To cite this article:

Verhees, M. W. F. T., Frérart, L., Horn, A. B., Inverardi, M., Kalkan-Cengiz, R. B., Righetti, F., Schoebi, D., Sels, L., Versyp, O., Kuppens, P., & Ceulemans, E. (in press). Dyadic compliance and report congruence in romantic couples' experience sampling data. *Psychological Assessment*.

Dyadic Compliance and Report Congruence in Romantic Couples' Experience Sampling Data

Martine W. F. T. Verhees[1*], Liesse Frérart[1], Andrea B. Horn[2,3], Michela Inverardi[1], Rana B. Kalkan-Cengiz[1], Francesca Righetti[4], Dominik Schoebi[5], Laura Sels[6], Otto Versyp[1], Peter Kuppens[1] & Eva Ceulemans[1]

[1] Research group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium

[2]Department of Psychology, Gerontopsychology and Gerontology, University of Zurich, Zurich, Switzerland

[3]Center for Gerontology, Healthy Longevity Center, University of Zurich

[4]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[5]Department of Psychology, University of Fribourg, Fribourg, Switzerland

[6]Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium

Funding statement: This work was supported by de Vlaamse Interuniversitaire Raad (VLIR iBOF/21/090), by the Research Foundations FWO and F.R.S.-FNRS under the Excellence of Science (EOS) program (Grant EOS 40007528/G0I2422N), the Netherlands Organization for Scientific Research (451-12-024), the Swiss National Science Foundation (PMPFP1_164470).

Declaration of conflicting interest: The author(s) declare no conflicts of interest.

Use of previously published data: Some of the datasets analyzed in the current study have been used in previous publications. However, in those publications, dyadic compliance and congruence were not the variables of interest, and the current study represents a novel analysis describing these variables and examining their relations with methodological predictors.

Corresponding author: Martine W. F. T. Verhees, Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Tiensestraat 102, 3000, Leuven, Belgium, martine.verhees@kuleuven.be

Abstract

Dyadic experience sampling method (ESM) studies offer unique insights into daily life interpersonal processes, but little is known about data quantity and quality in such designs. In this preregistered analysis of existing data, we examined overall levels and methodological predictors of dyadic compliance, both partners responding to ESM surveys, and report congruence, partners' alignment in responses regarding shared events. We analyzed eight studies comprising 1153 romantic couples and 41588 prompts. Most couples were different-gender; information on race, ethnicity, and socioeconomic status was not consistently available across datasets. Overall, dyadic compliance and congruence were high. Compliance was lower in studies with more surveys and shorter allowed response delays, and higher later in the day. Congruence was higher as the study progressed, and lower for retrospective or subjective events. These findings highlight the potential for high-quality dyadic ESM data. Yet, they also underscore that different factors can affect data quantity and quality, emphasizing the importance of careful attention to study design.

Keywords: dyadic data, experience sampling method, romantic couples, compliance, congruence

Public significance statement: This study shows that couples generally completed daily life smartphone surveys at high rates and provided responses aligned with each other regarding shared events, but factors like survey timing and content of the questions can influence response rates and agreement. These findings are important for researchers and clinicians who want to use

dyadic daily life data to understand how partners influence and respond to each other in real-world settings.

Dyadic Compliance and Report Congruence in Romantic Couples' Experience Sampling Data

Experience sampling methods (ESM) allow researchers to track people's emotions, thoughts, and contexts in the course of their daily lives, providing a wealth of ecologically valid and situationally sensitive data. When applied to dyads, ESM captures both dyad partners' experiences, allowing these to be mapped onto one another. This facilitates a truly interpersonal measurement of daily life, addressing calls to adopt an interpersonal perspective to increase our understanding of psychological (mal)adjustment (e.g., Bolis et al., 2023; Hopwood, 2018). Besides the opportunities dyadic ESM studies provide, they also present some challenges. Specifically, to map experiences of dyadic partners onto one another, they should report at the same timepoints and on the same events (Ji et al., 2025; Wrzus & Neubauer, 2023). This highlights the importance of (1) dyadic compliance, i.e., the proportion of ESM surveys, or prompts, answered by both partners at the same time, and (2) report congruence, or the alignment of partners' responses regarding shared events (note, however, that congruence may not always be expected or necessary, and, in some cases, incongruence itself may be of interest, see below). Despite their relevance, these aspects have not been studied, leaving a gap in our understanding of (a) their levels, and (b) the factors that may predict them.

To address this, we sought to describe overall levels of dyadic compliance (aim 1a) and report congruence for relationship-related events (aim 2a) across a set of ESM studies in which both partners of romantic couples participated. Moreover, to inform future studies' design choices, we aimed to explore factors that may predict dyadic compliance (aim 1b) and congruence (aim 2b), focusing on methodological factors at the study, prompt, and, for congruence specifically, item level. Note that there were no item-level predictors for compliance,

as compliance itself is situated at the prompt-level and thus does not vary in relation to specific items within prompts.

Dyadic compliance

ESM involves repeated assessment of participants in their daily lives, typically asking them to respond to multiple prompts per day over the course of several days. Participants usually have a limited window of time to respond to each prompt, known as the allowed response delay, after which the survey expires and is recorded as missing. Response compliance, the percentage of scheduled prompts that were completed by participants, is of relevance as low compliance can negatively impact both data quantity and quality (Eisele et al., 2022). Noncompliance may stem from not being able to respond at certain times (e.g., when driving a car), or a lack of motivation and engagement to respond to prompts, as ESM studies can be burdensome. A recent meta-analysis indicated that individual ESM compliance across different fields of research was on average 79% (Wrzus & Neubauer, 2023). Turning to dyadic compliance, this can be expected to be lower than individual compliance as a timepoint is missing for the dyad if at minimum one of the partners does not answer the prompt. However, to date, no research explicitly examined dyadic ESM compliance, nor which methodological factors may predict it.

Previous research did investigate study-level and prompt-level predictors of *individual* compliance. Regarding study-level predictors, findings have been mixed. For example, a study examining compliance across 488 ambulatory assessment studies found that total number of prompts and number of study days were negatively related to compliance (Ottenstein & Werner, 2022). However, another recent meta-analysis found little evidence for effects of study design-related factors (including number of prompts or study days), except for an effect of prompts per

day and incentives (Wrzus & Neubauer, 2023). Specifically, studies with fewer prompts per day showed borderline higher compliance rates. Additionally, studies with monetary incentives had higher compliance rates than those without.[1] Another study did not find any effects of sampling participants at fixed versus random times on individual compliance (Reiter et al., 2025). Thus, although findings are mixed and studies do not necessarily identify the same significant predictors, when significant effects are found, they suggest that study designs asking more from participants tend to be associated with lower compliance. Moreover, a study-level factor that to our knowledge has not been systematically investigated but may affect compliance is the allowed response delay, i.e., how quickly a survey request expires after it was sent to participants (Van Berkel et al., 2018). Longer allowed response delays give participants more time to respond, reducing the likelihood of missed prompts.

