

Supplementary materials: Current efforts in using open-ended items in ESM

For the manuscript “Experience sampling methods require more than numbers” (Laura F. Bringmann, Guðrún R. Guðmundsdóttir, Leonie Schorrepp, Leonie V. D. E. Vogelsmeier, Melissa M. De Smet, Charlotte Vrijen, Eeske van Roekel, Jessica Fritz, Jennifer X. Chen, Olivia Kirtley, Markus I. Eronen, Joanne M. Chung, Marie Stadel, and Katie Hoemann)

Method

At the outset of this project, anecdotal evidence led us to suspect that open-ended items are infrequently included in current ESM research. Even when such data are collected, it seems that researchers often do not systematically analyze them or report their use in published articles. To gain more insight into current practices in the field, we decided to: (1) survey ESM researchers about their use of open-ended items; (2) examine the use of open-ended items submitted to the ESM item repository (Kirtley et al., 2025); and (3) review the use and reporting of open-ended items in the published ESM literature.

We considered items open-ended if they provide open space for freely typed or recorded text-entry by participants. This also includes items with a fixed set of response options and an additional 'Other' option that provides such an open space for free text entry. However, we excluded “open-ended in disguise” items, where open-ended textboxes were actually used to record quantitative data, such as the number of times a type of behavior had been performed that day.

All appendices can be found on the Open Science Framework page of the project: <https://osf.io/jemrp>.

Survey

We conducted a brief online survey targeting ESM researchers. This study followed the relevant ethical guidelines and was exempt from regular review based on a low-risk self-assessment, using criteria established by the Ethics Committee of the Faculty of Behavioural and Social Sciences of the University of Groningen (PSY-2425-S-0199). The survey was made in Qualtrics and was distributed via the authors' professional networks, aiming to include as many researchers conducting ESM studies as possible. Specifically, the survey was shared within the [Measurement is the new black \(MITNB\) consortium](#), the [Dutch-Belgian ESM network](#), and the [Society for Ambulatory Assessment \(SAA\)](#). The survey link was also posted on Bluesky and LinkedIn. In addition, we applied the snowball method, encouraging researchers to share the survey with their colleagues. The survey included questions about the use of open-ended items within ESM research, specifically, if open-ended questions were used at all, and if so, for what purpose and which analysis approaches were employed. We further asked whether open-ended items and associated analyses were described in published articles and prompted researchers to add the reference for the publication. Moreover, we included open-ended follow-

up questions about the reasons for not using open-ended items. The full survey can be found in Appendix 1. In total, N = 78 researchers responded.

ESM Item Repository Search

We screened the ESM Item Repository (Kirtley et al., 2025; <https://esmitemrepositoryinfo.com/>) for open-ended item entries. We exported all items from the repository on the 21st of January 2025 and manually extracted all (potentially) open-ended items (N = 392). Thereby, we noticed that some open-ended items were follow-up questions that were only understandable when also including the stem question. Consequently, we also extracted relevant stem questions. First, we reviewed all exported items against our definition of open-ended items, resulting in a final set of relevant items (N = 157). Subsequently, we reviewed all items and labelled them according to two dimensions: item structure and content. Regarding the structure of the item, we differentiated “fully open” (i.e., open text box for free text), “partially open” (categorical/multiple choice items with an editable ‘other’ option), “fully open/list” (open text box for free text but phrased to invite a list or categories rather than running text). We additionally specified whether the items were stand-alone or follow-up items. Regarding the item content, we descriptively coded the construct or phenomenon the item intended to measure, resulting in a content analysis to summarise the topics that have been assessed open-endedly in ESM research to date. To illustrate, the item “Think about the most important event that will take place today. What is it?” was thus labeled as “fully-open stand-alone” (item type_structure) and “prospection (day) important event” (item type_content). Coding was done by three coders (i.e., LS, MS, and MDS) and discussed until consensus was reached. In addition, we tried to identify any connected publications to track down whether the items were mentioned and analysed. An overview of the final set of items, organized by item topic, can be found in Appendix 2, tab “grouped by item topic”; the selected items that we could find reported and analysed in research papers (N = 19) are listed in Appendix 2, tab “items mentioned and analysed”. After collecting open-ended items from the survey and the literature review, the coding was also applied to items found with these sources.

