Draft version 01/06/2024. This manuscript is a preprint and has not been peer-reviewed.

Investigating dynamics in attentive and inattentive responding together with their contextual correlates using a novel mixture IRT model for intensive longitudinal data

Leonie V.D.E. Vogelsmeier[1,*], Irina Uglanova[2], Manuel Rein[1], and Esther Ulitzsch[3,4,*]

[*]L.V.D.E.V. and E.U. contributed equally to this work

[1]Tilburg University

[2]IPN—Leibniz Institute for Science and Mathematics Education

[3]Centre for Educational Measurement, University of Oslo

[4]Centre of Research on Equality in Education, University of Oslo

Author Note

Correspondence concerning this article should be sent to either Leonie V. D. E. Vogelsmeier, Department of Methodology and Statistics, Tilburg University, Warandelaan 2, 5037 AB Tilburg, Netherlands, or to Esther Ulitzsch, Centre for Educational Measurement, University of Oslo, Gaustadalleen 21, 0349 Oslo, Norway. Online materials, including the data used for this article, can be found at the OSF and are available via the following link: https://osf.io/uj6sr/. Email: l.v.d.e.vogelsmeier@tilburguniversity.edu or esther.ulitzsch@cemo.uio.no. This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 33160.

Abstract

In ecological momentary assessment (EMA), respondents answer brief questionnaires about their current behaviors or experiences several times per day across multiple days, resulting in intensive longitudinal data. The frequent measurement enables a thorough grasp of the dynamics inherent in psychological constructs, but it also increases respondent burden. To lower this burden, respondents may engage in careless and insufficient effort responding (C/IER), leaving data contaminated with responses that do not reflect what researchers want to measure. We introduce a novel approach to investigate C/IER in EMA data. Our approach combines a confirmatory mixture item response theory model separating C/IER from attentive behavior with latent Markov factor analysis. This allows for (1) gauging the occurrence of C/IER and (2) studying transitions among states of different response behaviors as well as their contextual correlates. The approach can be implemented using R packages. In an empirical application, we showcase the efficacy of this approach in both pinpointing C/IER instances in EMA and gaining insights into their underlying causes. In a simulation study investigating robustness against unaccounted changes in measurement models underlying attentive responses, the approach proved robust against heterogeneity in loading patterns but not against heterogeneity in the factor structure. Extensions to accommodate the latter are discussed.

Keywords: Careless responding; ecological momentary assessment; mixture modeling; latent Markov modeling; three-step approach; time series; intensive longitudinal data

Intensive longitudinal data collected via methods like ecological momentary assessment (EMA) have great potential for studying the dynamics of psychological constructs such as well-being in daily life (Scollon et al., 2009). In EMA, respondents answer brief questionnaires about their current behaviors or experiences in daily-life situations several times per day over a prolonged period of time via mobile phone apps. The frequent and repetitive measures provide researchers and practitioners with valuable insights, enabling a thorough grasp of the dynamics inherent in psychological traits or constructs (Myin-Germeys & Kuppens, 2021; Wrzus & Mehl, 2015).

EMA's demanding sampling schedule, however, also increases respondents' burden and, as a consequence, may decrease their willingness and/or ability to invest effort into completing the study, rendering EMAs vulnerable to respondent non-compliance (Dejonckheere & Erbas, 2021). In the present study, we focus on careless and insufficient effort responding (C/IER) as a prominent form of respondent non-compliance. C/IER occurs when respondents complete questionnaires without investing effort into carefully evaluating the administered items (Huang et al., 2015; Ulitzsch et al., 2024). As a consequence, EMA data may be contaminated with responses that do not reflect what researchers want to measure.

Having techniques that allow to identify and investigate (dynamics in) C/IER in EMA data is important for at least two reasons. First, from a methodological perspective, detecting C/IER is vital for accurate inferences about the dynamics of psychological constructs (DeSimone et al., 2018; Huang et al., 2012; Kam & Meyer, 2015; Schmitt & Stuits, 1985; Woods, 2006). C/IER rates of as low as 5% that remain undetected and unaccounted for can comprise psychometric properties of the questionnaires and introduce bias in the correlations between measured constructs (Huang et al., 2015; McGrath et al., 2010). Consequently, ignoring C/IER can result in inaccurate interpretations and hinder the replication of research findings (Curran, 2016).

Second, from a substantive point of view, C/IER identification allows to study person and contextual characteristics related to its occurrence. This may raise researchers' awareness of contexts that, due to an increased risk of C/IER and the resulting lowered data quality, are hard to study and for which further incentives may be introduced to increase data quality. This is particularly relevant for EMA because individuals complete questionnaires in daily life, and may be exposed to environmental distractions that may impede their capacity to respond attentively. For instance, if respondents cannot concentrate, perhaps because they are preparing for an important work presentation later that day, they may engage in C/IER.[1] Further, understanding determinants of C/IER may ultimately aid in designing EMA studies that curb its occurrence and yield data of higher quality. For instance, researchers can investigate how sampling schedules affect C/IER occurrence and consider conclusions drawn from these investigations when designing EMAs.

Although studied excessively in cross-sectional research (Bowling et al., 2016; Huang et al., 2015; Maniaci & Rogge, 2014; McKay et al., 2018; Nichols & Edlund, 2020), only a few studies have focused on C/IER in EMA (Eisele et al., 2022; Hasselhorn et al., 2022, 2023; Jaso et al., 2022), with a potential reason being that tools for identifying and investigating C/IER in EMA data are still scarce. Research aiming to detect C/IER in EMA data has predominantly relied on established behavioral indicators for C/IER detection (see Curran, 2016; Meade & Craig, 2012; Niessen et al., 2016, for overviews in the context of cross-sectional data). These approaches are limited, however, in that they rely on somewhat arbitrary threshold settings separating attentive from C/IER behavior and do not take C/IER identification uncertainty into account. While mixture modeling approaches overcoming these limitations constitute a rich body of research in the cross-sectional context (e.g., Arias et al., 2020; Ulitzsch, Pohl, et al., 2022; Ulitzsch, Yildirim-Erbasli, et al., 2022; van Laar & Braeken, 2022), model-based approaches tailored

[1] Of course, respondents may also opt not to respond. Non-response is another form of non-compliance (see Jones et al., 2019; Morren et al., 2009; Ono et al., 2019; Wrzus & Neubauer, 2023). Whether individuals engage in non-response or C/IER may depend on specific factors, such as the desire to achieve a high completion rate of questionnaires for a higher monetary reward. In this article, we exclusively focus on C/IER and not on non-response.

to the EMA context emerged only recently and have, so far, only been developed for collateral information from screen time data (Ulitzsch et al., 2024).

In this study, we aim to expand EMA researchers' toolkit with a response-pattern-based mixture modeling approach for investigating dynamics in attentive and inattentive responding together with their contextual correlates. To this end, we blend *confirmatory* mixture item response theory (IRT) models developed for cross-sectional research (Uglanova et al., in preparation) with latent Markov factor analysis (LMFA; Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019), which has originally been developed for EMA to support investigations of transition patterns between *exploratorily* identified changes in the measurement model (i.e., changes in the extent to which items measure constructs).

In what follows, we first review current approaches for C/IER detection in EMA data. Next, we provide a short introduction to LMFA and present an approach that substitutes LMFA's exploratory identification of measurement model changes with theory-based IRT component models for attentive and inattentive behavior. We delineate how the proposed approach can easily be implemented with standard R packages. Using real EMA data, we illustrate the method and showcase how it allows studying transition patterns between attentive and inattentive responding and their contextual correlates. In two simulation studies, we evaluate robustness against violations of the key assumption that the same measurement model holds across all attentive observations. From these simulation studies, we conclude that C/IER identification is robust against unaccounted heterogeneity in attentive loading patterns across observations, but not against unaccounted heterogeneity in the factor structure underlying attentive responses. We conclude by discussing extensions to accommodate changes in attentive measurement models.

C/IER Detection in EMA Data

Attention Check and Bogus Items

Attention check and bogus items are items that elicit a specific attentive response (such as disagreement with "I enjoy eating cement" or compliance with the prompt "Choose response option 5"), such that deviations from this response can be assumed to signal C/IER. However, these items are unsuitable for fine-grained monitoring of C/IER in EMAs, as repeated and extensive administration may confuse attentive respondents (Meade & Craig, 2012). In addition, attention check and bogus items can be suspected to have very low power. For instance, in an EMA survey with 163 respondents, Eisele et al. (2022) found only four respondents failing instructional manipulation check items more than once, while multiple other indicators signaled markedly higher C/IER proportions.

Indicators Based on Response Patterns and Screen Times

Behavioral indicators based on response patterns of collateral information scan these for aberrances presumably signaling C/IER. Using the long-string index (Johnson, 2005) to scan for suspiciously low response variability or the exclusion of observations associated with screen times that are too short to properly evaluate the administered items (see Hasselhorn et al., 2022, for an application to EMA data) are well-known examples. When using such indicators, researchers have to decide on thresholds separating attentive behavior from C/IER. This decision is ultimately an arbitrary one, with even minor differences in threshold settings often heavily impacting conclusions on C/IER contamination (Niessen et al., 2016; Ulitzsch, Domingue, et al., 2023; Ulitzsch, Shin, & Lüdtke, 2023).

Data-driven approaches to identify C/IER observations with response-pattern-based and screen time indicators mitigate the subjectivity of threshold settings. Jaso et al. (2022) suggested setting thresholds such that the cleaned data set minimizes implausible positive correlations among psychometric antonyms (e.g., relaxed and anxious). Further, latent class analysis on one or more indicators can be employed (see Kam & Meyer, 2015; Maniaci

& Rogge, 2014, for applications in the cross-sectional context; and Hasselhorn et al., 2023, for a multilevel extension for EMA data), entirely avoiding the setting of thresholds. In this approach, classes with aberrant conditional means on the analyzed indicators are labeled as C/IER. Note, however, that post-hoc interpretation of the obtained latent classes may not always be straightforward (see Ulitzsch et al., 2024, for further examples and discussions). Hence, despite these recent advances, a common limitation of approaches relying on behavioral indicators remains that they are only loosely grounded in subject-matter theory on respondent behavior. This renders them susceptible to ambiguity and arbitrariness—either in setting thresholds or post-hoc interpretation of latent classes.

Confirmatory Mixture Model Based on Screen Times

As an alternative to post-hoc interpretations of latent classes as attentive or careless, *confirmatory* mixture models can be employed. These translate theoretical considerations on respondent behavior into two mixture component models—one representing an assumed attentive and the other one an assumed inattentive data-generating process. Posterior class probabilities obtained from the resultant two-class mixture model can then be employed to draw conclusions on the attentiveness of a given observation.

