Personality Psychology

On the Role of Response Styles in the Study of Intraindividual Variability

Steffen Nestler1,2, Katharina Geukes1, Tobit Zaun1, Theresa Eckes1

1 Psychology, Universität Münster, Münster, Germany

Keywords: well-being, within-person variability, response styles, intraindividual variability

In a number of psychological disciplines, experience sampling studies are used to investigate when and why individuals vary in their cognitions, feelings, or personality states, to name just a few examples. However, to validly answer such research questions, researchers have to use reliable and valid measures of such intraindividual or within-person variability. Following Baird, Lucas, and Donnellan (2017), the present study uses data from two experience sampling studies to investigate whether response style variability can be assessed in a reliable way, whether variability measures are related to such response styles, and whether it is important to control for response styles in later analyses. We found that individuals differ in their response style variability and that these differences can be reliably assessed. Furthermore, response styles were moderately related to personality indicators, affect, and self-esteem variability. They were not associated with any of the outcome measures, so that most results remained unchanged when controlling for these differences. Nevertheless, our results indicate that researchers should additionally assess response styles when studying phenomena involving intraindividual variability.

There is growing interest in psychological research in investigating the determinants and consequences of individuals’ variations in their cognitions, feelings, and behaviors. For example, personality psychologists examine whether certain personality traits such as narcissism go along with higher self-esteem variability (e.g., Geukes, Nestler, Hutteman, Dufner, et al., 2017) or whether variability in states or in personality traits across social roles is related to well-being or adjustment (e.g., Baird et al., 2006; Magee et al., 2018). In life-span psychology, researchers investigate whether age predicts differences in positive and negative affect variability (Röcke & Brose, 2013). Moreover, emotional psychologists are interested in whether people vary in their application of emotion regulation strategies and whether this type of variability is related to trait negative affect (e.g., Blanke et al., 2020).

In these, but also in all other studies of within-person variability (also known as intraindividual variability), reliable and valid measures of variability are needed to properly investigate the research questions at hand. Interestingly, however, there are only few studies that address the measurement properties of intraindividual variability indicators. One exception is the study by Baird, Lucas, and Donnellan (2017) in which individuals were asked to judge several neutral objects, allowing the authors to measure individuals’ preferences for certain response categories. Thereafter, participants took part in a daily diary study in which they repeatedly rated adjectives describing themselves (e.g., their self-esteem). The results showed that the response style measure was moderately associated with intraindividual variability in the substantive variables. This suggests that failure to control for response styles can lead to erroneous conclusions about the hypotheses being investigated – for example, falsely concluding that intraindividual variability in a substantive variable is associated with a certain determinant or consequence.

The aim of the present research is to replicate and extend the results by Baird et al. (2017) using data from two large experience sampling studies in which participants were repeatedly asked to provide ratings on different personality, affect, and self-esteem states. Specifically, we investigate the extent to which the intraindividual variability indicators between substantive variables are correlated, whether and how strongly they are correlated with response style variability measures, and finally, whether they are related to several adjustment indicators even when controlling for response styles.

Measuring Intraindividual Variability

Several approaches are used in psychological research to measure an individual’s intraindividual variability. First,
one approach is to employ questionnaires, that is, to explicitly ask people to judge how variable they are. Several scales have been constructed for this purpose, including the Self-pluralism Scale (McReynolds et al., 2000; example item: "I act and feel essentially the same whether at home, at work, or with friends."), the Stability-of-self Scale (Rosenberg et al., 1989; example item: "My opinion of myself tends to change a good deal instead of always remaining the same. [reverse scored]"), or the Self-Concept Clarity Scale (Campbell et al., 1996; example item: "My beliefs about myself often conflict with one another." [reverse scored]).