Literature Review

We searched for literature that used open-ended items in their study design. We began by conducting a targeted literature search of peer-reviewed studies reporting the use of open-ended items in their ESM protocol, which also identified relevant daily diary studies. We searched Google Scholar, APA PsycArticles, and APA PsycINFO using a combination of search terms for ESM (e.g., “experience sampling method”, “ecological momentary assessment”) and open-ended items (e.g., “open text box”, “open ended”, “qualitative data”, “mixed-method”). We included additional relevant articles (4) that we knew of but the original search had not captured. We retained entries if an open-ended item was reported as part of the study design, regardless of whether these data had been analyzed or results reported. As an exception to our peer-review criterion, we included an influential dissertation that has substantially shaped this field of research (Delespaul, 1995). In addition to the exact phrasing of the open-ended items, we coded for the topic studied, sample characteristics, study design, the nature of the open-ended

item, data processing, and analysis. We also coded the items for content and structure following the same procedure as for the ESM item repository (see above). An overview of the complete search strings and all identified studies can be found in Appendix 3.

Results

Use of open-ended items

One clear pattern from all three sources (survey, repository, literature search) is the discrepancy between collecting open-ended items and using or reporting them. Most of the 78 survey respondents reported including some form of open-ended item in their ESM research (N = 56; see Figure 1). It is likely that the researchers responding to the survey call were already more interested in using open-ended items than the average researcher, but even in this select group, most of those who collected open-ended item data did not further analyze or report them (N = 37; note however that some studies were still ongoing). We see a similar pattern in our analysis of the ESM Item Repository, where we found 392 open-ended items, of which only 19 items were actually analyzed in the publications (10 unique publications, out of which two used the same dataset, resulting in 9 unique studies). Similarly, when we look at the literature review, we see that often papers did not even mention the open-ended items, and of the 47 ESM papers that did mention open-ended items, only 28 also analyzed them. Combined with the nine papers found through the ESM Item Repository search, and three papers we discovered via the survey, this resulted in 40 ESM publications where open-ended items were actually mentioned and analyzed (see Table 1). Nearly half (N = 18) of the ESM papers where open-ended items were mentioned and analyzed were published in the past 5 years (2020 or later).