Such models have been predominantly developed for cross-sectional item response data (e.g., Arias et al., 2020; Kam & Cheung, 2023; Ulitzsch, Yildirim-Erbasli, et al., 2022; van Laar & Braeken, 2022). In these models, attentive item responses are assumed to reflect the to-be-measured traits, i.e., to follow standard measurement models for item response data—also referred to as latent trait models—such as confirmatory factor analysis models (Arias et al., 2020; Kam & Cheung, 2023) or IRT models for polytomous data (Ulitzsch, Pohl, et al., 2022, 2023; Ulitzsch, Yildirim-Erbasli, et al., 2022; van Laar & Braeken, 2022). Inattentive item responses, in contrast, are assumed to be unrelated to the items' content (as respondents do not invest effort into evaluating the items) and the traits to be measured (because respondents do not invest effort in selecting relevant response options). Instead, inattentive item responses are assumed to be driven by respondents' category preferences (Arias et al., 2020), to be purely random (van Laar & Braeken, 2022), or both (Kam & Cheung, 2023; Uglanova et al., in preparation).

Only recently, Ulitzsch et al. (2024) provided a confirmatory mixture modeling approach for C/IER tailored to the EMA context. This model is formulated for screen times instead of for item responses. Screen times can easily be recorded in electronically administered EMAs and indicate how much time respondents required to interact with the items presented on a given screen. Hence, the component models for this confirmatory mixture model reflect theoretical considerations on how this time may evolve in EMAs. For attentive screen times, the model assumes an exponential decay process, capturing possible speed-ups due to familiarization and practice with the EMA delivery system and the administered measures. Inattentive screen times are assumed to randomly fluctuate and to be, on average, shorter than screen times of attentive respondents who familiarized themselves with the EMA. The model allows attentiveness to vary on the respondent-by-occasion level and can be enriched with person- and occasion-level covariates, thereby providing a sophisticated tool to investigate time-invariant (e.g., affinity to technology) and time-varying (e.g., time of the day) covariates of C/IER occurrence.

Nevertheless, this model-based approach is of somewhat limited practical applicability. First, it comes with strong requirements for EMA data collection. For instance it is not clear yet how to handle routing designs. These pose a challenge because they result in questionnaires that vary in length across respondents and, as such, in the time attentive respondents require. Second, due to its complexity, the approach is not very easy to apply and estimation issues are likely to occur. Third, having only recently been developed, validity evidence that this approach indeed accurately captures C/IER is still pending. In the present study, we, therefore, aim to complement this approach with an easily applicable approach leveraging item responses. To this end, we integrate confirmatory mixture IRT models for cross-sectional data with LMFA originally developed for exploratory investigations of EMA measurement model changes.

Exploratory Mixture Model Based on Item Responses for Exploring Measurement Model Changes

The recently proposed LMFA (Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019) is an exploratory mixture model that classifies individual- and timepoint-specific observations from EMA data into latent states based on differences in response behavior. Different response behaviors are modeled with measurement models identified with either exploratory factor analysis for continuous data (Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019) or exploratory IRT models for ordinal data (Vogelsmeier, Vermunt, Keijsers, & De Roover, 2021). Transitions between the latent states representing the different response behaviors are captured via a latent Markov model (LMM), which sheds light on the overall probabilities of transitioning to a particular state. These probabilities can be related to covariates capturing individual or situational characteristics. LMFA thus holds promise in unveiling the moments when individuals transition between different response behaviors and identifying correlates with individual and situational characteristics. Its major advantages lies in its flexibility, as it is capable of uncovering manifold changes in measurement models. Vogelsmeier, Cloos, et al. (2023), for instance, identified two affect structures and identified contextual (negative event intensity) and person characteristics (neuroticism) related to transition patterns between affect structures. In principle, LMFA could also be applied to uncover C/IER: Exploring what the state-specific measurement models and the relationships between the covariates and state memberships look like may reveal that one state corresponds to attentive responding and another to C/IER. However, as highlighted by Vogelsmeier (2022), LMFA is currently not tailored to distinguish between attentive responding and C/IER as the method merely differentiates between any apparent difference in response behavior, and post-hoc interpretation of the latent states is likely ambiguous.

Proposed Method: LMFA with Confirmatory Mixture IRT models

In this article, we combine LMFA (originally developed to identify and explore measurement model changes and their contextual correlates in EMA data) with theory-based confirmatory mixture IRT models for C/IER (originally developed for detecting C/IER in cross-sectional data). The new methodology optimally leverages item information by clustering observations into two states that are in advance defined as attentive responding and C/IER using specific measurement model constraints. This overcomes the need for post-hoc interpretation of possibly ambiguous latent states resulting from the exploratory models used in regular LMFA. As previously developed mixture modeling approaches to C/IER in EMA data (Hasselhorn et al., 2023; Ulitzsch et al., 2024), this new method does not require decisions on threshold settings and takes C/IER classification uncertainty into account. The method is applicable to EMA studies employing multiple items with ordinal response scales to measure psychological constructs. Possible extensions are discussed in the Discussion Section.

Method

The proposed methodology builds upon the traditional LMFA model that comprises two parts: (1) the state-specific measurement models determining the nature of the relations between items and constructs, which are obtained via exploratory mixture factor analysis models (Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019) or exploratory ordinal IRT models (Vogelsmeier, Vermunt, Keijsers, & De Roover, 2021) and (2) the latent Markov transition model describing how individuals transition between the latent measurement-model states over time and how these transitions correlate with individual- or context-specific characteristics. What distinguishes our proposed methodology from the regular LMFA is the measurement model specification, which is no longer exploratory but confirmatory. Specifically, the presence of relationships between items and constructs are determined in advance using theory-based confirmatory mixture IRT modeling specifically tailored to capture attentive responding versus C/IER. In the following, we first explain the data structure, then detail the measurement model specification, and finally explain the transition model specification.

Data Structure

We assume intensive longitudinal data with observations nested within individuals. For every measurement occasion, we consider scales with $G + 1$ ordered response categories. We denote with $y_{ijt} \in \{0, \dots, G\}$ the response of respondent $i \in \{1, \dots, N\}$ to item $j \in \{1, \dots, J\}$ on measurement occasion $t \in \{1, \dots, T\}$. The J items are assumed to measure a set of M substantive traits. Note that the number of timepoints T typically differs across respondents. However, we omit the index i in T_i for simplicity. The observations are collected in the $1 \times J$ vectors $\mathbf{y}_{it} = (y_{i1t}, \dots, y_{iJt})$, which are collected in the $T \times J$ subject-specific data matrices $\mathbf{Y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iT})'$. The data matrices are concatenated in the dataset $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_N)'$ with $\sum_{i=1}^{N} T_i$ rows.

Measurement Model

To disentangle inattentive from attentive responding, we employ the two-state confirmatory mixture IRT model proposed by Uglanova et al. (in preparation), which incorporates previous IRT models for C/IER (Ulitzsch, Pohl, et al., 2022, 2023; van Laar & Braeken, 2022) as special cases. We assume that at each measurement occasion t, respondent i can be in one of two latent states $k \in \{1, 2\}$. The state memberships are indicated via the binary indicators s_{itk}. These are equal to 1 for state k and equal to zero for the other state. Specifically, $s_{it1} = 1$ denotes an attentive state membership and $s_{it2} = 1$ denotes an inattentive state membership.

Attentive responses are assumed to follow a (possibly multidimensional) graded response model (GRM; Samejima, 1969, 2016). That is, the probability that an attentive respondent i selects response option g or higher on item j on measurement occasion t is modeled as

$$\begin{aligned} p(y_{ijt} \geq g \mid s_{it1} = 1) = \frac{\exp(\sum_m^M \alpha_{jm}\theta_{imt} + \kappa_{jg})}{1 + \exp(\sum_m^M \alpha_{jm}\theta_{imt} + \kappa_{jg})} \qquad & \text{for } g \in \{1, \ldots, G\} \\ p(y_{ijt} \geq g \mid s_{it1} = 1) = 1 \qquad & \text{for } g = 0 \\ p(y_{ijt} \geq g \mid s_{it1} = 1) = 0 \qquad & \text{for } g = G + 1, \end{aligned} \tag{1}$$

where α_{jm} gives the loading parameter of item j on the substantive trait m, θ_{imt} gives respondent i's location on trait m at the tth measurement occasion, and κ_{jg} is the gth category threshold for item j. Note that raw item responses are modeled, i.e., item responses to negatively worded items are not re-coded. Hence, in the case that there are negatively worded items in the administered scales, loading parameters α_{jm} can be expected to take negative values. For model identification, latent trait expectations and variances are set to 0 and 1, respectively.

The probability that an attentive respondent i selects response option g on item j on measurement occasion t can simply be obtained as

$$p(y_{ijt} = g \mid s_{it1} = 1) = p(y_{ijt} \geq g \mid s_{it1} = 1) - p(y_{ijt} \geq g + 1 \mid s_{it1} = 1). \tag{2}$$

When being inattentive ($s_{it2} = 1$), respondents are assumed to provide responses that are not reflective of the to-be-measured substantive traits. Instead of being elicited by the items' content, inattentive responses are assumed to be driven by mere category preferences that generalize across all items administered at measurement occasion t. In the inattentive measurement model, this assumption is incorporated by substituting the possibly multidimensional substantive traits with a unidimensional trait that captures respondents' category preferences. Hence, a unidimensional GRM measuring respondent i's category preferences ξ_{it} at measurement occasion t is assumed, where all loadings are set to

1 and thresholds are set to be the same across items, i.e.,

$$p(y_{ijt} \geq g \mid s_{it2} = 1) = \frac{\exp(1\xi_{it} + \kappa_g)}{1 + \exp(1\xi_{it} + \kappa_g)} \qquad \text{for } g \in \{1, \ldots, G\} \tag{3}$$

Note that, because inattentive respondents are assumed to not pay attention to the items' content, loadings for both positively and negatively worded items are set to 1. Hence, analyzing scales with both positively and negatively worded items facilitates the detection of C/IER because the attentive and inattentive component models will have markedly different loading structures, especially in conditions that otherwise challenge C/IER identification, for instance, when unidimensional scales are modeled. For model identification, the latent trait expectation is set to 0, while its variance is freely estimated.