Concerning prompt-level predictors of compliance, it has been found that individual compliance can vary throughout a study. Specifically, compliance tends to decrease over the course of the study (e.g., Ono et al., 2019; Rintala et al., 2019). In addition, compliance may vary dependent on time and type of day that the prompt was sent, although again, findings have not been consistent. Specifically, individual compliance has been found to be lower early in the day (Rintala et al., 2019), as well as later in the day (Sun et al., 2021), although this seems to depend on the start and end time of the sampling schedule (Silvia et al., 2013). Overall, individuals seem most compliant around the middle of the day (Rintala et al., 2019; Silvia et al., 2013; Sun et al., 2021). Moreover, compliance was found to be lower on weekend versus weekdays in some studies (Rintala et al., 2019), but others reported no significant differences (e.g., Sun et al., 2021; Van Berkel et al., 2020).

[1] In the current study we could not examine the effect of monetary vs non-monetary incentives, as the included studies did not vary in this regard (i.e., they all provided monetary incentives).

Dyadic report congruence

Besides dyadic compliance, report congruence among partners' reports on relationship-related events is important as it allows to select data based on certain events or examine the effects of the occurrence of such events on partners. For example, one may want to assess whether emotions covary between partners, but only in moments when they have recently been in contact, or compare covariation of partner emotions between contact and no-contact moments. When partners' reports of such events (in this example, having had contact with the partner) are incongruent, this forces researchers to either exclude the occasion from analyses, or choose to 'believe' the report of one of the partners and not the other (note, however, that sometimes incongruence itself may actually be of interest; e.g., De Los Reyes et al., 2019; Ehrlich et al., 2019). Additionally, congruence in reporting about events is important when such events are used for branching (i.e., asking different follow-up questions based on an event report). In such cases, incongruence between partners will lead to different follow-up questions for both partners, resulting in missing data for those follow-up questions at the dyad level (Ji et al., 2025). Discrepancies between partner reports may stem from responding without sufficient regard to the questions (careless or insufficient effort responding, potentially due to a reduced ability or motivation to provide attentive responses to prompts) by one or both partners, differences in response times, forgetfulness, or differences in interpretation of the question.

To our knowledge, no previous research specifically examined report congruence in dyadic ESM or its methodological predictors. However, the effects of design choices (e.g., questionnaire length, prompts per day) on (in)attentive responding in individual ESM, which may be of relevance for congruence, have been investigated by a few studies (e.g., Eisele et al.,

2022; Hasselhorn et al., 2021; Reiter et al., 2025). These studies provided mixed results, potentially attributable to different indicators of (in)attentive responding being examined. Again, however, if significant results were found in studies examining study-level predictors, they point to reduced attentiveness with more intensive study designs (e.g., longer questionnaires, higher number of prompts per day; Eisele et al., 2022; Hasselhorn et al., 2021; Ulitzsch, Viechtbauer, et al., 2024). Research additionally suggests that participants' attentiveness can vary within a study, as participants seem to get more inattentive (Ulitzsch, Nestler, et al., 2024), more inaccurate (Verbeij et al., 2021), and faster (Dejonckheere et al., 2024; Eisele et al., 2023) over the course of an ESM study. These findings align with a so-called fatigue effect: participants may become tired of repeatedly answering the same questions, reducing their accuracy as the study progresses (Verbeij et al., 2021). Moreover, participants seem to get more inattentive (Ulitzsch, Viechtbauer, et al., 2024) and faster (Dejonckheere et al., 2024) over the course of an individual study day. Additionally, one could predict that longer inter-response times between partners may negatively impact congruence[2] as chances increase that the actual occurrence of relationship-related events diverges when the timeframes partners report on differ. For example, if partners are asked whether they have had contact since the last prompt, their responses may diverge if one partner responds immediately, before any interaction occurred, while the other responds later, after they did have contact. Finally, item-related factors such as the timeframe a question captures or the type of event (and how much interpretation goes into this) could be relevant predictors. Specifically, due to recall biases, items asking individuals to remember what has happened in a certain time interval (e.g., in the last hour or since the last prompt) may be expected to relate to increased incongruence compared to items assessing events happening at

[2] Note that we do not explore inter-response time between partners as a predictor for compliance as this information is not available for non-compliant prompts

the current moment for which no recall is needed. Moreover, items asking about more objective events such as contact with the partner may be expected to lead to higher congruence between partners than more subjective events such as whether or not a conflict with the partner occurred.

The current study

This study aimed to examine dyadic compliance and report congruence in romantic couples. Across eight different ESM studies in romantic couples, we (1a) examined the overall level of dyadic compliance; and (1b) explored whether study-level predictors (i.e., total number of ESM days, total number of prompts, prompts per day, questionnaire length, and allowed response delay) and prompt-level predictors (study day, time of day, type of day) predict dyadic compliance. Regarding study-level predictors of dyadic compliance, we expected that compliance may be lower in studies with a higher number of study days, prompts, prompts per day and questionnaire length (Ottenstein & Werner, 2022; Wrzus & Neubauer, 2023), and shorter allowed response delays. Regarding prompt-level predictors we expected that compliance may decrease over the course of the study (Ono et al., 2019; Rintala et al., 2019), and may be highest for prompts around the middle of the day (Rintala et al., 2019; Sun et al., 2021) and during the week (Rintala et al., 2019).

Concerning congruence, we (2a) examined overall report congruence between partners concerning the occurrence of relationship-related events; and (2b) explored whether study-level predictors (i.e., total number of days, total number of prompts, prompts per day, questionnaire length, and allowed response delay), prompt-level predictors (study day, time of day, type of day, inter-response time between partners), and item-level predictors (i.e., timeframe of item, subjectiveness of event) predicted report congruence. Concerning study-level predictors, we

expected that we may find lower congruence for studies with a higher number of study days, prompts, prompts per day and questionnaire length (Eisele et al., 2022; Hasselhorn et al., 2021; Ulitzsch, Viechtbauer, et al., 2024). Regarding the effect of the allowed response delay, we anticipated that longer allowed delays might relate to reduced congruence, as they could cause partners to report on different timeframes. Moreover, concerning prompt-level predictors,we expected that congruence may decrease over the course of the study and over the course of an individual day (e.g., Ulitzsch, Nestler, et al., 2024; Ulitzsch, Viechtbauer, et al., 2024; Verbeij et al., 2021), and with longer inter-response times between partners. We had no specific predictions concerning effects of day of the week the prompt was sent. Finally, at the item-level, we expected lower congruence when items assess the occurrence of events retrospectively (e.g., in the past hour, since the last prompt) versus currently (e.g., at this moment, right now) and when items assess more subjective (e.g., conflict) versus more objective (e.g., contact) events.