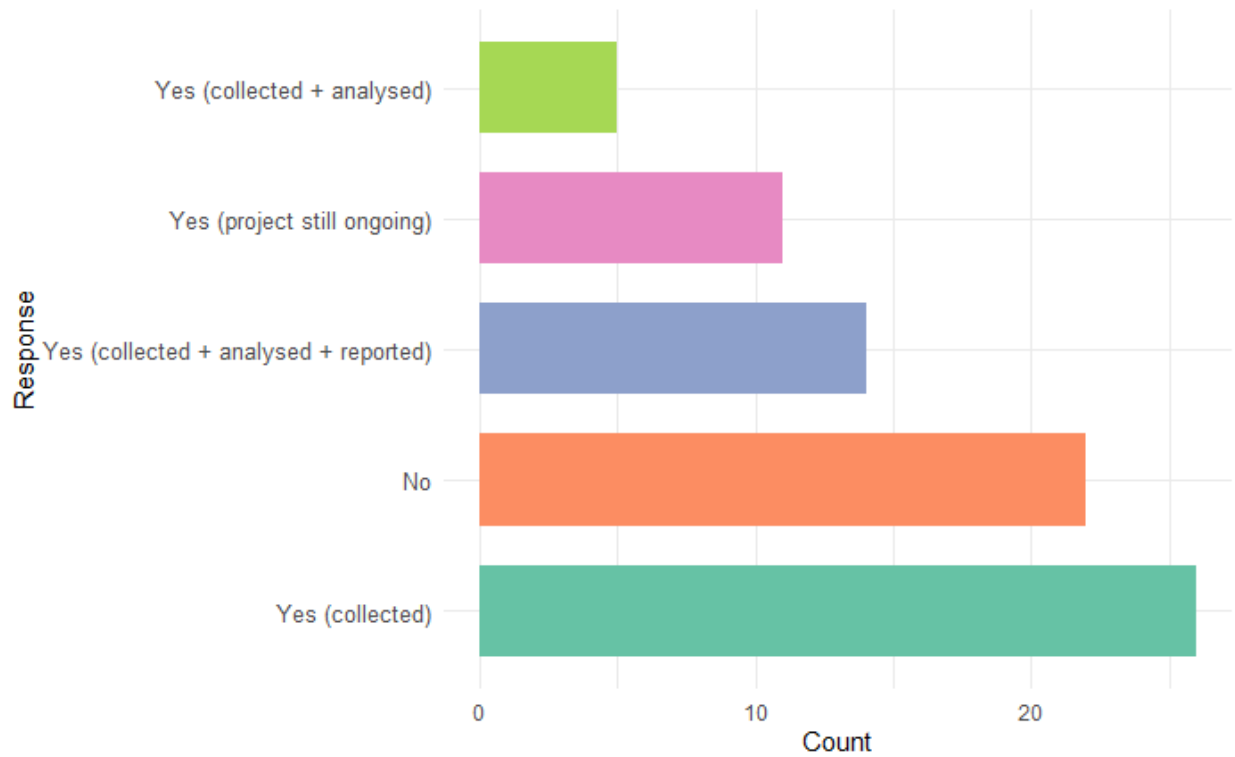


Figure 1. Use of open-ended items. This histogram shows the responses to the question “Did you ever use an open-ended item during your ESM research?” See Appendix 1 for the full survey.

Reasons for not using or reporting open-ended items

In the survey, we asked the follow-up question “What are your reasons for never using open-ended items during your ESM research?” to those respondents who reported never having used an open-ended item in their ESM research (N = 22; see Figures 1 and 2). The most common reason given was not knowing how to analyze them. Several respondents explicitly pointed out that they perceive them as harder to analyze than closed questions. Other reasons mentioned included high participant burden and privacy/ethical concerns.