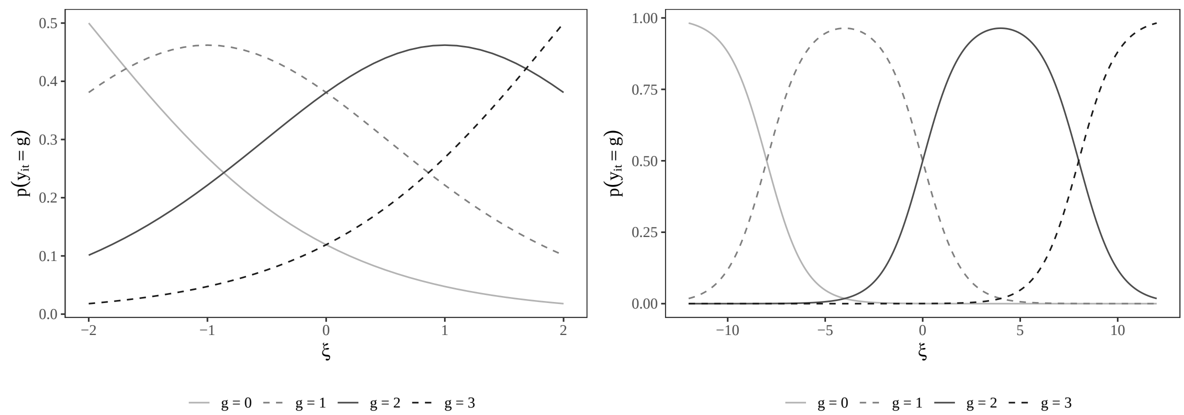
Based on theoretical considerations and simulation evidence, Uglanova et al. (in preparation) delineated that the inattentive component model can accommodate different types of careless behavior—spanning both random responding as well as pronounced respondent-specific category preferences. For instance, when all respondents exhibit random responding, $var(\xi) \approx 0$. In this case, with the latent trait expectation set to zero, Equation 3 essentially reduces to

$$p(y_{ijt} \geq g \mid s_{it2} = 1) = \frac{\exp(\kappa_g)}{1 + \exp(\kappa_g)} \qquad \text{for } g \in \{1, \ldots, G\} \tag{4}$$

and thresholds κ_g determine category probabilities of random response behavior.

Figure 1 illustrates different scenarios of inattentive respondents differing in their category preferences. In Figure 1a, inattentive respondents exhibit mildly pronounced category preferences, i.e., depending on their location on the latent category preference trait, tend to slightly favor lower, middle, or upper response categories. Figure 1b, in contrast, depicts a scenario where inattentive respondents have a strong preference for a specific category, i.e., exhibit behavior that closely resembles straight-lining. Here, category boundaries are further spaced apart and $var(\xi)$ is larger than in the scenario depicted in

Figure 1a.



(a) Mild category preferences (b) Strong category preferences

Figure 1. Illustrative item characteristic curves for different types of C/IER behavior. ξ: category preference

Uglanova et al. (in preparation) further illustrated that their proposed component model is well capable of capturing a blend of different types of inattentive behavior, for instance, when some respondents exhibit random responding and others have strongly pronounced category preferences. In this case, the C/IER component model will be misspecified, but still capable of “absorbing” the mixture of different inattentive response patterns.

Transition Model

The transition model describes how individuals transition between the attentive and C/IER states and is obtained using as LMM (e.g., Bartolucci et al., 2014), which is defined by three distinct types of parameters: The initial state, transition, and response probabilities. Together, the parameters form the joint distribution of the observations and states for subject i:

$$
\begin{aligned}
p(\mathbf{Y}_i, \mathbf{S}_i \mid \mathbf{Z}_i) &= p(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT}, \mathbf{s}_{i1}, \ldots, \mathbf{s}_{iT} \mid \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}) \\
&= \overbrace{p(\mathbf{s}_{i1} \mid \mathbf{z}_{i1})}^{\text{initial state probabilities}} \prod_{t=2}^{T} \overbrace{p_{\kappa_{ti}}(\mathbf{s}_{it} \mid \mathbf{s}_{it-1}, \mathbf{z}_{it})}^{\text{transition probabilities}} \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_{it} \mid \mathbf{s}_{it})}^{\text{response probabilities}} .
\end{aligned} \tag{5}
$$

Note that the 2×1 vectors $\mathbf{s}_{it} = (s_{it1}, s_{it2})'$ contain the binary indicators s_{itk}. The $U \times 1$ vectors $\mathbf{z}_{it} = (z_{it1}, \ldots, z_{itU})'$ comprise the covariate values z_{itu}, with $u = 1, \ldots, U$, referring to the subject- and timepoint-specific covariates possibly influencing the initial or transition probabilities as described below. The state-specific response probabilities $p(\mathbf{y}_{it} \mid s_{itk} = 1)$ indicate the probability for the attentive and C/IER response patterns at timepoint t given the state membership at that timepoint, $s_{i1k} = 1$. These probabilities depend on the two state-specific models as described above.

The initial state probabilities indicate the probabilities of starting in state k at timepoint $t = 1$. The probabilities can depend on covariate values at the first timepoint, indicated as $\mathbf{z}_{i1}$, and the probabilities are collected in a 2×1 probability vector $\boldsymbol{\pi}$ with elements $\pi_k = p(s_{i1k} = 1|\mathbf{z}_{i1})$ and $\sum_{k=1}^{2} \pi_k = 1$. The initial state probabilities are typically modeled via a logit model to prevent parameter range restrictions:

$$
\log \left(\frac{p(s_{i1k} = 1 \mid \mathbf{z}_{i1})}{p(s_{i11} = 1 \mid \mathbf{z}_{i1})} \right) = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{z}_{it=1}, \tag{6}
$$

where the initial state intercepts are denoted by β_{0k} and the initial state slopes that quantify the effect of the covariates on the initial state memberships are captured by the vectors $\boldsymbol{\beta}_k' = (\beta_{k,Z_{i11}}, \ldots, \beta_{k,Z_{i1U}})'$ for $k = 2$ and with $k = 1$ as the reference category.

The transition probabilities describe the probabilities of staying in the attentive or C/IER state or of transitioning to the respective other state. Hence, they indicate the probabilities of being in state k at timepoint $t > 1$ conditional on state $l \in \{1, 2\}$ at $t - 1$. Note that the regular so-called discrete-time (DT-)LMM assumes the intervals between measurements, δ_{ti}, to be equal. In contrast, the intervals can differ across timepoints and individuals in the so-called continuous-time (CT-)LMM (Böckenholt, 2005; Jackson & Sharples, 2002; Vogelsmeier, Vermunt, Böing-Messing, & De Roover, 2019). Because differences in intervals are more realistic in EMA (and the CT-LMM generalizes to the DT-LMM if intervals are equal), in this article, only the CT-LMM is applied and described (for a detailed description of the DT-LMM, we refer to Vogelsmeier, Vermunt, van Roekel, and De Roover (2019).

The transition probabilities $p_{\delta_{ti},lk} = p_{\delta_{ti}}(s_{itk} = 1|s_{it-1,l} = 1, \mathbf{z}_{it})$ are collected in the 2×2 matrix $\mathbf{P}_{\delta_{ti}}$, where the row sums of $\mathbf{P}_{\delta_{ti}}$, $\sum_{k=1}^{2} p_{\delta_{ti},lk}$, are equal to 1. In the CT-LMM, the transition probabilities $\mathbf{P}_{\delta_{ti}}$ are a function of the interval δ_{ti} and the “transition intensity matrix” $\mathbf{Q}$. The 2×2 matrix Q contains the transition intensities (or rates) q_{lk} defining the transitions from the origin state l to the destination state k per a very small time unit. The intensities for the two off-diagonal elements in the matrix $\mathbf{Q}$ (i.e., $k \neq l$) are

$$q_{lk} = \lim_{\delta \to 0} \left(\frac{p(s_{itk} = 1|s_{it-\delta,l} = 1, \mathbf{z}_{it})}{\delta} \right) \tag{7}$$

and the two diagonal elements are equal to $-\sum_{k \neq l} q_{lk}$ (Cox & Miller, 1965). Taking the matrix exponential of $\mathbf{Q} \times \delta_{ti}$ results in the transition probabilities $\mathbf{P}_{\delta_{ti}}$, implying that the probability of transitioning to the respective other state instead of staying in either the attentive or C/IER state at two consecutive measurement occasions (i.e., $k \neq l$) is more likely for longer intervals. As can be seen from Equation (7), like the initial state probabilities, the transition intensities (and, in turn, the transition probabilities) can depend on covariates, $\mathbf{z}_{it}$. Typically, a log-linear model for the transition intensities is used (again for $k \neq l$):

$$\log q_{lk} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk}\mathbf{z}_{it} \tag{8}$$

with γ_{0lk} as transition intercepts and $\boldsymbol{\gamma}'_{lk} = (\gamma_{lk,Z_{i11}}, \ldots, \gamma_{lk,Z_{i1U}})'$ as transition slopes that quantify the covariate effects on transitioning compared to staying.

Estimation

The parameters of the proposed model are obtained via maximum likelihood estimation. We chose a step-wise estimation approach as recommended by Vogelsmeier, Vermunt, Bülow, and De Roover (2021). The three steps are (1) estimating the state-specific measurement models, (2) assigning observations to the most probable state based on the posterior state-membership probabilities and calculating classification error inherent to the modal assignment, and (3) evaluating the transition model (including covariate effects). In this third step, the state-specific measurement model parameter estimates from step 1 remain fixed, and the classification error from step 2 is taken into account. To assess which covariates are significantly related to the transition model parameters (and, thus, whether they should be included in the model), one may employ backward selection utilizing Wald tests (Agresti, 1990), which is also the chosen approach for this study. For technical details of this three-step approach (including the details about how the classification errors are computed and accounted for), we refer to Vogelsmeier, Vermunt, Bülow, and De Roover (2021).

The step-wise approach is especially favorable when investigating covariate effects on transitions between states, as only the transition model requires re-estimation (i.e., step 3) when adding or removing covariates, while the estimates for the measurement models remain the same. Additionally, the step-wise estimation can be performed using the open-source software R (R Core Team, 2021), while the alternative one-step full information maximum likelihood estimation (e.g., Vogelsmeier, Vermunt, Böing-Messing, and De Roover, 2019) requires proprietary software like Latent GOLD (Vermunt & Magidson, 2021), which we perceive as a notable drawback for moving open science forward. More specifically, step 1 can be performed using the R package ‘mirt’ (Chalmers, 2012), step 2 can be manually coded, and step 3 can be conducted utilizing the R package ‘lmfa’ (Vogelsmeier & De Roover, 2021). The syntax for the following empirical application is provided at OSF (https://osf.io/uj6sr/).