Second, another way to measure variability is to examine how strongly a person varies in her responses to items within a scale assessing a construct or how strongly she varies across different scales within a test battery. For example, researchers computed participants’ standard deviation across responses to items in an extraversion scale in order to measure intraindividual variability in extraversion (e.g., Reise & Waller, 1995). Third, a related approach is to ask participants to judge the variable under consideration (e.g., extraversion) across a number of different social roles, relationships, or contexts (e.g., extraversion with friends, with family etc., see e.g., Baird et al., 2006, 2017). Here as well, intraindividual variability is measured by computing the standard deviation of participants’ responses to the items across social roles, relationships, or contexts.

Fourth and finally, another approach to measuring intraindividual variability is to assess a person’s thoughts, feelings, or behaviors across multiple days or multiple times per day over a period of several days in an experience sampling or daily diary study (e.g., Baird et al., 2006; Fleeson, 2001; Geukes, Nestler, Hutteman, Dufner, et al., 2017; Geukes, Nestler, Hutteman, Küfner, et al., 2017). The indicator of intraindividual variability for the variable in question is then obtained by computing an individual’s standard deviation for the respective item across all measurement occasions. If multiple indicators or items for a construct are assessed, one could average the standard deviations across items. Another approach would be to first average the respective items within each measurement occasion and compute the standard deviation of this average for an individual over time. Independently of how one proceeds, the experience sampling/daily diary approach is perhaps the most face-valid approach to assessing intraindividual variability in a range of different constructs. Due to increasing opportunities to collect and thus the increasing availability of experience sampling and daily diary data, in our experience, this approach is also used most frequently across psychological disciplines to examine intraindividual variability phenomena. Therefore, we focus on this final approach in the present paper.

Limitations of Daily Diary Measures of Intraindividual Variability

Despite their strengths, intraindividual variability measures based on daily diary data are not without problems. A first (mathematical) problem with standard deviations (or variances) is that they depend on a person’s mean score on the respective item¹. For example, when a person’s average for an extraversion item is 5 and the item was assessed on a 1 to 5-point scale, no variance in extraversion is mathematically possible. However, a person who has a mean of 3 could have a wider range of extraversion scores and hence vary more strongly. Thus, a more extreme average is linked to a smaller standard deviation, while the maximum standard deviation value can be achieved by persons whose average corresponds with the scale mean. This implies that persons’ standard deviations are correlated with their means and that one should account for this correlation when investigating research questions regarding intraindividual variability (see Baird et al., 2006).

A second problem concerns the effects of response styles, that is, a person’s tendency to prefer specific response scale options or categories over others (see Henninger & Meiser, 2020). For example, individuals may tend to choose the highest or lowest response categories, or they may tend to choose the middle category. Research shows that such individual differences in response styles can have a strong influence on the measurement of traits (e.g., Baumgartner & Steenkamp, 2001), occur consistently across traits (e.g., Weijters et al., 2010), and are stable across time (e.g., Wetzel et al., 2016). As response styles influence aggregate measures, they may also affect intraindividual variability indicators that are computed across repeated measurements. For instance, when people tend to more often use the extreme scale categories (either high or low), this can lead to a great and potentially overestimated intraindividual variability score, whereas when people often use the middle scale category, this can lead to a small and potentially underestimated intraindividual variability score.

To the best of our knowledge, Baird et al. (2017, Study 2) is the only study to directly examine whether intraindividual variability scores are contaminated by response styles. In Baird et al. (2017), participants took part in a daily diary study in which they were asked to rate different adjectives for the Big Five traits, positive and negative affect, and self-esteem once a day for 14 days. Prior to the daily diary phase of the study, they also completed a response style measure, and after the study, they reported on different measures of well-being and adjustment, for example, trait self-esteem and life satisfaction. The response style measure consisted of a neutral object rating task in which participants were asked to indicate their satisfaction with 25 neutral objects

¹ Another problem is that the construct the item measures might change in a linear or non-linear way over time. A remedy to this problem is to detrend the data for each person prior to data analyses or include a time trend variable when computing the standard deviations (see Geukes, Nestler, Hutteman, Dufner, et al., 2017). However, as such construct changes occur relatively slowly and therefore infrequently in daily diary studies, we do not discuss this issue further here.
On the Role of Response Styles in the Study of Intraindividual Variability