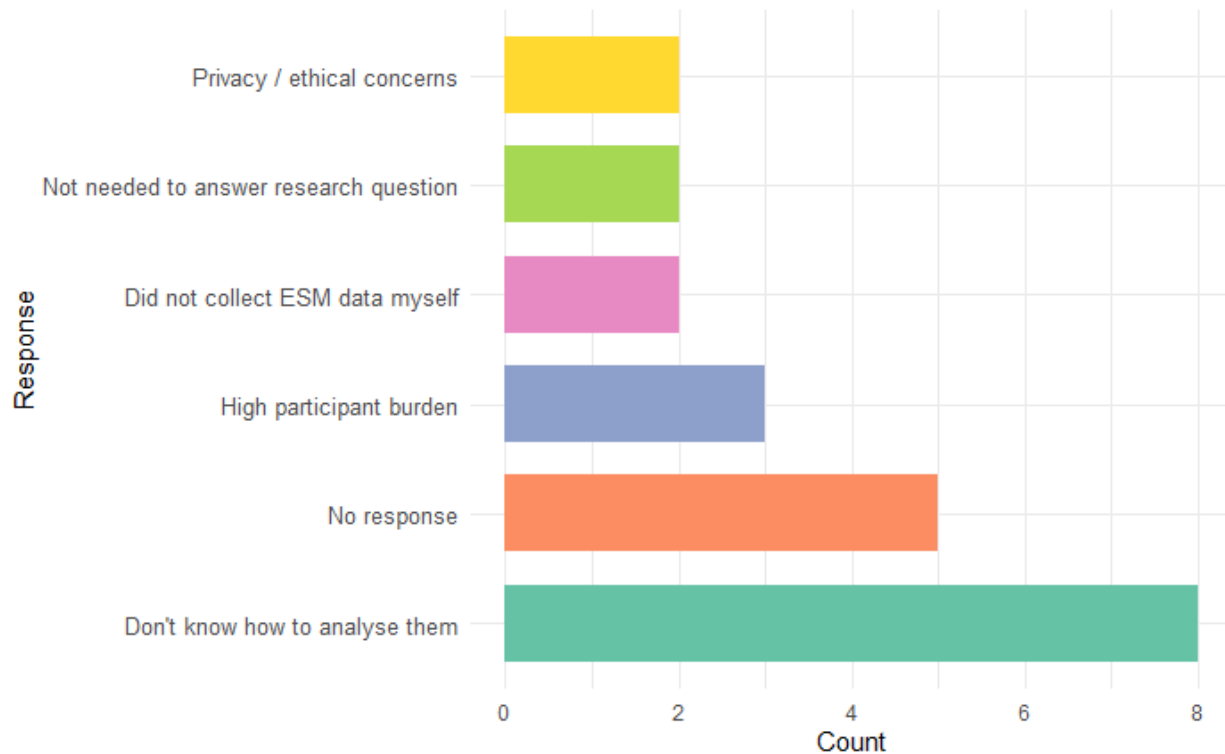


Figure 2. *Reasons for not using open-ended items. This histogram shows the responses to the follow-up question “What are your reasons for never using open-ended items during your ESM research?” to those respondents who reported never having used an open-ended item in their ESM research (N=22). “No response” means that the question was left unanswered. See Appendix 1 for the full survey.*

We also asked the follow-up question “What were your reasons for not analysing the collected data?” to those respondents who only collected open-ended item data and did not analyze or report it (N = 37; see Figure 3). The most common response was to simply leave this question unanswered, followed by the response that the project is still ongoing. Other reasons given included not needing them to answer the research question, and lack of time and resources. Some respondents also mentioned that the open-ended item data were used for giving feedback to participants based on their own data, but was not reported in the publications.

There was also a small group of respondents (N = 5) who analyzed the open-ended item data but did not report the analyses in the publication. The reasons for this were, for example, that the results of the analyses were in the end not needed to answer the research question, but were still useful for the participants.



Figure 3. *Reasons for not analyzing open-ended items. This histogram shows the responses to the follow-up question “What were your reasons for not analyzing the collected data?” to those respondents who only collected open-ended item data and did not analyze or report it (N = 37). “No response” means that the question was left unanswered. See Appendix 1 for the full survey.*

How open-ended items have been used and analyzed

The content analysis based on the three sources (i.e., survey with ESM researchers, ESM Item Repository search, literature review) showed that open-ended items were used for capturing similar kinds of information across studies. The content of the items can roughly be grouped into a few broader content categories. An overview of the items and their coding can be found in Table 1. The most commonly used open-ended items asked for event descriptions, specifically the most important, negative or positive events since the last assessment or during the day (e.g., “Since the last beep, the most important thing that happened for me was...”). In some cases, events were also assessed prospectively for the upcoming day (e.g., “Think about the most important event that will happen today. What is it?”). While in most studies any kind of event was considered, some studies prompted for specific kinds of events (e.g., events related to being part of a marginalised group). Open-ended items were also used to characterize the current situation of a participant (i.e., provide context for an experience). These items probed for current activities, location, and company (e.g., “What did I do just before the beep went off?”, “Where am I?”, “Who am I with?”). The characterisation of the current situation was

complemented with open-ended items asking for thoughts and feelings in the moment (e.g., “What was I thinking just before the beep went off”). These items were sometimes relative to specific experiences (e.g., the context when feeling high stress) or were used as follow-up items on event items.