Empirical Application

The EMA data that we use to illustrate the proposed methodology stem from a previous study, which presented novel strategies for assessing the psychometric properties of EMA questionnaires using a "momentary satisfaction with life" scale (Rein et al., 2022). We applied the novel model-based C/IER detection approach to get insight into (1) how many and which observations are classified into the C/IER and attentive latent states, respectively, (2) how individuals transition between the latent states over time, and (3) potential situation-/participant-level predictors of transitions among the latent states. Note that no analyses or hypotheses were pre-registered beforehand. The primary objective of this application is to showcase the implementation and interpretation of the novel methodology, as well as to explore potential predictors associated with transitioning to the latent C/IER state. Therefore, any conclusions should be validated in future research. The secondary data analysis was approved by the School of Social and Behavioral Sciences Ethics Review Board of Tilburg University, The Netherlands (reference number: RP_FT29). In the following, we describe the study-design characteristics, participant information, and measures most relevant to this empirical application. For a complete description, we refer to the original study.

Study Design and Participants

Participation in the study took 15 days. The study consisted of three parts: an introductory survey on day 1, a 14-day EMA on days 2 to 15, and a final survey on the evening of day 15. The study was implemented in m-path (www.m-path.io). Upon signing up, participants received an information letter and gave their informed consent. As the EMA could pose a burden for the participants, they were informed that participation was voluntary and that they could withdraw from the study at any time. As a token of appreciation for participating in the study, participants were offered the opportunity to receive personalized feedback and to participate in a raffle for Amazon vouchers. It is important to highlight that the participants' chances to win a voucher was higher if they filled in more questionnaires (i.e., with a higher compliance). This may have incentivized participants to engage in C/IER rather than to omit questionnaires when they were not fully motivated at a given measurement occasion.

After consenting to participate in the study, the participants completed the introductory survey, which comprised demographic variables and questions concerning the study participation (e.g., whether participants wished to receive personalized feedback). On the next morning, the EMA began, which employed a signal-contingent sampling design (Scollon et al., 2009) with six prompts per day (i.e., up to 84 EMA surveys per participant in total). Participants were notified once randomly within two-hour blocks (e.g., the first notification was sent between 8 am and 10 am, the second between 10 am and 12 pm, and so on). These surveys comprised items on momentary satisfaction and, on some days, instructional manipulation check items. To reduce the time needed for filling in the surveys and thus reduce participant burden, a planned-missingness design (Silvia et al., 2014) was used. The median time required to finish the EMA surveys was 37 seconds. On the evening of the 15th day, participants were asked to fill out the final survey, which concluded their participation in the study. The final survey assessed the participants' study experience.

The data comprised 71 participants between 18 and 49 years old ($M = 25.1$, $SD = 7.38$), of which 24 identified as male and 46 as female (one did not disclose their gender). On average, participants filled in 61.1 out of 84 EMA surveys ($\approx$ 72.7% compliance rate, 4,335 observations total).

Measures

In the following, we describe two types of measures: time-varying (i.e., situation-level) measures from the EMA surveys and time-constant (i.e., person-level) measures from the final survey.

EMA surveys. Participants received seven out of ten items measuring the unidimensional construct “momentary satisfaction” at every measurement occasion. Specifically, participants were shown a consistent anchor item along with a randomized selection of six items from the remaining nine. For each item, the participants evaluated their answer on a Likert scale ranging from 0 = “strongly disagree” to 6 = “strongly agree”. The scale was shown to have a high reliability across both time when averaging across persons ($r = .91$) and across persons when averaging across time ($r = .99$). The items, along with their abbreviations, are provided in Table 1. In the following, these ten items are referred to as the *content items*.

Table 1
Momentary Satisfaction Questionnaire

Item	Abbreviation
1. In this moment, I consider myself happy.	consider_happy
2. My current activity makes me satisfied with life.	satisfied_life
3. I would like to change many things about my current situation.*	change_many
4. In this moment, I am experiencing life close to my ideal.	life_ideal
5. My current activity makes me happy.	makes happy
6. All things considered, I am satisfied with my current situation.	satisfied_situation
7. If I could relive this moment, I would change nothing.	change_nothing
8. My current activity leaves a lot to be desired.*	leaves_desired
9. I get the important things that I want in life from my current situation.	important_things
10. This moment is in line with how I want my life to be.	line_life

Note: * denotes negatively worded items. Item 10 is the anchor item presented at every measurement occasion. Instructions for the scale are: “Please indicate how much the statements describe your feelings and experiences in the moment right before you received this notification. Please answer honestly and spontaneously. There are no correct or incorrect answers.”

Additionally, participants received instructional manipulation check items at specific measurement occasions to monitor signs of careless responding. Each participant was shown one such item on days 4, 7, 10, and 13 (i.e. at four of 84 measurement occasions). More specifically, the participants were asked to select one specific answer option, for example, "Please select ‘disagree’ for this question." For this application, 18 participants who specified the wrong answer at least once during participation were considered careless. The other 53 participants were considered attentive. In the following, this dichotomous participant-level measure is referred to as *CR_check* (careless responders according to check) with the two categories 1 = “yes” and 0 = “no”.

In addition to the questionnaire-based measures, two situation-level measures were considered for each measurement occasion: First, the number of the current observation, *n_obs*, was computed for each participant and occasion, counting only the completed questionnaires while excluding skipped ones. This metric ranges from 1 to the total number of observations recorded per participant. Second, the hour of the day, *hour_day*, was derived from the submission time of the questionnaires. This metric ranges from 8 to 21, as questionnaires were exclusively sent during this timeframe.

Final Survey. After the final EMA questionnaire, participants were asked two questions about their study experience: (1) “If you were to participate in a similar study again: Would you prefer a different study length (fewer or more days)?” and (2) “If you were to participate in a similar study again: Would you prefer a different number of surveys per day?”. For both questions, participants were given the answer options 1 = “yes, fewer [days/surveys per day]”, 2 = “yes, more [days/surveys per day]”, and 3 = “no”. For this application, participants who indicated option 1 were considered overly burdened by the study length and/or the number of surveys per day. This applied to 51 and 34 participants, respectively. Participants who indicated options 2 or 3 were not considered overly burdened, which applied to the remaining 16 and 33 participants, respectively. In the following, these two participant-level predictors are referred to as *burdened_sl* (burdened by the study length) and *burdened_sf* (burdened by sampling frequency) with the two categories 1 = “yes” and 0 = “no”, respectively.

In summary, the data comprised ten content items, three participant-level measures (*CR_check*, *burdened_sl*, and *burdened_sf*), and two situation-level measures (*n_obs* and *hour_day*). The content items were utilized as indicators of the unidimensional construct “momentary satisfaction” in step 1. All other measures were considered potential predictors of latent state transitions in step 3.[2]

Analysis Strategy

Our analysis adhered to the three consecutive steps outlined in the Estimation Section. In step 1, we analyzed the raw content item response data with the confirmatory mixture IRT model for identifying C/IER described above. We assumed a unidimensional model to hold for attentive item responses. We ran the model with 50 sets of random starting values. The solution with the smallest log-likelihood was replicated in roughly 40% of replications. Next, in step 2, the observations were assigned to either the attentive or the C/IER state based on their most probable state membership. Furthermore, the classification error was calculated. Finally, in step 3, the transition model was estimated to investigate the transitions between the attentive and the C/IER states, and the effects of five predictors on the transition probabilities were explored. From the variables available in the data, we chose the three participant-level measures *CR_check*, *burdened_sl*, and *burdened_sf*, and the two situation-level measures *n_obs* and *hour_day* as potential predictors of transitioning to the C/IER state. Firstly, instructional manipulation check items can act as a preliminary gauge of participants' overall attentiveness (Curran, 2016; Meade & Craig, 2012). Therefore, we expected that participants identified as careless through these checks ($CR\_check = 1$) show higher probabilities of transitioning to and remaining in the C/IER state than participants identified as attentive ($CR\_check = 0$).

Secondly, participant burden is expected to reduce attentiveness (Hasselhorn et al., 2023; Ulitzsch et al., 2024). The data for this application encompass two distinct forms of burden: one related to the duration of the study and the other associated with the sampling frequency. While the relation between study length and C/IER remains largely unexplored in previous studies, researchers experimentally manipulated sampling

[2] Note that the data contained additional measures. These are, however, not relevant to this application and thus not further described. The interested reader is referred to the publicly-available codebook: https://osf.io/ab579.

frequencies in EMA studies and did not find significant effects on C/IER (Eisele et al., 2022; Hasselhorn et al., 2023). Nevertheless, in this application, we included both perceived burden due to study length and due to sampling frequency as predictors of transition probabilities. Assessing (predictors of) C/IER is still a newly emerging stream of research, and it remains important to investigate the effects of design choices in different studies, especially when using differing measures of sampling frequency (burden) and novel methodologies to identify C/IER. We expected participants who were burdened by either the study length ($burdened\_sl = 1$) or the sampling frequency ($burdened\_sf = 1$) to have higher probabilities of transitioning to and staying in the C/IER state than participants who were not burdened ($burdened\_sl = 0$ and $burdened\_sf = 0$).

Finally, attentiveness can generally decrease over time—throughout the course of participation in general but also within a day in particular—because of getting bored from repeatedly completing the same questionnaires (i.e., fatigue effects; Eisele et al., 2023). Attentiveness within a day may additionally decline because of reduced cognitive skills throughout the day (Schmidt et al., 2007; West et al., 2022). Previous research indeed indicated a decline in attentiveness with study duration (Denison, 2022; Ulitzsch et al., 2024), but discernible within-day trends did not emerge (Ulitzsch et al., 2024). However, the effect of daytime on C/IER is still largely unexplored in EMA research. In this application, we included *hour_day* and *n_obs* as predictors of the transition probabilities, where the latter served as a proxy for elapsed duration of participation. We expected higher probabilities of transitioning to and staying in the C/IER state for larger values on both predictors than for lower ones.

For the predictor selection, we used backward selection; that is, we started with a transition model including all predictors and removed them one by one until only those with significant Wald-test statistics were in the model.[3] After selecting the model, we first

[3] As explained in Vogelsmeier, Vermunt, and De Roover (2023), different selection procedures can lead to different results. Therefore, we performed a sensitivity analysis with forward selection; that is, we started with one predictor at a time and stopped as soon as the added predictors were no longer significant.

examined the transition probabilities for a one-hour interval[4] and predictor values corresponding to their sample means to obtain an overall impression of the stability of state memberships. Subsequently, for each predictor, we compared the transition probabilities for a one-hour interval for the lowest score in the sample to the probabilities for the highest score, keeping the predictor values for the other two predictors equal to their sample means.[5] For the categorical predictors, the lowest and highest values simply refer to the meanings of the values 0 and 1.