Upon reviewers' suggestions, we performed additional exploratory analyses. Specifically, we explored whether item-level predictor event valence related to report congruence. Moreover, we explored whether dyad-level predictors age of the couple and cohabitation status related to dyadic compliance and congruence. Concerning age, some previous studies on individual compliance found lower compliance rates in younger (adult) participants (Ono et al., 2019; Rintala et al., 2019); although meta-analytical evidence suggested no effects of age on compliance (Wrzus & Neubauer, 2023). Cohabitation status may serve as a proxy for time spent together and could be hypothesized to relate to higher compliance if partners would alert each other to the occurrence of prompts.

Methods

Transparency and Openness

The hypotheses, methods, and analysis plan for this study were preregistered at https://osf.io/9gmzu, after data collection for the individual studies was completed but before we had access to all datasets for this project and before any analyses were conducted across datasets. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The data used in this study were obtained from multiple researchers and institutions, and are subject to varying data sharing restrictions; therefore, the raw data for this study are not publicly available. The analysis code is available at https://osf.io/ue4bh. Research materials (items about relationship-related events) are described under 'Measures' below. Data were analyzed using the lme4 package, version 1.1.-27.1 (Bates et al., 2015) in R, version 4.1.2 (R Core Team, 2021). Generative AI ChatGPT (OpenAI, 2025) was used to assist with language and phrasing. After using this tool, the authors reviewed and edited the content as needed.

Sample and Procedure

Analyses were conducted using a dataset comprising data from eight studies employing a signal or interval-contingent design (i.e., prompts were sent at (quasi-)random times or at fixed intervals). Datasets were obtained from different sources, i.e., via the RE-Connect consortium and the love consortium[3]. The former is a research consortium in which some of the authors are involved, while the latter is a global network that created a research dataverse from which we identified and requested access to relevant datasets for this project. Data could be included if

[3] Note that the current study was not a systematic review of all available empirical research, but rather multiple datasets that were available to the researchers were analyzed.

they contained ESM data (minimum of two measurements per day over the course of several days) and both partners participated in the study and received surveys at the same time.

The final dataset was based on the availability of existing data that met our inclusion criteria, resulting in a total of 1153 couples and 41588 prompts. The majority of couples across studies were different-gender couples (99%), cohabiting (82%) and had children (62%). Relationship duration was on average 122 months (SD = 125). The experience sampling procedure differed between studies in terms of number of ESM days, number of prompts (overall and per day), questionnaire length, allowed response delay and time of sampling. Information on study design factors per study is presented in Table 1. All study procedures were approved by local ethical committees. Participants in all studies provided written informed consent.

Table 1

Study Design Characteristics and Overall Compliance and Congruence per Study

Study	N (couples)	Number of days	Total number of prompts	Prompts per day	Questionnaire length (items)	Allowed response delay (minutes)	Sampling time(s)	Overall compliance	Overall congruence	Reference
1	50	7	70	10.00	25	1.5	10 am – 10 pm	87%[ab]	87%[a]	(Sels et al., 2017)
2	94	7	58	8.29	24	1.5	10 am – 10 pm	88%[a]	91%[b]	(Sels et al., 2020)
3	50	10	78	7.80	18.5	5	10 am – 10 pm	57%[c]	94%[c]	n/a
4	106	14	42	3.00	15.5	8	10 am – 10 pm	68%[d]	97%[d]	n/a
5	119	8	48	6.00	25	60	9am and 9pm	79%[e]	89%[a]	(Righetti et al., 2016)

6	116	21	63	3.00	52	120	8am – 8pm	82%[be]	86%[a]	n/a
7	54	14	28	2.00	29	540	5am and 6pm	95%[f]	93%[ce]	(Horn et al., 2021)
8	564	7	21	3.00	56	120	Noon – 9pm	87%[a]	91%[be]	(Schoebi et al., 2010)

Note. Prompts per day was calculated as total number of prompts divided by total number of days, as some studies' sampling frequency varied across the study (e.g., more prompts during weekend days). Overall compliance and overall congruence were calculated per study. Superscripts in the Overall compliance and Overall congruence columns indicate pairwise differences between studies as tested in a multilevel model with Study as fixed effect (with different studies as reference categories): studies sharing at least one superscript do not differ significantly, whereas studies without shared superscripts differ significantly ($p < .05$).

Measures

Outcomes

Dyadic compliance was operationalized as a binary variable indicating whether both partners answered the prompt (as reflected by a response of both partners to the last (non-conditional) item of the ESM questionnaire; coded as 1) or not (coded as 0).

Congruence on relationship-related events was operationalized as a binary variable indicating whether both partners agreed on the occurrence of the event (both partners reported the event did happen, or both partners reported the event did not happen; coded as 1) or not (only one of the partners reported the event happened; coded as 0). Events were included if they involved a shared experience with the partner (e.g., contact, conflict, pleasant moment with partner), with both partners responding to the same item, and their answers were in a 'did/did not occur' format or could be recoded to this format (e.g., for some contact items, participants indicated whether they had seen each other, called, texted, or not had contact; these responses were recoded into a binary contact/no-contact variable). A total of 20 different events were included, with each study containing between one and five events (see Table 2 for the items and their coding by study).