Less frequent and more study-specific were items asking about feelings and their perceived causes (e.g., “What caused you the most anger today? Briefly describe”; Kashdan et al., 2016), coping strategies, meal descriptions, health-related matters (e.g., substance and medication use and complaints), goal setting, and cognitive failures. Finally, open-ended items also covered explanations for measurement-related matters such as response processes, missing beeps, or choosing not to share information with the researchers (e.g., “Reasons for not filling in” or “Reasons for turning off the watch”).

Regarding the structure of the items, both stand-alone and follow-up items (i.e., asked after a closed-ended item) were used in roughly equal proportions. Most of the items were in the category fully open-ended, followed by the category fully open/list (i.e., phrased to invite a list or categories). The data were mostly textual, with some studies supplementing textual data with other kinds of data: For example, Steinberg et al. (2023) and Cutler et al. (2018) collected voice recordings that were later transcribed to text, and König & Renner (2018) and König et al. (2021) included photos of meals in combination with open-ended text descriptions of the meals.

Studies varied in how the data from open-ended items were processed. The data were mostly coded manually, either using predefined categories (e.g., Ha et al., 2024; Hintzen et al., 2010; Levy et al., 2009), which appeared to be the slightly more common approach (18 studies), or using a data-driven method in which categories were derived from the data itself (14 studies; e.g., Farmer et al., 2017; McCombie et al., 2024; Shamasunder et al., 2022). In three cases, both methods were combined (Delespaul, 1995; deVries & Delespaul, 1989; Kashdan et al., 2016). Although not yet very common in the ESM literature, some of the papers used automated coding methods, such as Large Language Models (LLMs), Natural Language Processing (NLP) techniques, or Linguistic Inquiry and Word Count (LIWC). LLMs were used in one paper to classify participants as depressed vs. not depressed based on their language (Shin et al., 2024), while NLP techniques were used to quantify the occurrence of certain (types of) words (Hoemann et al., 2024; Ottenstein & Lischetzke, 2020; Stadel et al., 2025) and the semantic distances between closed- and open-ended responses (Stadel et al., 2025). LIWC was used to quantify the presence of words in predetermined categories, such as words related to negative or positive emotions, or to various life domains (e.g., family, work; Tov et al., 2013; Vakhrusheva et al., 2019).

There was also considerable variation in the statistical analyses that were applied. Some analyses were purely descriptive (e.g., length of transcribed voice recordings, Doherty et al., 2014; theme frequencies, Farmer et al., 2017), while others used inferential statistics to answer their research questions. For example, it was common for studies to use coded categories as variables to predict other outcomes (e.g., how stressful experiences varied by type and threat appraisals, Pow et al., 2016; which anger triggers predicted anger-related behaviors, Kashdan et al., 2016). Studies with clinical groups often used codes to structure comparisons with non-clinical controls (e.g., what kinds of experiences schizophrenic vs. non-schizophrenic individuals

are having, Delespaul et al., 2002; what kinds of activities depressed vs. non-depressed individuals are engaging in, Barge-Schaapveld & Nichols, 2002). Only a few papers engaged with fully qualitative, interpretative analyses (e.g., where example quotes are presented and dissected). For example, Flynn et al. (2024) and Sullivan et al. (2024) presented the themes that emerged from positive and negative identity-related experiences of sexual and gender minority adolescents.

Lastly, the responses to open-ended items were presented with a broad range of visualizations, including word clouds, pie charts, or Venn diagrams. For example, Delespaul (1995) used pie charts to illustrate the time use patterns of activities of participants, and Hoemann et al. (2024) used word clouds to illustrate how often emotion words were used in open-ended items. Such visualisations are common with open-ended items, often as feedback to the participants (see, e.g., von Klipstein et al., 2023; Bringmann et al., 2021; Stadel et al., 2024 that used open-ended items for feedback only, but did not analyze or summarize them further).