The analyses were performed in R (R Core Team, 2021, R version 4.2.3). Step 1 was performed using the R package 'mirt' version 1.41 (Chalmers, 2012), step 2 was manually coded, and step 3 was conducted utilizing the R package 'lmfa' version 0.1.3 (Vogelsmeier & De Roover, 2021).[6] The data and the code for the data analysis are publicly available at OSF (https://osf.io/uj6sr/).

Results

Measurement Model. Approximately eight percent of the observations were assigned to the C/IER state. The classification-error probabilities were close to zero for the attentive state (0.01) but considerably higher for the C/IER state (0.29). This implies that the classification of observations assigned to the C/IER state was relatively less certain compared to the classification of observations assigned to the attentive state.

Figures 2 and 3 display item characteristic curves (ICCs) for the attentive and

Results were the same for forward and backward selection. In the following, we report results obtained with backward selection only, but provide code for both procedures in the analysis script.

[4] Note that researchers are free to choose other intervals in their applications. For instance, when dealing with a sampling design with less frequent measurement occasions, larger intervals would make sense.

[5] Note that interpreting transition probabilities for a specific interval and comparing them for different predictor values is the suggested approach by Vogelsmeier, Vermunt, and De Roover (2023) because the estimates of the CT-LMM parameters—that is, the (effects on) transition intensities—are hardly interpretable.

[6] It is important to note that including participant-level predictors for the transition probabilities in the latent Markov model is only possible by repeating the participant-level scores at each of the participant's measurement occasions. The limitations of this approach are detailed in the Discussion Section.

C/IER states. Recall that for the C/IER state, item parameters are constrained to be the same across items and, hence, only a single ICC is obtained. As was to be expected, in the attentive state, for the positively worded items, the probability of choosing a higher response category increased with a higher location on the momentary satisfaction trait but decreased for negatively worded items. In the C/IER state, category probabilities for the lower two (0 and 1) and the highest category (6) were essentially zero across the latent category preference continuum. Instead, when being in the C/IER state, participants tended to favor one of the upper middle categories (categories 3 to 5), depending on their location on the category preference continuum. This pattern mirrors previous findings (Ulitzsch, Pohl, et al., 2022; Ulitzsch, Yildirim-Erbasli, et al., 2022) and is in line with cognitive theories on edge aversion in decision-making processes when items do not need to be (or, as in the present case, are not) processed (Bar-Hillel, 2015).

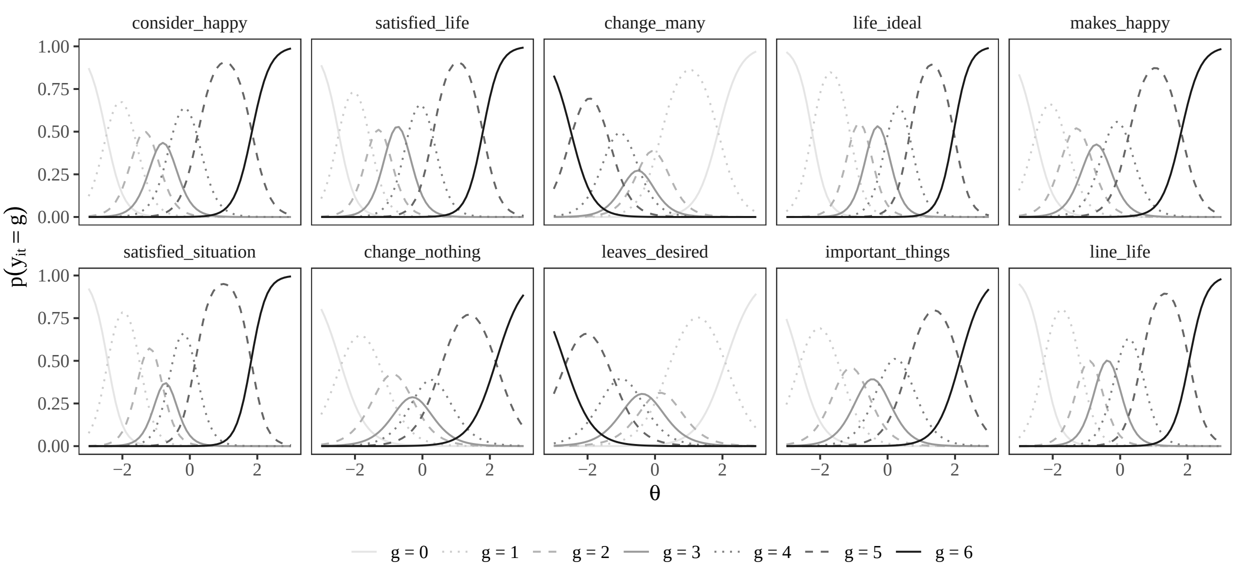


Figure 2. Item characteristic curves in the attentive state for the 10 content items. θ: momentary satisfaction

Transition Model. Starting from the model with all predictors included, we first removed *burdened_sl* and then *n_obs* because the Wald-test statistics were non-significant. Thus, contrary to our expectations, these predictors were unrelated to the transitions

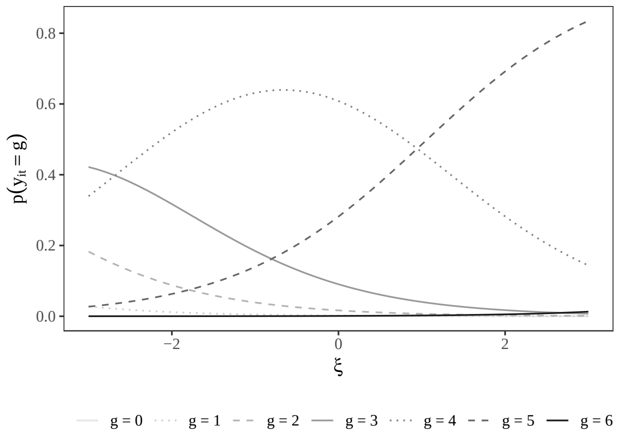


Figure 3. Item characteristic curve for the C/IER state. ξ: category preference

between the attentive and the C/IER state. In line with our expectations, the remaining three predictors were significantly related to the transition probabilities (*CR_check*: $\chi^2(2) = 27.45$, $p < .001$; *hour_day*: $\chi^2(2) = 6.57$, $p = 0.04$; *burdened_sf*: $\chi^2(2) = 12.75$, $p < .001$).

Investigating the transition probabilities for a one-hour interval and predictor values corresponding to their sample means (i.e., 0.29 for *CR_check*, 13.81 for *hour_day*, and 0.51 for *burdened_sf*) indicated that the staying probabilities were equal to .98 and .82 for the attentive and C/IER states, respectively. Hence, the probabilities of transitioning from the attentive to the C/IER state and in the reverse direction were .02 and .18., respectively. Therefore, staying in a state was generally more probable than transitioning away from it. However, the stability was more pronounced in the attentive state. In line with this, 15 of the 71 individuals consistently remained in the attentive state. None of the participants consistently stayed in the C/IER state.

The comparisons of the transition probabilities for the lowest and highest predictor values are shown in Table 2 and the results for each predictor are described below.

CR_check. Comparing the transition probabilities for participants who were flagged as attentive according to the instructional manipulation check items to the probabilities for participants who were flagged as careless showed that, in line with our

Table 2

Transition probabilities for low and high values on the three predictors while holding all other predictor values equal to their sample means

Predictor (low value → high value)	From/To	Attentive		C/IER	
CR_check (attentive → careless)	Attentive	0.99 → 0.96	(−)	0.01 → 0.04	(+)
	C/IER	0.18 → 0.18	( )	0.82 → 0.82	( )
hour_day (8 h → 21 h)	Attentive	0.99 → 0.96	(−)	0.01 → 0.04	(+)
	C/IER	0.10 → 0.35	(+)	0.90 → 0.65	(−)
burdened_sf (no → yes)	Attentive	0.96 → 0.99	(+)	0.04 → 0.01	(−)
	C/IER	0.37 → 0.09	(−)	0.63 → 0.91	(+)

Notes: The interval length for which the transition probabilities were compared was equal to one hour. The value on the left of the arrow represents the transition probability associated with the lowest (categorical or continuous) predictor value in the sample while holding all other predictor scores equal to their sample means; the value on the right of the arrow indicates the transition probability associated with the highest predictor value in the sample while holding all other predictor scores equal to their sample means; the (meanings of) the lowest and highest values for each predictor are provided in the first column. The sign in parentheses indicates the increment/decrement between the two values; the parentheses are empty if the value does not differ. The cell color is green if the transition probability rises with an increase in the predictor value, red if it decreases, and white if it stays the same. The strength of the color indicates the magnitude of the change.

expectations, the probability of transitioning to the C/IER state was slightly higher for the careless participants (0.04) than for the attentive participants (0.01). However, the probabilities of staying in the C/IER state were the same for careless and attentive participants (0.82). Overall, this indicates that the predictor is practically rather uninformative for the probabilities of participants being in the C/IER state, although it was significantly related to the transition probabilities. This is likely the result of an artificially inflated sample size for the participant-level predictors, which is explained in more detail in the Discussion Section.

Hour_day. The comparison of the transition probabilities when completing the EMA questionnaires at 8 am with the probabilities when completing the questionnaires at 9 pm showed that the probabilities of transitioning to the C/IER state were only slightly higher in the evening (0.04) than in the morning (0.01) and that the probabilities of remaining in the C/IER state were considerably lower in the evening (0.65) than in the morning (0.90). Contrary to our expectation, this indicates that it is generally less likely to be in the C/IER state in the evening than in the morning.

Burdened_sf. Comparing the transition probabilities for participants who were not burdened by the sampling frequency with the probabilities of participants who were burdened, we found that the probability of transitioning to the C/IER state was slightly lower for burdened participants (0.01) than for non-burdened participants (0.04), but, more importantly, that the probability of remaining in the C/IER state was significantly higher for burdened participants (0.91) than for non-burdened participants (0.63). Overall, this shows that, in line with our expectations, experienced burden regarding the sampling frequency is related to a higher probability of being in the C/IER state.

Conclusions

Approximately eight percent of observations were classified as stemming from C/IER. This is a noteworthy amount considering that the data were collected using a planned-missing design to reduce participant burden by shortening the questionnaire because a shorter questionnaire length was shown to be associated with more attentive responding than long questionnaires in EMA (Eisele et al., 2022). This highlights the need to model and understand C/IER in EMA data, even under seemingly favorable study designs.