Table 2

Items on Relationship-Related Events per Study and Their Coding

Study	Item	Timeframe	Event subjectiveness	Valence	Frequency of occurrence
1	Are you with your partner at the moment?	0	0	0	39%
	Have you been in contact with your partner since the last beep?	1	0	0	73%
	Did something positive happen involving your partner since the last beep?	1	1	1	20%
	Did something negative happen involving your partner since the last beep?	1	1	-1	7%
	Have you talked with your partner about a problem that one or both of you encountered since the last beep?	1	1	0	20%
2	Are you with your partner at the moment?	0	0	0	60%
	Have you been in contact with your partner since the last beep?	1	0	0	79%
	Have you had a conflict with your partner since the last beep?	1	1	-1	5%
	Have you had a nice moment with your partner since the last beep?	1	1	1	51%
3	Are you with your partner at the moment?	0	0	0	52%
	Have you been in contact with your partner since the last beep?	1	0	0	88%
4	Are you with your partner at the moment?	0	0	0	43%
	Have you been in contact with your partner since the last beep?	1	0	0	97%
	Have you been sexually active with your partner since the last beep?	1	0	0	16%
5	In the past hour, have you encountered a situation of divergence of interests?	1	1	0	10%

6	Have you been in contact with your partner since this morning?	1	0	0	95%
	I have talked with my partner about what happened today.	1	0	0	83%
7	I have been in contact with my partner today.	1	0	0	94%
	I have talked with my partner about what happened today.	1	0	0	94%
8	Where there discussions, tensions or conflicts with your partner in the last 4 hrs.?	1	1	-1	10%

Note. Timeframe is coded as 0 = current, 1 = retrospective; Event subjectiveness is coded as 0 = more objective, 1 = more subjective.

Frequency of occurrence reflects how often the event was reported to have occurred across all participant reports.

Predictors

Study-level predictors were number of days, number of prompts, prompts per day, questionnaire length and allowed response delay. Number of days and prompts reflect the total number of ESM days and the total number of prompts participants received across the sampling period, respectively. Prompts per day was operationalized as the number of prompts divided by the number of days. Questionnaire length reflects the number of items asked at each prompt. In case conditional questions were used and therefore the number of items asked differed per prompt we took the average of the minimum and maximum number of items asked. Upon reviewer suggestion, we also explored the effect of average number of actually answered items across participants per study as indicator of questionnaire length.Finally, allowed response delay was the time in hours[4] after which the ESM survey expired after it was sent to participants.

Prompt-level predictors were study day, type of day, time of day, and inter-response time between partners. Study day reflects the moment a prompt was sent in continuous study days with zero reflecting the first day of sampling. Type of day reflects whether the prompt was sent during a week or weekend day (0 = week, i.e., Monday through Friday; 1 = weekend, i.e., Saturday and Sunday). Time of day indicates the time the prompt was sent within a day in hours, with zero indicating 12 noon (meaning that times in-between midnight and noon receive a negative value). [5] Inter-response time between partners reflects the absolute difference in minutes between the times a prompt was answered by each of the partners.

[4] We preregistered that allowed response delay would reflect the time in minutes but the model was nearly unidentifiable so we rescaled the predictor to hours.

[5] We deviated from our preregistration here in two ways: we preregistered that we would have time of day coded in minutes rather than hours, and that we would examine both the linear and curvilinear effects of time of day by including a linear and quadratic effect of time. However, due to scaling and convergence issues when running the models, we rescaled the variable to hours and only included the linear effect of time.

Item-level predictors were timeframe and event subjectiveness. Timeframe reflects whether the item asks about a current or retrospective event (0 = current; 1 = retrospective, e.g., since the last prompt, in the past hour). Event subjectiveness indicates whether the item asked about a more objective or subjective event (0 = more objective, e.g., contact; 1 = more subjective, e.g., conflict). Event valence was coded from the question wording: positive (1) if the item explicitly contained positive terms (e.g., a positive event with partner), negative (-1) if it contained a negative term (e.g., conflict), and neutral (0) if it concerned an item without emotional language (e.g., contact). Event subjectiveness and valence were coded by two researchers independently, and in case their coding diverged (which happened three times) discussed to consensus (see Table 2 for the coding of the events).

In exploratory analyses we explored the effect of dyad-level predictors couple age and cohabitation status on dyadic compliance and congruence. Couple age was set to the two partners' average age in years. Cohabitation status reflected whether couples cohabited (0 = no, 1= yes).

Analyses

Missing data

Prompts were excluded when partners did not receive them at approximately the same time (e.g., due to technical issues). We used a cut-off of 90 seconds[6], the shortest allowed response delay across studies. Some dyads did not receive any prompts at the same time and were thus completely excluded (n = 17). In Study 5, data on the time of day when the prompt was sent was missing in case both partners missed the prompt. To address this, the missing time was imputed based on the sequence of prompts within the day (e.g., first, second, or x-th

[6] This specific cut-off was not preregistered.

prompt), using the average time for that specific position in the sequence across participants and days. Study 7 had no information on inter-response time between partners and was thus left out of the analyses with inter-response time as predictor. Additionally, Study 4 had no information about cohabitation status and was thus not included in the exploratory analyses with cohabitation status as predictor. Finally, for congruence, responses to items for which individuals were not congruent with themselves were deleted (e.g., in items allowing multiple responses, sometimes participants selected both 'having had contact' and 'having had no contact', n = 70 responses). This was not preregistered as these instances were only discovered during data analysis.

Main analyses

We fitted multilevel logistic regression models using the lme4 package in R (Bates et al., 2015). Effects for which p-values <.05 were considered statistically significant. For dyadic compliance as criterion variable, we ran three-level models (with prompts nested within dyads who are nested within studies). We first fitted an empty intercept-only model to describe the overall dyadic compliance across datasets (aim 1a). Next, we ran separate regression models for each predictor to examine the effects of the study- and prompt-level predictors (aim 2a). All models included random intercepts for study and dyads within study, and models with prompt-level predictors additionally include random slopes for study and dyads within study. As an example, we present the equation for compliance as criterion and study day as prompt-level predictor:

Level 1:

$$\text{Dyadic compliance}_{tij} = b_{0ij} + b_{1ij}\,\text{Study day}_{tij} + e_{tij}$$

Level 2:

$$b_{0ij} = b_{00j} + u_{0ij}$$

$$b_{1ij} = b_{10j} + u_{1ij}$$

Level 3:

$$b_{00j} = B_{000} + v_{00j}$$

$$b_{10j} = B_{100} + v_{10j}$$

t = timepoint, i = dyad, j = study

Next, we fitted the models with report congruence as criterion variable. We ran three-level models with item-prompt combinations nested within dyads who are nested within studies. Again, we first fitted an empty model to describe overall report congruence across all events and datasets (aim 2a). Then, we ran separate regression models for each predictor to examine whether congruence can be predicted by the different study-, prompt- and item-level predictors (aim 2b). All models included random intercepts for study and dyads within study. We preregistered that all models with prompt- and item-level predictors would additionally include random slopes for study and dyads within study. However, we deviated from this in three cases due to singular fit issues: for the model with inter-response time between partners as predictor, the random slope for dyads was removed, and for the models with timeframe and event subjectiveness as predictors the random slope for study was removed.