The present research focuses on the role of response styles in studies of intraindividual variability. The aim of this article is to replicate and extend the results by Baird et al. (2017) using two large experience sampling studies. We deem successful replications important in the light of the current replication crisis in psychology generally, but also specifically because researchers routinely examine hypotheses concerning intraindividual variability in many psychological disciplines. It is therefore crucial to know—optimally, based on robust, that is, successfully replicated, empirical findings—whether measures of intraindividual variability are influenced by response styles. If this is the case, researchers would have to control for response style differences in their analyses to avoid drawing the wrong conclusions regarding their hypotheses. Of course, this would necessitate simple and reliable measures of response styles, whereby these properties should ideally have been replicated across multiple independent studies. With our research, we aim to provide such a first replication. However, we also want to extend the findings of Baird et al. (2017) by investigating a second response tendency measure and by examining the retest reliability of the measure. Furthermore, we also examine the associations of the response style measures across different operationalization of intraindividual variability and we provide an important extension regarding the associations with well-being measures.

We use data from two daily diary studies. In both studies, participants took part in an experience sampling study in which they completed state measures of different personality traits, affect, and self-esteem. After this experience sampling portion of the study, they also provided ratings on different trait measures concerning well-being. In the first study, called FLIP (fluctuations in personality), participants completed different measures of response tendencies after the experience sampling portion of the study, while in the second study, called FLUX (fluctuations), participants answered these measures at both the start and the end of the study. This allows us to investigate the stability of participants’ response style variability scores. Consistent with the findings by Baird et al. (2017), we assume that, first, participants display interindividual differences in response style variability and that those differences can be reliably captured and are relatively stable over time. Second, we assume that interindividual differences in response styles are related to differences in intraindividual variability in the substantive variables. Third and finally, we examine the associations between the variability measures and well-being measures after controlling for response styles. The results of these analyses depend on whether response style variability is related to the well-being outcomes we are studying. If that is the case, we predict that the correlations will disappear.

Method

FLIP Study

Sample

The sample consisted of 102 participants. Six participants did not complete the second online survey in which we assessed response styles and outcome measures. These participants were therefore excluded from the analyses, resulting in a final sample of 96 individuals. Most of the participants were women (81%) and their age ranged between 18 and 65 years ($M = 22.46$ years, $SD = 5.59$ years). The number of completed daily reports per participant ranged from 33 to 82 ($M = 75.98$, $SD = 8.84$, $Mdn = 79.5$). Participants received monetary compensation (about 50 Euro) for their participation. The FLIP study was approved by the ethics committee of the University of Leipzig (331/17-ek).

Procedure

The data collection phase took place from October to December 2017. On each day during this period (82 consecutive days), participants were asked to retrospectively rate their experiences and behaviors on this day with regard to Big Five personality states, their self-esteem, a number of motivational states, and an affect grid. Before and after this daily diary phase, participants completed online surveys to obtain demographic information as well as measures of personality traits (i.e., Big Five personality traits, agency and communion, narcissism, motives), well-being measures (self-esteem, positive and negative affect, life satisfaction), and measures of response styles. Here, we only use the daily diary measures of personality states and self-esteem and the questionnaire data from the second assessment (after the daily diary portion) concerning self-esteem, positive and negative affect, life satisfaction, and response styles.

Measures

Daily diary variability measures. Personality states were measured with 19 adjectives. We adapted these adjectives from Borkenau and Ostendorf (1998) and the interpersonal adjective list (Jacobs & Scholl, 2005). Three adjectives were indicators of neuroticism (nervous, relaxed, irritable), four were indicators of extraversion (assertive, unsociable, shy, outgoing), three indicated openness (curious, creative, witty), five indicated agreeableness (hostile,
compliant, sensitive, friendly, cynical, helpful), and three indicated conscientiousness (diligent, organized, negligent). Furthermore, one item measured daily self-esteem ("I am satisfied with myself"). Participants were instructed to rate how well the respective adjective or item matched their experiences on that particular day ("Today I was..."). All ratings were made on scales ranging from 1 ("strongly disagree") to 6 ("strongly agree"). Note that the personality state items were presented in a randomized order on each day and that the self-esteem item was assessed after the personality states.