The overall within-person stability of attentive responding and C/IER was high. However, situation-level predictors altered the probabilities of engaging in both types of responding. Specifically, the perceived burden due to sampling frequency and participation at earlier times of the day were related to a higher probability of C/IER. The latter finding did not align with our initial expectations. A possible explanation is that the hour of the day serves as more than just a timestamp; it might also reflect the location of the participants and the time available to complete the EMA attentively. Considering the overall young age of the participants, it is plausible that they have greater mental capacity to respond attentively in the later hours of the day, after their study or work commitments have ended. However, further research, including experimental manipulation, is needed to better understand the effects of the time of the day on C/IER in EMA. Likewise, future research may explore more complex (e.g., quadratic) relationships between time of the day and C/IER.

Evaluating Robustness Against Violations of Measurement Invariance

The proposed method rests on the assumption that the attentive and C/IER mixture component model parameters generalize across all observations. For trustworthy C/IER detection, we believe that this assumption is less of an issue for the C/IER component model. The C/IER component model by Uglanova et al. (in preparation) is well capable of handling multiple types of C/IER behavior across observations, occurring, for instance, when some respondents select their careless responses uniform randomly, while others tend to opt for some specific categories. Unaccounted violations of measurement invariance for the attentive component model across individuals and/or time, however, may entail artifactual conclusions on C/IER rates. This is especially problematic because such violations are plausible to occur (Adolf et al., 2014; Brose et al., 2015; McNeish et al., 2021; Vogelsmeier, Cloos, et al., 2023; Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019): Because participants fill out their questionnaires in various contexts and situations, the structure of measurement may vary across the course of an EMA study.

To determine the degree of violations that still allows obtaining trustworthy conclusions on C/IER occurrence, we conducted two simulation studies. Study I evaluated the robustness of C/IER detection against unaccounted heterogeneity in attentive loading patterns across observations. Thereby, Study I mimics a scenario of reprioritization in attentive responding (Oort et al., 2005), where single items change in their importance for the measured constructs.

Study II investigated the consequences of unaccounted heterogeneity in factor structure across observations. Such heterogeneity occurs when the nature of the measured constructs changes across the course of the EMA. For instance, when responding to items assessing affective experiences, respondents may shift from a valence to an arousal focus in some contexts, causing the underlying factor structure to shift from positive affect and negative affect factors to high and low arousal factors (Feldman, 1995). Likewise, dampened emotional granularity (i.e., lower differentiation between emotions) can be assumed to result in increased correlations among emotional facets (Krone et al., 2018), with the most extreme scenario being that emotional facets collapse into a single factor for observations with low granularity.

In our evaluations, we focused on the trustworthiness of C/IER detection in step 1 of the proposed approach. This is a key prerequisite for accurate conclusions drawn based on the subsequent steps 2 and 3, which serve to get a better understanding of respondent and occasion characteristics associated with the occurrence of (presumed) C/IER.

Methods

To evaluate C/IER detection under realistic research conditions, we mimicked the EMA study of the empirical application reported above. We generated data for $I = 75$ respondents being administered $J = 10$ 7-point Likert-scale items at $T = 60$ measurement occasions. Data for each observation (i.e., respondent-by-occasion interaction) were generated independently.[7]

In all simulation conditions, 10% of the observations were simulated to stem from C/IER. To simulate C/IE responses, we followed recommendations by Curran and Denison (2019) and simulated different C/IE response patterns. Doing so allows us to showcase that the proposed approach can indeed deal with the simultaneous occurrence of different C/IE response patterns. C/IER observations were randomly assigned to two groups: For the first group of purely random responders, C/IE responses were generated uniform randomly. For

[7] Recall that in the simulation study, we were interested in the quality of parameter estimates in step 1. For this, autocorrelation between factor scores or the Markov structure are irrelevant (Vogelsmeier, Vermunt, Bülow, & De Roover, 2021).

the second group of responders with category preferences, C/IE responses were generated according to the constrained GRM component model given in Equation 3, with loadings and thresholds set to the values obtained in the empirical application (see Table A1). C/IER category preference traits were drawn from a standard normal distribution.[8]

We simulated attentive responses according to a uni-dimensional GRM, employing the loadings and thresholds obtained for the attentive component model in the empirical application (see Table A2) and drawing respondents' traits from a standard normal distribution. In the baseline condition, all attentive item responses were generated from the same model. To study robustness, the attentive measurement model was manipulated for some observations, inducing heterogeneity in loading patterns (Study I) or factor structures (Study II).

For each simulation condition, we generated 100 data sets. In line with step 1 of the proposed approach, each data set was analyzed with the confirmatory mixture IRT model proposed by Uglanova et al. (in preparation, Equations 1 and 3) treating all repeated observations as independent, using the R package 'mirt' version 1.41 (Chalmers, 2012).

As evaluation criteria, we considered convergence, bias and variability of C/IER rate estimates as well as sensitivity (i.e., the proportion of correctly identified C/IER observations out of all observations classified as C/IER) and specificity (i.e., the proportion of correctly identified attentive observations out of all observations classified as attentive) of observation-level C/IER classification using modal assignment. Further, we inspected mean posterior C/IER state probabilities for true positive, true negative, false positive, and false negative C/IER observations to gauge the classification error for these types of observations. All analyses were conducted using R (R Core Team, 2021, R version 4.2.3).

Study I: Heterogeneity in loading patterns. In Study I, we evaluated the effects of changing loading patterns, while the overall factor structure remained intact across attentive respondents and measurement occasions. We randomly selected a varying

[8] For evaluations of other C/IER patterns, we refer to Uglanova et al. (in preparation).

number of items $(2; 5; 10)$ that were affected by an exponential decay in their loadings across measurement occasions, assuming that more change occurs at the beginning of the study.[9] Then, for the affected items, we gradually decreased loadings across measurement occasions as $\alpha_{jt} = \alpha_{j1} \cdot d^{(t-1)}$, varying the severity of exponential decay d $(0.99; 0.98; 0.95)$. Figure 4 illustrates the trajectory of three item loadings with $\alpha_{j1} = 2$ and exponential decay of $d = 0.99$, $d = 0.98$, and 0.95 across $T = 60$ measurement occasions. As can be seen, for $d = 0.99$, original item loadings were approximately halved by the last measurement occasion. In the most extreme condition of $d = 0.95$, item loadings of affected items were essentially zero at the last measurement occasion.

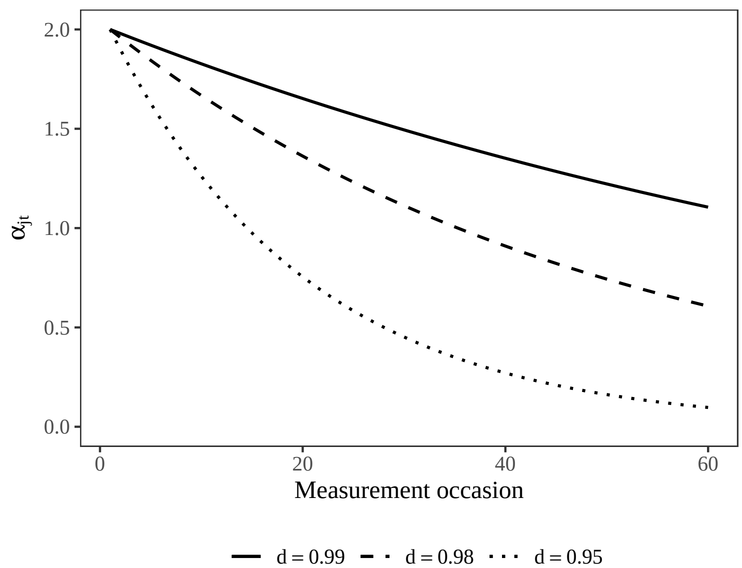


Figure 4. Change in item loadings across measurement occasions for different severity of exponential decay d.

Study II: Heterogeneity in factor structure. In Study II, we evaluated the effect of changes in attentive factor structure across the EMA. We randomly selected a varying proportion of observations $(0.20; 0.30; 0.50)$ for which a two- instead of a uni-dimensional attentive measurement model with varying latent correlation ρ

[9] Having loadings decrease according to an exponential decay process is a somewhat arbitrary choice. Note, however, that because the order of observations is not taken into account in step 1, the specific form of changes in loading patterns will not impact results. What matters is the proportion of observations with specific loading patterns.

$(0.80; 0.60; 0.40)$ was simulated. For the two-dimensional attentive measurement model, the first five items measured the first and the remaining items the second factor. Data-generating values for loadings and thresholds were the same as for the uni-dimensional model.

Results

Study I: Heterogeneity in loading patterns. Table 3 displays convergence rates and quality of C/IER identification for Study I. Conditions with extreme exponential decay ($d = 0.95$) challenged the estimation procedure, resulting in increasing non-convergence rates with an increasing number of affected items. When all $J = 10$ items were affected, only 83% of the replications converged. Note, however, that this condition is rather extreme and, arguably, not realistic, as it resulted in attentive loadings of essentially zero for all items by the end of the EMA (i.e., none of the items was reflective of the to-be-measured trait). In conditions with less extreme exponential decay, all replications converged. For converged replications, the quality of C/IER detection resembled those of the baseline condition without violations of measurement invariance. Bias and variability of C/IER proportion estimates were low, while sensitivity and specificity were high.

Mean posterior C/IER state probabilities for true positive (i.e., true C/IER observations classified as C/IER with modal assignment) and true negative (i.e., truly attentive observations classified as attentive) C/IER observations were close to 1 and 0 respectively, indicating very low classification error for these observations. For false positive (i.e., truly attentive observations classified as C/IER) and false negative (i.e., true C/IER observations classified as attentive) C/IER observations, mean posterior state probabilities were further away from 1 and 0, respectively, indicating that for wrongly classified observations, classification error is higher. From these results, we conclude that C/IER detection exhibits robustness against the studied types of heterogeneity in attentive loading patterns.