Sensitivity and exploratory analyses

First, we ran two multiple regression models (one with compliance and one with congruence as criterion) including simultaneously all predictors that were significant in the single-predictor regression models, to explore their unique effects. Second, for congruence specifically, we did not account for the nesting of items within prompts in our main analyses, that is, in some studies, multiple item-specific congruence values were obtained within the same prompt. For example, in Study 3, participants were asked at each prompt whether they (1) were together with their partner at the moment, and (2) had been in contact with their partner since the last beep. This means that for each completed prompt, two values for congruence were obtained, one for item 1 and one for item 2. Therefore, we ran sensitivity analyses to examine whether findings converged when we only included one item per prompt, such that items (rather than item-prompt combinations as in the main analyses) were nested within dyads within studies. To this aim, we created a series of 500 datasets in which we randomly selected one item per prompt and reran the analyses across these datasets. Third, we explored the effects of dyad-level predictors couple age and cohabitation status. To this aim, for each outcome (dyadic compliance and report congruence), we ran separate regression models with couple age or cohabitation status as the predictor. In these models, random intercepts were included for study and dyads within study as in our main analyses. Random slopes for study were additionally included for the models with cohabitation status as predictor, but not for those with couple age predictor as in these models including this random slope led to singular fit issues. Fourth, we explored the effect of item-level predictor event valence on congruence. In this model, random intercepts were included for study and dyads within study and a random slope for dyads within study (we removed a random slope for study due to singular fit issues).

Results

Dyadic Compliance

Overall dyadic compliance across studies amounted to 87.86%. Using the latent variable approach (Goldstein et al., 2002), ICCs showed that for compliance, 27% of variance was situated at the study level, 16% at the dyad level, and 57% at the prompt level. The fixed effects of study-and prompt-level predictors on dyadic compliance are presented in Table 3. The random effect variances/covariances are reported in supplementary Table 1. Study-level predictor number of prompts related negatively to compliance, indicating that studies with more prompts in total had lower compliance rates. Additionally, a positive effect of allowed response delay indicated that studies in which surveys expired faster after being sent to participants obtained lower compliance rates. No other predictors at the study level were significantly associated with compliance in our main analyses. At the prompt-level, time of day positively predicted compliance indicating that dyads become more compliant over the course of a day. Prompt-level predictors study day and type of day (week vs weekend) were not significantly related to compliance. When we included all significant predictors in one model, the effect of study-level predictor number of prompts became non-significant ($b = -0.02$, $p = .109$), whereas the effects of allowed response delay ($b = 0.24$, $p = .036$) and time of day ($b = 0.04$, $p < .001$) remained significant. An exploratory analysis of questionnaire length operationalized as average number of actually answered items across participants per study indicated that a higher number of items was associated with higher compliance ($b = 0.05$, $p = .039$). Note that this effect became non-significant in an analysis controlling for other predictors significantly related to compliance, which may suggest that this effect may be (partly) explained by its correlation with other study-level design features. In addition, exploratory analyses of dyad-level predictors indicated that

couple age was positively associated with compliance (b = 0.02, p < .001), whereas cohabitation status did not relate to compliance (b =0.46, p = .071).

Report Congruence

Overall report congruence across studies and relationship-related events was 93.16%. ICCs indicated that 7% of variance in congruence was situated at the study level, 12% at the dyad level, and 79% at the item-prompt level. The fixed effects of predictors on congruence are presented in Table 4. The random effect variances/covariances are reported in supplementary Table 2. None of the tested study-level predictors were significantly related to congruence. At the prompt level, study day positively predicted congruence, indicating that dyads became more congruent over the course of a study. No other prompt-level predictors were significantly associated with congruence. At the item-level, both timeframe and event subjectiveness were relevant for congruence. Items assessing current events obtained higher congruence than items assessing retrospective events. Moreover, more objective events obtained higher congruence than more subjective events. Including all significant predictors of congruence in a multiple regression model, all predictors remained significant (i.e., study day: b = 0.04, p = .006 ; item timeframe: b = -1.10, p < .001; item subjectiveness: b = -0.51, p<.001). Rerunning the analyses across 500 datasets, in which randomly one event per prompt was included, indicated that the significant effects were robust across the majority of datasets (see supplementary Figure 1). Exploratory analyses of dyad-level predictors indicated that couple age related positively to congruence (b = 0.02, p < .001), whereas cohabitation status was not associated with congruence (b = 0.09, p = .665). Additionally, exploratory analyses of the item-level predictor event valence indicated that neutral events obtained higher congruence than positive events (b = 1.13, p <.001)

and negative events obtained higher congruence than neutral events (b = 0.29, p <.01) and positive events (b = 1.42, p <.001).

Table 3

Effects of Study-, and Prompt -Level Predictors on Dyadic Compliance

Predictor (Model)		b	Z	*p*	OR	95% CI OR
Study-level						
Number of days	Intercept	1.79	2.08	<.05	5.98	[1.11, 32.28]
	Number of days	0.02	0.23	.819	1.02	[0.88, 1.18]
Number of prompts	Intercept	3.90	6.11	<.001	49.27	[14.12, 171.88]
	Number of prompts	-0.04	-2.93	**.003**	0.96	[0.94, 0.99]
Prompts per day	Intercept	2.81	3.45	<.001	16.69	[3.36, 82.76]
	Prompts per day	-0.16	-1.14	.256	0.86	[0.65, 1.12]
Questionnaire length	Intercept	1.31	1.73	.084	3.72	[0.84, 16.48]
	Questionnaire length	0.02	0.90	.368	1.02	[0.97, 1.07]
Allowed response delay	Intercept	1.29	5.06	<.001	3.63	[2.2, 5.98]
	Allowed response delay	0.39	4.87	**<.001**	1.48	[1.27, 1.74]
Prompt-level						
Study day	Intercept	2.22	6.47	<.001	9.17	[4.69, 17.96]
	Study day	-0.05	-1.35	.178	0.95	[0.89, 1.02]
Type of day	Intercept	1.96	5.00	<.001	7.13	[3.3, 15.41]
	Type of day (weekend vs weekday)	0.26	1.55	.122	1.31	[0.93, 1.83]

Time of day	Intercept	1.85	3.92	<.001	6.36	[2.52, 16.04]
	Time of day	0.04	2.26	**.024**	1.04	[1.00, 1.07]

Note. Separate models were run per predictor. Significant p-values are displayed in bold; b = estimate value of coefficient; Z = Z-value; p = p value; OR = Odds ratio (calculated as exp(b)); 95% CI OR = Wald confidence intervals for ORs. The reference categories for the predictors are as follows: Type of day (weekday).

