr 14, 2025;389:e081123. [doi: [10.1136/bmj-2024-081123](https://doi.org/10.1136/bmj-2024-081123)] [Medline: [40228833](https://pubmed.ncbi.nlm.nih.gov/40228833/)]
44. Thomander SD, Krosnick JA. Question and questionnaire design. In: Edlund JE, Nichols AL, editors. *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences.* 1st ed. Cambridge University Press; 2024:352-370. [doi: [10.1017/9781009000796.017](https://doi.org/10.1017/9781009000796.017)] ISBN: 978-1-009-00079-6]
45. Crozier SE, Cassell CM. Methodological considerations in the use of audio diaries in work psychology: adding to the qualitative toolkit. *J Occup Organ Psychol.* Jun 2016;89(2):396-419. [doi: [10.1111/joop.12132](https://doi.org/10.1111/joop.12132)] [Medline: [27708502](https://pubmed.ncbi.nlm.nih.gov/27708502/)]
46. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. *Assessment.* Apr 2023;30(3):825-846. [doi: [10.1177/10731911211067538](https://doi.org/10.1177/10731911211067538)] [Medline: [35016567](https://pubmed.ncbi.nlm.nih.gov/35016567/)]
47. Janssens KAM, Bos EH, Rosmalen JGM, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. *BMC Med Res Methodol.* Nov 16, 2018;18(1):140. [doi: [10.1186/s12874-018-0579-6](https://doi.org/10.1186/s12874-018-0579-6)] [Medline: [30445926](https://pubmed.ncbi.nlm.nih.gov/30445926/)]
48. Koch TK, Romero P, Stachl C. Age and gender in language, emoji, and emoticon usage in instant messages. *Comput Human Behav.* Jan 2022;126:106990. [doi: [10.1016/j.chb.2021.106990](https://doi.org/10.1016/j.chb.2021.106990)]
49. Wu R. It's fun to play: emoji usage in interactive marketing communication. In: Wang CL, editor. *The Palgrave Handbook of Interactive Marketing.* Springer International Publishing; 2023:893-913. [doi: [10.1007/978-3-031-14961-0\\_38](https://doi.org/10.1007/978-3-031-14961-0_38)] ISBN: 978-3-031-14960-3
50. Meng H, Lu X, Xu J. The impact of chatbot response strategies and emojis usage on customers' purchase intention: the mediating roles of psychological distance and performance expectancy. *Behav Sci (Basel).* Jan 23, 2025;15(2):117. [doi: [10.3390/bs15020117](https://doi.org/10.3390/bs15020117)] [Medline: [40001748](https://pubmed.ncbi.nlm.nih.gov/40001748/)]
51. Nguyen M, Casper Ferm L, Quach S, Pontes N, Thaichon P. Chatbots in frontline services and customer experience: an anthropomorphism perspective. *Psychology and Marketing.* Nov 2023;40(11):2201-2225. URL: <https://onlinelibrary.wiley.com/toc/15206793/40/11> [doi: [10.1002/mar.21882](https://doi.org/10.1002/mar.21882)]
52. Bai Q, Dan Q, Mu Z, Yang M. A systematic review of emoji: current research and future perspectives. *Front Psychol.* 2019;10:2221. [doi: [10.3389/fpsyg.2019.02221](https://doi.org/10.3389/fpsyg.2019.02221)] [Medline: [31681068](https://pubmed.ncbi.nlm.nih.gov/31681068/)]
53. Dohrenwend B. Some effects of open and closed questions on respondents' answers. *Hum Organ.* Jun 1965;24(2):175-184. [doi: [10.17730/humo.24.2.u5838w33858455u3](https://doi.org/10.17730/humo.24.2.u5838w33858455u3)]
54. Friberg O, Rosenvinge JH. A comparison of open-ended and closed questions in the prediction of mental health. *Qual Quant.* Apr 2013;47(3):1397-1411. [doi: [10.1007/s11135-011-9597-8](https://doi.org/10.1007/s11135-011-9597-8)]
55. Lambert MD, Sorensen TJ, Elliott KM. A comparison and analysis of preservice teachers' oral and written reflections. *JAE.* Oct 30, 2014;55(4):85-99. [doi: [10.5032/jae.2014.04085](https://doi.org/10.5032/jae.2014.04085)]
56. Multiple choice question – manual. m-Path. 2025. URL: <https://manual.m-path.io/knowledge-base/multiple-choice/#the-other-option-and-self-learning-multiple-choice-questions> [Accessed 2025-06-16]
57. Bakker AB, Demerouti E. Job demands-resources theory: frequently asked questions. *J Occup Health Psychol.* Jun 2024;29(3):188-200. [doi: [10.1037/ocp0000376](https://doi.org/10.1037/ocp0000376)] [Medline: [38913705](https://pubmed.ncbi.nlm.nih.gov/38913705/)]
58. Liu J, Gu J, Tong M, et al. Evaluating the agreement between ChatGPT-4 and validated mental health scales in older adults: a cross-sectional study. *Am J Geriatr Psychiatry.* Oct 2025;33(10):1049-1061. [doi: [10.1016/j.jagp.2025.04.213](https://doi.org/10.1016/j.jagp.2025.04.213)] [Medline: [40393915](https://pubmed.ncbi.nlm.nih.gov/40393915/)]
59. Maurya RK, Montesinos S, Bogomaz M, DeDiego AC. Assessing the use of chatgpt as a psychoeducational tool for mental health practice. *Couns and Psychother Res.* Mar 2025;25(1):e12759. [doi: [10.1002/capr.12759](https://doi.org/10.1002/capr.12759)]
60. Vossen P, Báez Santamaría S, Baier T. A conversational agent for structured diary construction enabling monitoring of functioning & well-being. In: Lorig F, Tucker J, Dahlgren Lindström A, Dignum F, Murukannaiah P, Theodorou A, Yolum P, editors. *Frontiers in Artificial Intelligence and Applications* IOS Press. 2024. [doi: [10.3233/FAIA240204](https://doi.org/10.3233/FAIA240204)]
61. Kim T, Bae S, Kim HA, et al. MindfulDiary: harnessing large language model to support psychiatric patients' journaling. Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024:1-20; Honolulu HI USA. May 11, 2024. URL: <https://dl.acm.org/doi/proceedings/10.1145/3613904> [doi: [10.1145/3613904.3642937](https://doi.org/10.1145/3613904.3642937)]
62. Savage N. AI listens for health conditions. *Nature New Biol.* May 22, 2025. [doi: [10.1038/d41586-025-01598-8](https://doi.org/10.1038/d41586-025-01598-8)] [Medline: [40404963](https://pubmed.ncbi.nlm.nih.gov/40404963/)]