References

- Barge-Schaapveld, D. Q., & Nicolson, N. A. (2002). Effects of antidepressant treatment on the quality of daily life: an experience sampling study. *The Journal of Clinical Psychiatry*, 63(6), 477-485. <https://doi.org/10.4088/jcp.v63n0603>
- 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*, 30(11), 3179–3188. <https://doi.org/10.1007/s11136-020-02701-4>.
- Cutler, S., Doherty, S., & Carmichael, B. (2018). The experience sampling method: examining its use and potential in tourist experience research. *Current Issues in Tourism*, 21(9), 1052-1074. <https://doi.org/10.1080/13683500.2015.1131670>.
- Delespaul, P. (1995). *Assessing schizophrenia in daily life: The experience sampling method* [PhD thesis]. University of Limburg.
- Delespaul, P., deVries, M., & van Os, J. (2002). Determinants of occurrence and recovery from hallucinations in daily life. *Social Psychiatry and Psychiatric Epidemiology*, 37(3), 97-104. <https://doi.org/10.1007/s001270200000>
- deVries, M. W., & Delespaul, P. A. (1989). Time, context, and subjective experiences in schizophrenia. *Schizophrenia Bulletin*, 15(2), 233–244. <https://doi.org/10.1093/schbul/15.2.233>.
- Doherty, S. T., Lemieux, C. J., & Canally, C. (2014). Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. *Social Science & Medicine*, 106, 83-92. <https://doi.org/10.1016/j.socscimed.2014.01.048>.
- Farmer, S. , Mindry, D. , Comulada, S. W. & Swendeman, D. (2017). Mobile Phone Ecological Momentary Assessment of Daily Stressors Among People Living With HIV: Elucidating Factors Underlying Health-Related Challenges in Daily Routines. *Journal of the Association of Nurses in AIDS Care*, 28(5), 737-751. <https://doi.org/10.1016/j.jana.2017.04.001>.

Flynn, S. S., Touhey, S., Sullivan, T. R., & Mereish, E. H. (2024). Queer and transgender joy: A daily diary qualitative study of positive identity factors among sexual and gender minority adolescents. *Psychology of Sexual Orientation and Gender Diversity*. <https://doi.org/10.1037/sgd0000733>.

Ha, T., Iida, M., Quiroz, S. I., Maras, O., & Savord, A. (2024). Feeling loved in mixed-gender adolescent romantic relationships in the face of daily stress and support: A dyadic diary study. *Developmental Science*, 27(6), e13511. <https://doi.org/10.1111/desc.13511>.

Hintzen, A., Delespaul, P., van Os, J., & Myin-Germeys, I. (2010). Social needs in daily life in adults with Pervasive Developmental Disorders. *Psychiatry Research*, 179(1), 75-80. <https://doi.org/10.1016/j.psychres.2010.06.014>.

Hoemann, K., Warfel, E., Mills, C., Allen, L., Kuppens, P., & Wormwood, J. B. (2024). Using freely generated labels instead of rating scales to assess emotion in everyday life. *Assessment*, 32(6), 859–877. <https://doi.org/10.1177/10731911241283623>.

Kashdan, T. B., Goodman, F. R., Mallard, T. T., & DeWall, C. N. (2016). What triggers anger in everyday life? links to the intensity, control, and regulation of these emotions, and personality traits. *Journal of Personality*, 84(6), 737-749. <https://doi.org/10.1111/jopy.12214>.

König, L. M., Koller, J. E., Villinger, K., Wahl, D. R., Ziesemer, K., Schupp, H. T., & Renner, B. (2021). Investigating the relationship between perceived meal colour variety and food intake across meal types in a smartphone-based ecological momentary assessment. *Nutrients*, 13(3), 755. <https://doi.org/10.3390/nu13030755>.

König, L. M., & Renner, B. (2018). Colourful= healthy? Exploring meal colour variety and its relation to food consumption. *Food Quality and Preference*, 64, 66-71. <https://doi.org/10.1016/j.foodqual.2017.10.011>.