Table 4

Effects of Study-, Prompt, and Item-Level Predictors on Report Congruence

Predictor (Model)		b	Z	*p*	OR	95% CI OR
Study-level						
Number of days	Intercept	2.42	5.20	<.001	11.28	[4.52, 28.12]
	Number of days	0.02	0.44	.662	1.02	[0.94, 1.1]
Number of prompts	Intercept	3.07	5.78	<.001	21.48	[7.59, 60.85]
	Number of prompts	-0.01	-0.91	.361	0.99	[0.97, 1.01]
Prompts per day	Intercept	3.05	8.24	<.001	21.11	[10.22, 43.6]
	Prompts per day	-0.08	-1.33	.182	0.92	[0.82, 1.04]
Questionnaire length	Intercept	3.15	8.14	<.001	23.36	[10.94, 49.91]
	Questionnaire length	-0.02	-1.52	0.129	0.98	[0.96, 1.01]
Allowed response delay	Intercept	2.59	11.19	<.001	13.28	[8.44, 20.89]
	Allowed response delay	0.01	0.20	.838	1.01	[0.89, 1.16]
Prompt-level						
Study day	Intercept	2.45	12.87	<.001	11.62	[8.00, 16.89]

	Study day	0.05	2.96	**.003**	1.05	[1.02, 1.08]
Type of day	Intercept	2.62	14.39	<.001	13.69	[9.59, 19.56]
	Type of day (weekend vs weekday)	0.01	0.11	.915	1.01	[0.83, 1.23]
Time of day	Intercept	2.59	11.70	<.001	13.38	[8.66, 20.65]
	Time of day	0.01	0.50	.615	1.01	[0.97, 1.05]
Inter-response time	Intercept	2.61	11.10	<.001	13.64	[8.59, 21.66]
	Inter-response time	-0.07	-1.17	.243	0.94	[0.84, 1.05]
Item-level						
Timeframe	Intercept	3.90	20.03	<.001	49.63	[33.87, 72.73]
	Timeframe (retrospective vs current)	-1.41	-13.64	**<.001**	0.25	[0.20, 0.30]
Event subjectiveness	Intercept	2.96	17.24	<.001	19.27	[13.77, 26.98]
	Event subjectiveness (subjective vs objective)	-0.80	-8.98	**<.001**	0.45	[0.38, 0.54]

Note. Separate models were run per predictor. Significant p-values are displayed in bold; b = estimate value of coefficient; Z = Z-value; p = p value; OR = Odds ratio (calculated as exp(b)); 95% CI OR = Wald confidence intervals for ORs. The reference categories for the predictors are as follows: Type of day (weekday), Timeframe (current) and Event subjectiveness (objective).

Discussion

The current study aimed to examine overall levels of dyadic compliance (aim 1a) and report congruence between partners concerning relationship-related events (aim 2a) across a set of

dyadic ESM studies in romantic couples. We found overall high compliance (88%) across datasets and overall high report congruence across (93%) across events and datasets. Additionally, we explored study- and prompt-level predictors of dyadic compliance (aim 1b), and found some significant predictors. Specifically, studies with a higher number of prompts in total and shorter allowed response delays were characterized by lower compliance. Additionally, dyads became more compliant over the course of individual study days. Finally, we explored study-, prompt-, and item-level predictors of report congruence (aim 2b). Dyads became more congruent over the course of the study, and were more congruent in their reports about current versus retrospective events and about more objective versus more subjective events.

Overall compliance and congruence

The overall level of dyadic compliance as found in the current study was higher than we anticipated. Specifically, in a meta-analysis of individual ESM studies, the average compliance rate was 79% (Wrzus & Neubauer, 2023) and we expected a lower compliance rate at the dyad-level, as a dyad is not compliant if just one of the partners does not answer the prompt. Of course, these findings come from different sets of studies, so comparisons should be made with caution. Nevertheless, participating as a dyad may relate to higher compliance than what can be expected by multiplying individual compliance rates, because couples are likely not independent in their missing responses (e.g., shared activities may make it more likely that they miss prompts at the same time). Supporting this idea, an additional analysis revealed a significant association between partners' momentary (individual) compliance at the prompt level (see Supplementary files). Speculatively, compliance may even be higher than in individual ESM studies as partners may motivate each other to complete prompts or alert each other to prompts. Anecdotally, in a

previous study in triads (Verhees et al., 2024; not included in the present study as it did not focus on dyads), we gathered qualitative feedback about participants' experiences and one of them stated: "We encouraged each other as much as possible to be alert to the beeps". Another speculative explanation for the relatively high compliance rates observed here is that participants in dyadic ESM studies may be somewhat more motivated overall. Because participation requires mutual agreement from both partners, it may be a more deliberate commitment compared to individual ESM studies. Additionally, the high congruence in partners' reports on relationship-related events suggests they were generally attentive in their responses to the questions, further supporting the idea of motivated participants.

Predictors of compliance and congruence

The majority of examined study-level predictors were not significantly related to compliance or congruence. Exceptions were number of prompts and allowed response delay for compliance. The direction of these effects was as predicted, such that studies that ask more of participants resulted in lower congruence and compliance. The finding that number of prompts negatively related to compliance was in line with some previous findings in individual ESM studies (Ottenstein & Werner, 2022), but not others (Wrzus & Nuebauer, 2023). Given that we found no effect of study day on compliance, and that an additional exploratory analysis showed higher compliance when participants had already received more prompts that day (see Supplementary files), the association between the total number of prompts and compliance is unlikely to reflect cumulative burden. Instead, it may relate to participants' expectations of burden. It is important to note that number of prompts became non-significant in the multiple

regression model, suggesting that this effect may partly be explained by its correlation with other study-level design features such as allowed response delay.

Indeed, studies with shorter allowed response delays led to lower dyadic compliance, as predicted. In dyadic ESM studies, if the goal is to map partners' responses onto one another, it is important to balance allowed response delay: it should be short enough to keep partners' responses temporally close for meaningful alignment, yet long enough to maintain acceptable compliance rates. Notably, most studies included here achieved acceptable compliance rates, even with shorter response delays (e.g., a maximum of 5 minutes). Future research could further investigate the optimal balance between allowed response delay and compliance.