63. Schneider S, Junghaenel DU, Gutsche T, Mak HW, Stone AA. Comparability of emotion dynamics derived from ecological momentary assessments, daily diaries, and the day reconstruction method: observational study. *J Med Internet Res*. Sep 24, 2020;22(9):e19201. [doi: [10.2196/19201](https://doi.org/10.2196/19201)] [Medline: [32969835](https://pubmed.ncbi.nlm.nih.gov/32969835/)]

### Abbreviations

**EMA:** ecological momentary assessment

**LLM:** large language model

**MC:** multiple-choice

**TNO:** The Netherlands Organization for Applied Scientific Research

*Edited by Amaryllis Mavragani; peer-reviewed by Robert Spang; submitted 12.Sep.2025; final revised version received 22.Dec.2025; accepted 24.Dec.2025; published 11.Mar.2026*

*Please cite as:*

*Thielecke J, Bakhuis Roozeboom M, Niks I, de Korte E, Shahmohammadi S*

*Using Semiautomated WhatsApp Messages for Daily Stress Measurements: Integrated Usability and Feasibility Study*  
*JMIR Form Res 2026;10:e84032*

*URL: <https://formative.jmir.org/2026/1/e84032>*

*doi: [10.2196/84032](https://doi.org/10.2196/84032)*

© Janika Thielecke, Maartje Bakhuis Roozeboom, Irene Niks, Elsbeth de Korte, Sadegh Shahmohammadi. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 11.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.