We followed the approach by Baird et al. (2017; see also Baird et al., 2006) and calculated one intraindividual variability score across all personality states. To this end, we first calculated the within-person mean and within-person standard deviation for each of the nineteen adjectives. The standard deviations for each adjective were then regressed on each person's corresponding mean and the square of this mean. The residuals of these regressions, reflecting mean-corrected variability scores, were then averaged to obtain a score of daily personality variability. The nineteen standard deviation residuals were also used to calculate Cronbach's alpha for this variability. In addition to this overall measure of intraindividual personality variability, we also calculated an intraindividual variability score for each of the Big Five dimensions by aggregating the respective mean-corrected variability scores. For self-esteem, we followed the same approach as for the personality state items (without aggregating, as there was only one item). To calculate the reliability of self-esteem variability, we computed the corrected standard deviation for the first half and second half of each participant's daily reports (e.g., Day 1 to Day 40 and Day 41 to Day 80). The correlation between the two standard deviations was then used to compute the reliability score.

We used the mean-corrected standard deviation in our analyses to stay as close as possible to the approach used in Baird et al. (2017). However, we also implemented two further approaches to examine (and ensure) that identical result patterns occur with different operationalizations of intraindividual variability. The first alternative approach consisted of aggregating the daily diary items per dimension and using the aggregate to compute a (mean-corrected) standard deviation. The second approach consisted of computing the relative variability index (see Mestdagh et al., 2018) instead of the mean-corrected standard deviation for each adjective. For this index, the individual's standard deviation for a variable is divided by an individual's maximum possible standard deviation given her/his mean resulting in a measure of intraindividual variability that is less influenced by the mean than is the mere standard deviation.

**Self-concept Clarity.** As an explicit measure of intraindividual variability, participants were asked to complete the Self-Concept Clarity Scale (Campbell et al., 1996; German version: Stucke, 2002) at the second assessment point. This scale consists of twelve items, each answered on a scale ranging from 1 ("strongly disagree") to 5 ("strongly agree", $\alpha = .90$).

**Outcome measures.** For global self-esteem, participants answered the Rosenberg Self-esteem Scale (Rosenberg, 1989, German version: von Collani & Herzberg, 2003), which consists of 10 items to be answered on a scale ranging from 1 ("strongly disagree") to 4 ("strongly agree", $\alpha = .92$). Furthermore, as a measure of life satisfaction, they were asked to fill out the Satisfaction with Life Scale (Diener et al., 1985; German version: Glaesmer et al., 2011). This scale consists of 5 items to be answered on a scale ranging from 1 ("strongly disagree") to 7 ("strongly agree", $\alpha = .87$). Additionally, participants answered a short version of the Positive and Negative Affect Schedule (Watson et al., 1988) containing 5 items measuring positive affect ($\alpha = .71$; active, optimistic, determined, in a good mood, relaxed) and two items measuring negative affect ($\alpha = .57$; inhibited, nervous). All seven items were answered on a scale ranging from 1 ("does not apply") to 6 ("applies perfectly").