Table 3

Convergence rates and quality of C/IER identification for different conditions of heterogeneity in attentive loading patterns

d	J_d	Conv	Bias π_2	SD π_2	Sens	Spec	$\bar{\pi_2}^{TP}$	$\bar{\pi_2}^{TN}$	$\bar{\pi_2}^{FP}$	$\bar{\pi_2}^{FN}$
Baseline		1.00	-0.002	0.005	0.992	0.975	0.978	0.005	0.721	0.173
	2	1.00	-0.001	0.004	0.992	0.975	0.977	0.005	0.714	0.177
0.99	5	1.00	-0.003	0.005	0.992	0.976	0.977	0.005	0.719	0.180
	10	1.00	-0.002	0.005	0.992	0.975	0.978	0.005	0.716	0.179
	2	1.00	-0.002	0.005	0.992	0.975	0.977	0.005	0.724	0.181
0.98	5	1.00	-0.001	0.005	0.992	0.973	0.975	0.005	0.717	0.182
	10	1.00	-0.002	0.004	0.992	0.977	0.977	0.005	0.719	0.176
	2	1.00	-0.003	0.005	0.992	0.975	0.976	0.005	0.718	0.181
0.95	5	0.97	0.002	0.005	0.991	0.935	0.977	0.006	0.795	0.179
	10	0.83	-0.000	0.004	0.993	0.962	0.979	0.005	0.799	0.176

Notes: d: exponential decay; J_d: number of affected items; Bias π_2: bias of the C/IER mixture proportion estimate; SD π_2: standard deviation of the C/IER mixture proportion estimate across replications; Sens: sensitivity of observation-level C/IER classification using modal assignment; Spec: specificity of observation-level C/IER classification using modal assignment; $\bar{\pi_2}^{TP}$: mean posterior C/IER state probability for true C/IER observations classified as C/IER; $\bar{\pi_2}^{TN}$: mean posterior C/IER state probability for truly attentive observations classified as attentive; $\bar{\pi_2}^{FP}$: mean posterior C/IER state probability for truly attentive observations classified as C/IER; $\bar{\pi_2}^{FN}$: mean posterior C/IER state probability for true C/IER observations classified as attentive.

Study II: Heterogeneity in factor structure. Table 4 displays convergence rates and quality of C/IER identification for Study II. No convergence issues were encountered in Study II. However, there was an increasing upwards bias in C/IER proportion estimates with an increasing degree of violation of measurement invariance, i.e., with an increasing proportion of observations for which a two- instead of the modeled uni-dimensional attentive measurement model held as well as with a decreasing correlation ρ between the two factors. For highly correlated factors, bias remained minimal, with bias for $\rho = 0.80$ remaining below 1%. For the most extreme condition with 50% of attentive observations being affected and $\rho = 0.40$, however, bias was as high as 0.06. That is, researchers would conclude that 16% instead of the data-generating 10% of observations stem from C/IER. The increasing upwards bias was accompanied by decreasing specificity. Further, mean posterior C/IER state probabilities for false positive C/IER observations approached 1, indicating that for observations wrongly classified as inattentive, the classification error tended to reflect misclassification to a lesser extent than in Study I. From these results, we conclude that the approach exhibits robustness only to mild heterogeneity in factor structures.

Table 4

Convergence rates and quality of C/IER identification for different conditions of heterogeneity in attentive factor structure

ρ	Obs	Conv	Bias π_2	SD π_2	Sens	Spec	$\bar{\pi}_2{}^{\mathrm{TP}}$	$\bar{\pi}_2{}^{\mathrm{TN}}$	$\bar{\pi}_2{}^{\mathrm{FP}}$	$\bar{\pi}_2{}^{\mathrm{FN}}$
Baseline		1.00	-0.002	0.004	0.992	0.975	0.978	0.005	0.723	0.173
0.80	0.20	1.00	0.003	0.005	0.992	0.930	0.975	0.007	0.791	0.181
	0.30	1.00	0.005	0.005	0.991	0.911	0.975	0.008	0.792	0.177
	0.50	1.00	0.008	0.005	0.990	0.886	0.971	0.010	0.794	0.183
0.60	0.20	1.00	0.015	0.005	0.992	0.834	0.976	0.009	0.849	0.185
	0.30	1.00	0.022	0.005	0.991	0.788	0.974	0.011	0.851	0.190
	0.50	1.00	0.034	0.006	0.990	0.718	0.970	0.015	0.845	0.193
0.40	0.20	1.00	0.027	0.005	0.992	0.754	0.976	0.010	0.880	0.187
	0.30	1.00	0.040	0.006	0.992	0.684	0.975	0.013	0.880	0.182
	0.50	1.00	0.062	0.007	0.991	0.595	0.970	0.019	0.877	0.195

Notes: ρ: correlation of latent factors; Obs: proportion of affected observations; Bias π_2: bias of the C/IER mixture proportion estimate; SD π_2: standard deviation of the C/IER mixture proportion estimate across replications; Sens: sensitivity of observation-level C/IER classification using modal assignment; Spec: specificity of observation-level C/IER classification using modal assignment; $\bar{\pi}_2{}^{\mathrm{TP}}$: mean posterior C/IER state probability for true C/IER observations classified as C/IER; $\bar{\pi}_2{}^{\mathrm{TN}}$: mean posterior C/IER state probability for truly attentive observations classified as attentive; $\bar{\pi}_2{}^{\mathrm{FP}}$: mean posterior C/IER state probability for truly attentive observations classified as C/IER; $\bar{\pi}_2{}^{\mathrm{FN}}$: mean posterior C/IER state probability for true C/IER observations classified as attentive.

Discussion

In this study, we introduced a novel method to distinguish between attentive responding and C/IER in EMA while uncovering contextual correlates of C/IER. Our confirmatory mixture modeling approach is suitable for scales with multiple indicators assessed via ordered response categories. It maximizes the leverage of item information by identifying C/IER based on pre-defined measurement models for attentive and C/IER responses. This eliminates the need for possibly ambiguous post-hoc interpretation of latent states resulting from exploratory mixture modeling. As previous mixture modeling approaches for C/IER in EMA data (Hasselhorn et al., 2023; Ulitzsch et al., 2024), it does not require decisions on threshold settings and takes C/IER classification uncertainty account. Further, because C/IER is identified using content item responses, no attention check items need to be administered, resulting in shorter questionnaires. Nevertheless, it should be noted that the administered scales should contain negatively worded content items to facilitate trustworthy identification (Uglanova et al., in preparation). Our empirical findings underscore the efficacy of this novel approach in both pinpointing C/IER instances in EMA and gaining insights into their underlying causes. Understanding these reasons is crucial for mitigating C/IER in future studies and thus for optimizing research outcomes. For instance, incorporating additional motivational reminders in situations prone to C/IER occurrence could enhance attentive responding and in turn data quality.

We evaluated the method's robustness against the unaccounted presence of attentive measurement model changes in two simulation studies. We found the approach to be robust against unaccounted heterogeneity in attentive loading patterns across observations, but not against unaccounted heterogeneity in the factor structure underlying attentive responses.

Limitations and Future Directions

Future research may explore methods to accommodate changes in the attentive measurement model. If hypotheses on the type of measurement model changes exist, these can be translated into additional confirmatory measurement models. For instance, researchers may specify multiple attentive states between which correlations among latent factors are allowed to vary to accommodate hypothesized variations in emotional granularity, or they could specify attentive measurement models with positive and negative affect factors and high and low arousal factors, respectively. If researchers assume that the perception of a specific item may be context-dependent, component models with different loading structures could be formulated. For instance, depending on the context, reporting excitement can be reflective of both positive (when associated with enthusiasm) and negative (when associated with nervousness) affect. Likewise, researchers may easily accommodate more complex types of attentive response behavior. Response styles (i.e., idiosyncrasies in how respondents use rating scales), for instance, recently gained increased attention in EMA research (Deng et al., 2018). These can easily be considered by using an attentive component model that takes response styles into account (see Ulitzsch, Pohl, et al., 2023, for a mixture IRT model accommodating both response styles and C/IER). If no such hypotheses exist, researchers may combine the confirmatory C/IER component model with exploratory attentive measurement models. Such an approach may filter out observations exhibiting C/IER and explore measurement model changes in the remaining attentive observations. Note, however, that such an approach introduces additional complexity due to the need for model selection both in terms of the number and nature of the attentive models. The performance of model selection (e.g., using the Bayesian information criterion as proposed for regular LMFA; Vogelsmeier, Vermunt, van Roekel, & De Roover, 2019) has yet to be evaluated when combining exploratory attentive measurement models with the confirmatory C/IER component model.

Another limitation is that the method is currently suitable only for ordered response options (e.g., resulting from Likert scales). However, the proposed method has the potential to be extended to also accommodate continuous scales (e.g., resulting from visual analog scales). A straightforward approach is to build upon the mixture factor analysis model employed in the original LMFA and specify confirmatory factor models (see Kam & Cheung, 2023, for a constrained factor mixture model for C/IER) instead of exploratory ones within the states (which is possible in Latent GOLD; Vermunt & Magidson, 2021). However, simulation studies are required to evaluate the performance of this extension for distinguishing between attentive responding and C/IER in data gathered using continuous scales. An alternative approach worth exploring is employing a beta item response model accommodating the constrained range of visual analog scales (Noel & Dauvier, 2007) within the states. However, for beta item response models, C/IER component models still need to be developed.

To further aid the disentanglement of attentive from C/IER observations future research may enrich the proposed component models with the additional information on response behavior contained in item-level response times. This can easily be achieved by incorporating additional measurement models for attentive and inattentive response times alongside the models for item responses (Ulitzsch, Pohl, et al., 2022; Ulitzsch et al., 2020).

Obviously, the proposed approach is not applicable when EMAs administer single-indicator measures, as is often done to keep surveys as short as possible. In this case, the screen-time-based mixture modeling approach by Ulitzsch et al. (2024) may be a viable alternative.

One limitation of the application is that we included participant-level predictors for the transition probabilities in the latent Markov model, which was only possible by repeating the participant-level scores at every measurement occasion. This approach is not optimal—especially for analyses that extend beyond the exploratory purposes of the present study—because the sample size is artificially inflated. Thus, the accuracy of the findings is affected. This could be seen for the predictor *CR_check*, which was significantly related to transition probabilities but was practically uninformative regarding the likelihood of attentive responding and C/IER. A better approach would be to use a mixture latent Markov model (Crayen et al., 2017; Vermunt, 2008) with participant-level predictors in step 3. Mixture latent Markov models cluster participants based on the most prominent differences in the transition patterns (e.g., participants who frequently transition between states versus those who are always in the C/IER or attentive state, respectively). Participant-level predictors can subsequently be included to predict participant-specific cluster membership. However, to our best knowledge, it is currently not possible to estimate mixture latent Markov models while also accounting for unequal intervals using continuous-time latent Markov modeling in open-source software and, for this article, we aspired to present an open-source estimation of the proposed methodology. However, if desired, step 3, including the additional mixture component, could be performed in Latent GOLD (Vermunt & Magidson, 2021). Example applications of mixture latent Markov modeling in step 3, including syntax, can be found in Vogelsmeier, Vermunt, Keijsers, and De Roover (2021) and Vogelsmeier, Cloos, et al. (2023).

References

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra-and inter-individual model structures. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00883

Agresti, A. (1990). *Categorical data analysis*. John Wiley & Sons.