Levy, A., Nicholls, A., Marchant, D., & Polman, R. (2009). Organisational stressors, coping, and coping effectiveness: A longitudinal study with an elite coach. *International Journal of Sports Science & Coaching*, 4(1), 31-45. <https://doi.org/10.1260/1747-9541.4.1.31>.

McCombie, C., Ouzzane, H., Schmidt, U., & Lawrence, V. (2024). 'Physically it was fine, I'd eat what normal people do. but it's never like this in my head': A qualitative diary study of daily experiences of life in recovery from an eating disorder. *European Eating Disorders Review*, 32(1), 46-55. <https://doi.org/10.1002/erv.3018>.

Ottenstein, C., & Lischetzke, T. (2020). Development of a Novel Method of Emotion Differentiation That Uses Open-Ended Descriptions of Momentary Affective States. *Assessment*, 27(8), 1928–1945. <https://doi.org/10.1177/1073191119839138>.

Pow, J., Lee-Baggley, D., & DeLongis, A. (2016). Threats to communion and agency mediate associations between stressor type and daily coping. *Anxiety, Stress & Coping: An International Journal*, 29(6), 660-672. <https://doi.org/10.1080/10615806.2015.1126258>.

Shamasunder, B., Chan, M., Navarro, S., Eckel, S., & Johnston, J. E. (2022). Mobile daily diaries to characterize stressors and acute health symptoms in an environmental justice neighborhood. *Health & Place*, 76, 102849. <https://doi.org/10.1016/j.healthplace.2022.102849>.

Shin, D., Kim, H., Lee, S., Cho, Y., & Jung, W. (2024). Using Large Language Models to Detect Depression From User-Generated Diary Text Data as a Novel Approach in Digital Mental Health Screening: Instrument Validation Study. *Journal of Medical Internet Research*, 26, e54617. <https://doi.org/10.2196/54617>.

Stadel, M., Langener, A. M., Hoemann, K., & Bringmann, L. F. (2025). Assessing daily life activities with experience sampling methodology (ESM): Scoring predefined categories or qualitative analysis of open-ended responses? *Methods in Psychology*, 12, 100177. <https://doi.org/10.1016/j.metip.2025.100177>.

Stadel, M., Stulp, G., Langener, A. M., Elmer, T., van Duijn, M. A. J., & Bringmann, L. F. (2024). Feedback About a Person's Social Context—Personal Networks and Daily Social Interactions. *Administration and Policy in Mental Health and Mental Health Services Research*, 51(4), 476–489. <https://doi.org/10.1007/s10488-023-01293-8>.

Steinberg, N., Moshe-Cohen, R., Sheena, L. et al. (2023). Capturing a mentalized moment: A pilot study of the psychometric properties of a novel assessment method of mentalizing in daily life. *Current Psychology*. <https://doi.org/10.1007/s12144-023-04963-w>.

Sullivan, T. R., Flynn, S. S., Touhey, S., & Mereish, E. H. (2024). Through their eyes: A qualitative, daily diary exploration of oppression-based stress experiences among sexual and gender minority adolescents. *Psychology of Sexual Orientation and Gender Diversity*. <https://doi.org/10.1037/sgd0000779>.

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment*, 25(4), 1069-1078. <https://doi.org/10.1037/a0033007>.

Vakhrusheva, J., Khan, S., Chang, R., Hansen, M., Ayanruoh, L., Gross, J., & Kimhy, D. (2019). Lexical analysis of emotional responses to “real-world” experiences in individuals with schizophrenia. *Schizophrenia Research*, 216, 272–278. <https://doi.org/10.1016/j.schres.2019.11.045>.

von Klipstein, L., Servaas, M. N., Schoevers, R. A., van der Veen, D. C., & Riese, H. (2023). Integrating personalized experience sampling in psychotherapy: A case illustration of the Therap-i module. *Heliyon*, 9(3), e14507. <https://doi.org/10.1016/j.heliyon.2023.e14507>.