At the prompt level, we found that participants became more compliant over the course of a study day, aligning with some prior findings on individual compliance (Rintala et al., 2019) but not others (Sun et al., 2021). It seems unlikely that this pattern can be explained by participants being asleep during early prompts (as previously suggested, e.g., Rintala et al., 2019), as across studies, the first prompt could be completed until 10 am or later, by which time most people are awake (Walch et al., 2016). Higher compliance later in the day may reflect greater availability outside work hours, or increased partner presence, which could enhance the aforementioned potential benefits of participating together. Additionally, couples became more congruent in their reports about relationship-related events as the studies progressed. This contradicts the hypothesis that participants become tired of repeatedly answering the same questions, increasing inattentive or inaccurate responses over time (fatigue effect), but rather suggests a learning effect (Verbeij et al., 2021). Potentially, repeatedly answering the same questions heightens participants' awareness of relevant events, improving accuracy over time

(Fuller-Tyszkiewicz et al., 2013). Alternatively, partners may discuss and align their interpretations of the items as the study unfolds.

At the item level, results indicated that items asking about current (vs. retrospective) events and more objective (vs. subjective) events showed higher congruence rates, as predicted. Of note, retrospective and subjective items still achieved acceptable congruence rates in the current study (92.40% and 89.69%, respectively). Additionally, the number of current items in our study was limited and all current items were objective, restricting the conclusions about timeframe. This also means that timeframe and subjectiveness were related. When we simultaneously examined both predictors, however, they both remained significant indicating that the effects of timeframe cannot be explained by subjectiveness or vice versa. Exploratory analyses further indicated that event valence related to congruence, such that partners were more congruent in their reporting of neutral compared to positive events, and more congruent in their reporting of negative compared to neutral events. As for the difference between neutral and positive events, this may (partially) be explained by event subjectiveness and/or timeframe as all positive events were subjective and retrospective. However, negative events were also subjective and retrospective, and for these, congruence was higher than for neutral events. This may suggest that negative events are salient to both partners (Baumeister et al., 2001), increasing the likelihood that they perceive them in similar ways which may align their responses. Alternatively, base rate of event occurrence may play a role, as negative events occurrence was rather rare (see Table 2) and congruence is likely to be higher when events are very rare.

We additionally explored the effects of dyad-level predictors couple age and cohabitation status on dyadic compliance and report congruence. Only couple age positively related to both compliance and congruence. For compliance, this aligns with some previous studies finding

higher compliance rates in older participants (Ono et al., 2019; Rintala et al., 2019). Additionally, previous research on careless responding found that younger participants may be more likely to respond carelessly (Schneider et al., 2018), which aligns with the current finding that congruence was lower for younger couples. Speculatively, older couples' lives may be less demanding (e.g., they may be retired, or have older children), potentially giving them more capacity to respond consistently and attentively (Ono et al., 2019).

Limitations and future directions

The current findings should be interpreted in light of their limitations. First, we included a total of eight studies, which may have limited our power to detect study-level predictors. Moreover, individual studies may have had a substantial influence on the findings. For example, Study 8 had a large sample size in terms of included couples and assessed only one event-related item which occurred rarely. One could argue that this may have contributed to a higher overall congruency rate, as congruence is likely to be higher when events occur infrequently. However, it should be noted that some other studies obtained even higher congruency rates than Study 8 (see Table 1). Nevertheless, future research should aim to conduct a comprehensive meta-analysis of dyadic ESM studies to address these limitations. A challenge for such an analysis will be to include studies with lower compliance or congruence rates (if they exist), as these may be less likely be published or shared by researchers. In the current study, we may also have primarily included studies with higher compliance and congruence rates, leading to (potentially overly) optimistic estimates for these factors. Second, although we examined several predictors at different levels that could be relevant for compliance and congruence, additional predictors may warrant further investigation. For example, incentives (specifically, monetary vs non-

monetary; Wrzus & Neubauer, 2023) have been shown to matter for individual compliance. We were unable to assess this predictor here because all included studies provided monetary incentives. In addition, other study-design-related factors, such as whether and how participants were briefed about the study, including about the importance of compliance and the intended meaning of specific items, might be relevant for future studies to explore, although obtaining this information may prove more challenging. Relatedly, we only explored limited dyad- or sample-related characteristics as predictors for compliance and congruence. In individual ESM meta-analyses, sample-related characteristics have been found to not relate to compliance (Wrzus & Nuebauer, 2023; Ottenstein & Werner, 2022). Nevertheless, future research may consider examining features of the sample or dyad as predictors of compliance and congruence. For example, time spent together could be a relevant predictor for examining whether social dynamics affect dyadic ESM compliance, although we did not find any evidence that couples who cohabited (which may serve as a proxy for time spent togethers) showed higher compliance. Finally, our evaluation of study-level predictors was constrained by the variability in design choices among the included studies. Such design choices are often balanced in order to keep the study feasible in terms of participant burden (Wrzus & Neubauer, 2023). Experimental manipulation of single factors, as has previously been done for individual compliance and careless responding (Eisele et al., 2022), may provide more precise insights into how specific design choices affect dyadic ESM data quantity and quality.

Constraints on generality

Our study examined dyadic compliance and report congruence in romantic couples, the majority of whom were different-gender. Information on race, ethnicity, and socioeconomic

status was not consistently available across datasets. As a result, the findings may not generalize to more diverse populations or to couples with demographic characteristics not represented in these samples, nor to other types of dyads (e.g., parent-child, friends, colleagues).

Implications for future dyadic ESM studies

Despite the limitations noted above, the current study provides some implications for future dyadic ESM studies. First, it suggests that running a dyadic ESM study is feasible when looking at dyadic compliance and congruence rates, as these were generally high. To optimize compliance and congruence rates, future studies should aim to minimize participant burden. Specifically, longer allowed response delays may improve dyadic compliance. As mentioned previously, however, it remains important to keep allowed response delay short enough to keep partners' responses temporally close if the goal is to map these onto one another. Additionally, given that we observed higher compliance later in the day, future studies may consider scheduling prompts mostly later in the day, potentially after working hours or when partners are more likely to be together (which may also be beneficial to capture dyadic dynamics). Moreover, when using event-based items as branching items or to select data, researchers should consider using current and explicit events to optimize data quantity and quality. Potentially, explaining item meanings during participant briefing may clarify their intent and help prevent discrepancies due to differences in interpretation.