**Response style measures.** Participants completed two measures of response styles: the Simpsons characters rating task and a neutral object rating task. Both measures were suggested by Baird et al. (2017) but only one (the neutral object rating task) was assessed in their daily diary study. In the Simpsons task, participants were asked to rate how well ten adjectives describe four characters from the Simpsons (Marge, Homer, Bart, and Lisa). The ten adjectives were energetic, trustful, dependable, vulnerable, philosophical, talkative, cooperative, organized, irritable, and intelligent, each rated on a scale ranging from 1 ("does not describe this person") to 5 ("describes this person very well"). To obtain the Simpsons variability measures, we proceeded exactly like Baird et al. (2017): For each participant, we first computed the mean and the standard deviation for each of the ten items across the four Simpsons characters. Each of the ten cross-character standard deviations were then regressed on the respective mean and square of the mean. Thereafter, the ten standard deviation residuals were averaged to create an overall measure of variability for the ratings of the four Simpsons characters. We also used the ten standard deviation residuals to compute Cronbach's alpha for this measure. The idea behind this measure is that, because participants are judging the same stimuli, there should be a similar amount of cross-character variability in the sample. Hence, variability in a person's personality state or self-esteem variability should be unrelated to a variability measure concerning cartoon characters.

For the neutral object rating task, participants indicated their satisfaction with 25 items describing neutral objects (e.g. "the movies produced today" or "public transportation"). Each item was rated on a scale ranging from 1 ("very unsatisfied") to 5 ("very satisfied"). To obtain an index of intraindividual variability for these ratings, we also calculated a standard deviation and the mean of each person's ratings across the 25 objects. The standard deviations were then regressed on the mean and its squared means to obtain the corrected variability score. As an index of reliability, we computed a split-half reliability by taking the average split-half reliability from 1000 random splits of the 25 items (see Baird et al., 2017, for a similar approach). The underlying idea behind this measure is the same as for the Simpsons rating task: As most of these objects are relatively neutral, they should be rated similarly by different persons, with the consequence that participants should exhibit a similar amount of cross-object variability in the absence of interindividual differences in response styles.
On the Role of Response Styles in the Study of Intraindividual Variability

FLUX Study

Sample

This sample included 104 participants. Again, we excluded those participants with missing values in the response style and outcome measures in the second survey, resulting in a final sample of 95 participants. Most participants were female (85%); their age ranged between 18 and 37 years old (M = 25.27 years, SD = 6.45, Mdn = 53.0). Again, participants received monetary compensation for their participation (about 40 Euro). The FLUX study was approved by the ethics committee of the German Research Foundation (SN 072018).

Procedure

The data collection phase lasted from September to December, 2018. During this period, individuals participated in a measurement-burst experience sampling study consisting of four bursts of 14 days each (56 days in total). The first, second, and third burst were each followed by breaks of 14 days in which participants were not required to provide any ratings.

During each burst, participants were asked to rate their behavior on each day with regard to Big Five personality states, self-esteem, mood, and motivational states. Furthermore, prior to the first burst and after the last burst, they completed an online survey containing demographic measures, measures of personality traits (e.g., the Big Five personality traits), well-being (e.g., life satisfaction), and response styles, respectively. Here, we only use the daily diary measures for personality states, affect states (not assessed in FLIP), and self-esteem states and the outcome measures from the second assessment (after the daily diary portion) concerning self-esteem, positive and negative affect, life, and job satisfaction, and loneliness. Response style measures were completed in both the initial and final surveys, and we use the data from both time points in the analyses.

Measures

Daily diary variability measures. Personality states were assessed with twelve adjectives. One adjective measured neuroticism (emotionally stable), four adjectives measured neuroticism (emotionally stable), four adjectives measured extraversion (assertive, unsociable, shy, outgoing), one openness (versatility interested), four agreeableness (offending, hostile, compliant, sensitive), and one was an indicator of conscientiousness (reliable). Four adjectives were used to assess positive affect (active, happy, calm, relaxed) and three adjectives measured negative affect (annoyed, anxious, lonely). Finally, one item measured daily self-esteem ("I am satisfied with myself"). Similar to FLIP, participants were instructed to rate how well each adjective or item matched their behavioral experiences on that particular day ("Today I was... "). All ratings were made on scales ranging from 1 ("strongly disagree") to 6 ("strongly agree"). To obtain an overall measure of personality state variability, variables of the individual Big Five factors, positive and negative affect variability, and self-esteem variability, we followed the same statistical approach as described for the FLIP study. Please note that the personality state items and affect state items were presented in random order on each day and that the self-esteem item was assessed after the personality and affect states.