Arias, V. B., Garrido, L., Jenaro, C., Martinez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*, 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Bar-Hillel, M. (2015). Position effects in choice from simultaneous displays: A conundrum solved. *Perspectives on Psychological Science*, *10*(4), 419–433. https://doi.org/10.1177/1745691615588092

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Comments on: Latent markov models: A review of a general framework for the analysis of longitudinal data with covariates. *Test*, *23*, 473–477. https://doi.org/10.1007/s11749-014-0387-1

Böckenholt, U. (2005). A latent markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychological Methods*, *10*, 65–83. https://doi.org/10.1037/1082-989X.10.1.65

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*(2), 218.

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*, *29*(1), 55–71. https://doi.org/10.1002/per.1961

Chalmers, R. (2012). Mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. Chapman & Hall.

Crayen, C., Eid, M., Lischetzke, T., & Vermunt, J. K. (2017). A continuous-time mixture latent-state-trait markov model for experience sampling data. *European Journal of Psychological Assessment*, *33*, 296–311. https://doi.org/10.1027/1015-5759/a000418

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Curran, P. G., & Denison, A. J. (2019). Creating carelessness: A comparative analysis of common techniques for the simulation of careless responder data. https://doi.org/10.31234/osf.io/ge6fa

Dejonckheere, E., & Erbas, Y. (2021). Designing an experience sampling study. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing esm studies* (pp. 33–70). Center for Research on Experience Sampling; Ambulatory Methods, Leuven.

Deng, S., E. McCarthy, D., E. Piper, M., B. Baker, T., & Bolt, D. M. (2018). Extreme response style and the measurement of intra-individual variability in affect. *Multivariate behavioral research*, *53*(2), 199–218. https://doi.org/10.1080/00273171.2017.1413636

Denison, A. (2022). *Prevalence and predictors of careless responding in experience sampling research* [Thesis]. https://digitalcommons.usf.edu/etd/9341

DeSimone, J. A., DeSimone, A. J., Harms, P., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology*, *67*(2), 309–338. https://doi.org/10.1111/apps.12117

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

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

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology*, *69*(1), 153–166. https://doi.org/10.1037

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2022). 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

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2023). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*. https://doi.org/10.1037/met0000580

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

Jackson, C. H., & Sharples, L. D. (2002). Hidden markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine*, *21*, 113–128. https://doi.org/10.1002/sim.886

Jaso, B. A., Kraus, N. I., & Heller, A. S. (2022). Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods*, *27*(6), 958–981. https://doi.org/10.1037/met0000312

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

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

Kam, C. C. S., & Cheung, S. F. (2023). A constrained factor mixture model for detecting careless responses that is simple to implement. *Organizational Research Methods*. https://doi.org/10.1177/10944281231195298

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739–754. https://doi.org/10.1037/emo0000384

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*(3), 450–470. https://doi.org/10.1037/a0019216

McKay, A. S., Garcia, D. M., Clapper, J. P., & Shultz, K. S. (2018). The attentive and the careless: Examining the relationship between benevolent and malevolent personality traits with careless responding in online surveys. *Computers in Human Behavior*, *84*, 295–303. https://doi.org/10.1016/j.chb.2018.03.007

McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(5), 807–822. https://doi.org/10.1080/10705511.2021.1915788

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Morren, M., van Dulmen, S., Ouwerkerk, J., & Bensing, J. (2009). Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain*, *13*(4), 354–365. https://doi.org/10.1016/j.ejpain.2008.05.010

Myin-Germeys, I., & Kuppens, P. (Eds.). (2021). Center for Research on Experience Sampling; Ambulatory Methods Leuven; Leuven.

Nichols, A. L., & Edlund, J. E. (2020). Why don't we care more about carelessness? understanding the causes and consequences of careless participants. *International Journal of Social Research Methodology*, *23*(6), 625–638. https://doi.org/10.1080/13645579.2020.1719618

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, *63*, 1–11. https://doi.org/10.1016/j.jrp.2016.04.010

Noel, Y., & Dauvier, B. (2007). A beta item response model for continuous bounded responses. *Applied Psychological Measurement*, *31*(1), 47–73. https://doi.org/10.1177/0146621605287691

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

Oort, F. J., Visser, M. R., & Sprangers, M. A. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research*, *14*, 599–609. https://doi.org/10.1007/s11136-004-0831-x

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

Rein, M. T., Vogelsmeier, L. V. D. E., & Bolsinova, M. (2022). *Assessing well-being in everyday life: Developing the scale for ecological momentary assessment of life satisfaction* [Thesis]. https://doi.org/10.17605/OSF.IO/X94EK

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, *34*.

Samejima, F. (2016). Graded response models. In *Handbook of item response theory* (pp. 123–136). Chapman; Hall/CRC.

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, *24*(7), 755–789. https://doi.org/10.1080/02643290701754158

Schmitt, N., & Stuits, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*(4), 367–373. https://doi.org/10.1177/014662168500900405

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strengths and weaknesses. In E. Diener (Ed.), *Assessing well-being. the collected works of ed diener*. Springer. https://doi.org/10.1007/978-90-481-2354-4

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, *46*(1), 41–54. https://doi.org/10.3758/s13428-013-0353-y

Uglanova, I., Nagy, G., & Ulitzsch, E. (in preparation). A mixture IRT model for careless responding with flexible assumptions.

Ulitzsch, E., Domingue, B. W., Kapoor, R., Kanopka, K., & Rios, J. (2023). A probabilistic filtering approach to non-effortful responding. *Educational Measurement: Issues and Practice*. https://doi.org/10.1111/emip.12567

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., Pohl, S., Khorramdel, L., Kroehne, U., & von Davier, M. (2022). A response-time-based latent response mixture model for identifying and modeling careless and insufficient effort responding in survey data. *Psychometrika*, *87*(2), 593–619. https://doi.org/10.1007/s11336-021-09817-7

Ulitzsch, E., Pohl, S., Khorramdel, L., Kroehne, U., & von Davier, M. (2023). Using response times for joint modeling of careless responding and attentive response styles. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/10769986231173607

Ulitzsch, E., Shin, H.-J., & Lüdtke, O. (2023). Accounting for careless and insufficient effort responding in large-scale survey data—Development, evaluation, and application of a screen-time-based weighting procedure. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-02053-6

Ulitzsch, E., von Davier, M., & Pohl, S. (2020). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level nonresponse. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 83–112. https://doi.org/10.1111/bmsp.12188

Ulitzsch, E., Yildirim-Erbasli, S. N., Gorgun, G., & Bulut, O. (2022). An explanatory mixture IRT model for careless and insufficient effort responding in survey data. *British Journal of Mathematical and Statistical Psychology*, *75*, 668–698. https://doi.org/10.1111/bmsp.12272

van Laar, S., & Braeken, J. (2022). Random responders in the TIMSS 2015 student questionnaire: A threat to validity? *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.12317

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 33–51. https://doi.org/10.1177/0962280207081238

Vermunt, J. K., & Magidson, J. (2021). Upgrade manual for latent gold basic, advanced, syntax, and choice version 6.0. *Statistical Innovations Inc., Arlington*.

Vogelsmeier, L. V. D. E. (2022). *Latent markov factor analysis: A mixture modeling approach for evaluating within- and between-person measurement model differences in intensive longitudinal data* [Doctoral Thesis]. Tilburg University. Ridderprint. https://pure.uvt.nl/ws/portalfiles/portal/59625665/Vogelsmeier_Latent_14_01_2021.pdf

Vogelsmeier, L. V. D. E., & De Roover, K. (2021). 'lmfa': An r-package for exploring measurement invariance in intensive longitudinal data with continuous-time latent markov factor analysis. https://github.com/LeonieVm/lmfa

Vogelsmeier, L. V. D. E., Vermunt, J. K., Böing-Messing, F., & De Roover, K. (2019). Continuous-time latent markov factor analysis for exploring measurement model changes across time. *Methodology*, *15*, 29–42. https://doi.org/10.1027/1614-2241/a000176

Vogelsmeier, L. V. D. E., Vermunt, J. K., Bülow, A., & De Roover, K. (2021). Evaluating covariate effects on esm measurement model changes with latent markov factor analysis: A three-step approach. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2021.1967715

Vogelsmeier, L. V. D. E., Vermunt, J. K., & De Roover, K. (2023). How to explore within-person and between-person measurement model differences in intensive longitudinal data with the r package lmfa. *Multivariate Behavioral Research*, *55*, 2387–2422. https://doi.org/10.3758/s13428-022-01898-1

Vogelsmeier, L. V. D. E., Vermunt, J. K., Keijsers, L., & De Roover, K. (2021). Latent markov latent trait analysis for exploring measurement model changes in intensive longitudinal data. *Evaluation & the Health Professions*, *44*, 61–76. https://doi.org/10.1177/0163278720976762

Vogelsmeier, L. V. D. E., Cloos, L., Kuppens, P., & Ceulemans, E. (2023). Evaluating dynamics in affect structure with latent Markov factor analysis. *Emotion*. https://doi.org/10.1037/emo0001307

Vogelsmeier, L. V. D. E., Vermunt, J. K., van Roekel, E., & De Roover, K. (2019). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 557–575. https://doi.org/10.1177/0163278720976762

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & Stuss, D. T. (2022). Effects of time of day on age differences in working memory. *Journal of Gerontology*, *57*(1), 3–10.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 189–94. https://doi.org/10.1007/s10862-005-9004-7

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, *29*(2), 250–271. https://doi.org/10.1002/per.1986

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

Appendix

Data-Generating Item Parameter Values

Table A1

Data-Generating Item Parameter Values for Respondents With Category Preferences in the C/IER State

α	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
1	3.86	1.36	0.64	0.24	-0.59	-2.72

Notes: The parameter values were equal to the estimates of the application reported in the Empirical Application Section.

Table A2

Data-Generating Item Parameter Values for the Attentive State

	α	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
consider_happy	3.79	9.45	6.16	3.96	2.10	-0.93	-6.98
satisfied_life	4.10	10.20	6.46	4.20	1.84	-1.32	-7.34
change_many	-3.10	5.82	0.61	-1.02	-2.14	-4.31	-7.74
life_ideal	4.45	9.97	4.97	2.49	0.12	-2.97	-8.72
makes_happy	3.50	8.88	5.69	3.39	1.57	-0.95	-6.35
satisfied_situation	4.45	10.85	6.60	3.99	2.44	-0.74	-8.08
change_nothing	2.55	6.23	3.14	1.34	0.16	-1.48	-5.58
leaves_desired	-2.36	4.97	1.04	-0.26	-1.52	-3.19	-6.36
important_things	2.86	7.52	4.11	2.12	0.46	-1.80	-6.15
line_life	4.07	9.25	4.83	2.63	0.42	-2.55	-8.33

Notes: The parameter values were equal to the estimates of the application reported in the Empirical Application Section.