Conclusion

Overall, our findings suggest that, given the designs and approaches used in the studies examined here, dyadic compliance and congruence were generally high, despite some

fluctuations across and within studies. Our findings further indicate that a higher number of prompts across the study and shorter allowed response delays were associated with lower compliance. In addition, dyads were more likely to complete surveys later in the day and were more likely congruent later in the study, and for items that assess current (vs retrospective) and more objective (vs subjective) events. Nevertheless, further research is needed to examine the robustness of these predictors.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323-370. https://doi.org/10.1037/1089-2680.5.4.323

Bolis, D., Dumas, G., & Schilbach, L. (2023). Interpersonal attunement in social interactions: From *collective* psychophysiology to *inter- personalized* psychiatry and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1870), 20210365. https://doi.org/10.1098/rstb.2021.0365

De Los Reyes, A., Ohannessian, C. M., & Racz, S. J. (2019). Discrepancies Between Adolescent and Parent Reports About Family Relationships. *Child Development Perspectives*, *13*(1), 53–58. https://doi.org/10.1111/cdep.12306

Dejonckheere, E., Verdonck, S., Andries, J., Röhrig, N., Piot, M., Kilani, G., & Mestdagh, M. (2024). Real-time incentivizing survey completion with game-based rewards in experience sampling research may increase data quantity, but reduces data quality. *Computers in Human Behavior*, *160*, 108360. https://doi.org/10.1016/j.chb.2024.108360

Ehrlich, K. B., vanDellen, M. R., Felton, J. W., Lejuez, C. W., & Cassidy, J. (2019). Perceptions about marital conflict: Individual, dyadic, and family level effects. *Journal of Social and Personal Relationships*, *36*(11–12), 3537–3553. https://doi.org/10.1177/0265407519829846

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The Effects of Sampling Frequency and Questionnaire Length on Perceived

Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (2023). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. *Psychological Assessment*, *35*(1), 68–81. https://doi.org/10.1037/pas0001177

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning Variation in Multilevel Models. *Understanding Statistics, 1*, 223–231. https://doi.org/10.1207/S15328031US0104_02

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2021). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods*, *54*(4), 1541–1558. https://doi.org/10.3758/s13428-021-01683-6

Hopwood, C. J. (2018). Interpersonal Dynamics in Personality and Personality Disorders. *European Journal of Personality*, *32*(5), 499–524. https://doi.org/10.1002/per.2155

Horn, A. B., Holzgang, S. A., & Rosenberger, V. (2021). Adjustment of Couples to the Transition to Retirement: The Interplay of Intra- and Interpersonal Emotion Regulation in Daily Life. *Frontiers in Psychology*, *12*, 654255. https://doi.org/10.3389/fpsyg.2021.654255

Ji, Y., Revol, J., Schouten, A., Schreuder, M., & Ceulemans, E. (2025). Performance of the longitudinal actor-partner interdependence model in case of large amounts of missing values: Challenges and possible alternatives. *Psychometrika, 90,* 508-530. https://doi.org/10.1017/psy.2025.18

Ono, M., Schneider, S., Junghaenel, D. U., & Stone, A. A. (2019). What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis. *Journal of Medical Internet Research*, *21*(2), e11398. https://doi.org/10.2196/11398

OpenAI. (2025). ChatGPT [Large language model]. https://chat.openai.com/

Ottenstein, C., & Werner, L. (2022). Compliance in Ambulatory Assessment Studies: Investigating Study and Sample Characteristics as Predictors. *Assessment*, *29*(8), 1765–1776. https://doi.org/10.1177/10731911211032718

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reiter, T., Sakel, S., Scharbert, J., ter Horst, J., van Zalk, M., Back, M., Bühner, M., & Schoedel, R. (2025). Side Effects of Experience-Sampling Protocols: A Systematic Analysis of How They Affect Data Quality, Data Quantity, and Bias in Study Results. *Advances in Methods and Practices in Psychological Science, 8,* 25152459251347274. https://doi.org/10.1177/25152459251347274

Righetti, F., Gere, J., Hofmann, W., Visserman, M. L., & Van Lange, P. A. M. (2016). The burden of empathy: Partners' responses to divergence of interests in daily life. *Emotion*, *16*(5), 684–690. https://doi.org/10.1037/emo0000163

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

Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research, 27*, 1077-1088. https://doi.org/10.1007/s11136-017-1767-2

Schoebi, D., Wang, Z., Ababkov, V., & Perrez, M. (2010). Affective interdependence in married couples' daily lives: Are there cultural differences in partner effects of anger? *Family Science*, *1*(2), 83–92. https://doi.org/10.1080/19424620903471681

Sels, L., Ceulemans, E., & Kuppens, P. (2017). Partner-expected affect: How you feel now is predicted by how your partner thought you felt before. *Emotion*, *17*(7), 1066–1077. https://doi.org/10.1037/emo0000304

Sels, L., Ruan, Y., Kuppens, P., Ceulemans, E., & Reis, H. (2020). Actual and Perceived Emotional Similarity in Couples' Daily Lives. *Social Psychological and Personality Science*, *11*(2), 266–275. https://doi.org/10.1177/1948550619845927

Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed Beeps and Missing Data: Dispositional and Situational Predictors of Nonresponse in Experience Sampling Research. *Social Science Computer Review*, *31*(4), 471–481. https://doi.org/10.1177/0894439313479902

Sun, J., Rhemtulla, M., & Vazire, S. (2021). Eavesdropping on Missing Data: What Are University Students Doing When They Miss Experience Sampling Reports? *Personality and Social Psychology Bulletin*, *47*(11), 1535–1549. https://doi.org/doi.org/10.1177/01461672209646

Ulitzsch, E., Nestler, S., Lüdtke, O., & Nagy, G. (2024). A screen-time-based mixture model for identifying and monitoring careless and insufficient effort responding in ecological momentary assessment data. *Psychological Methods*. https://doi.org/10.1037/met0000636

Ulitzsch, E., Viechtbauer, W., Lüdtke, O., Myin-Germeys, I., Nagy, G., Nestler, S., & Eisele, G. (2024). *Investigating the effect of experience sampling study design on careless and insufficient effort responding identified with a screen-time-based mixture model*. https://doi.org/10.31234/osf.io/2s38a

Van Berkel, N., Ferreira, D., & Kostakos, V. (2018). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, *50*(6), 1–40. https://doi.org/10.1145/3123988

Van Berkel, N., Goncalves, J., Hosio, S., Sarsenbayeva, Z., Velloso, E., & Kostakos, V. (2020). Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies*, *134*, 1–12. https://doi.org/10.1016/j.ijhcs.2019.10.003

Verbeij, T., Pouwels, J. L., Beyens, I., & Valkenburg, P. M. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports*, *3*, 100090. https://doi.org/10.1016/j.chbr.2021.100090

Walch, O. J., Cochran, A., & Forger, D. B. (2016). A global quantification of "normal" sleep schedules using smartphone data. *Science Advances*, *2*(5), e1501705. https://doi.org/10.1126/sciadv.1501705

Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, *30*(3), 825–846. https://doi.org/doi.org/10.1177/107319112110675