Self-Concept Clarity. Participants also completed the Self-Concept Clarity Scale as an explicit measure of intradividual variability. In contrast to FLIP, the Self-Concept Clarity Scale was assessed prior to (α = .92) and after the daily diary phase of the study (α = .91).

Outcome measures. We used the same global self-esteem measure (α = .95) and the same life satisfaction measure (α = .90) as in the FLIP study. For positive and negative trait affect, participants answered the Positive and Negative Affect Schedule (Watson et al., 1988; German version: Krohne et al., 1996), which contains ten items measuring positive affect (α = .87) and ten items measuring negative affect (α = .88), each of which is rated on a scale ranging from 1 ("not at all") to 5 ("extremely"). We also assessed participants’ job satisfaction with five items (adapted from Brayfield & Rothe, 1951), each rated on a scale ranging from 1 ("does not at apply") to 5 ("applies quite a lot", α = .91). Moreover, we assessed loneliness with four items (UCLA Loneliness Scale, Version 3; Russell, 1996; α = .88) to be answered on a scale ranging from 1 ("never") to 4 ("always").

Response style measures. In the FLUX study, we used the same two response style measures (i.e., the Simpsons task and the neutral object rating task) as in the FLIP study. We also used the same analytical approach to compute response tendency variability measures. The FLIP and FLUX studies differed only in the timing of the response style assessment: While the FLIP study measured response styles only after the daily diary phase, the FLUX study assessed response style measures both prior to and after the experience sampling phase of the study. This allowed us to also compute retest reliabilities for the two measures.

Power Analysis

Baird et al. (2017) report an average correlation of r = .28 between the daily diary variability measures and the neutral object rating task measure (see their Table 2; the Simpson task was administered in the daily diary study in Baird et al., 2017). When we assume a population correlation of ρ = .25, α = .05, and a one-tailed test, the power to detect such an effect is 80% when the sample size is N = 96 (FLIP) and 79% when the sample size is N = 95 (FLUX). Thus, the statistical power for investigating our research questions is good.

Results

We provide the data and the R code necessary to reproduce the results in an OSF project https://osf.io/taid9/. In each of the following paragraphs, we first report the results concerning FLIP and then the results concerning FLUX. Furthermore, we follow standard conventions to interpret correlations in terms of effect sizes: a correlation of r = .10 is considered a weak or small effect, a correlation of r = .30 as a moderate effect, and a correlation of r = .50 or larger as a large effect. Finally, across the two studies we found that the results and conclusions regarding the influence of
response styles based on the two alternative approaches to operationalize intraindividual variability were largely consistent with the conclusions based on the approach used in Baird et al. (2017). We therefore report the results of the two approaches only in Appendix 1 (standard deviation across an aggregate) and Appendix 2 (relative variability index), respectively.

Reliability and Correlations between Variability Measures

Table 1 shows the correlations between the different variability measures in the FLIP sample. As can be seen in the table’s diagonal, all four variability measures were sufficiently reliable. For the personality variability measure, this suggests that the nineteen variability scores measure something common. Furthermore, and replicating earlier research (e.g., Baird et al., 2006, 2017), correlations between the Self-Concept Clarity Scale—as an explicit measure of variability—and the other variability measures were small. The only exceptions were the small to moderate correlation with the variability measure stemming from the neutral object ratings and the conscientiousness variability measure. The two response style variability measures were positively correlated. However, this small to moderate correlation was not significantly different from zero. Importantly, moderate to large correlations were found between the two response style measures and the variability measures of the substantive variables. This indicates that response styles may play an important role in intraindividual variability.

Table 2 shows the correlations for FLUX. Again, all measures were sufficiently reliable in terms of internal consistency (i.e., Cronbach’s alpha, split-half reliability). Furthermore, the Self-Concept Clarity Scale and the two response scale measures were also substantially correlated across time, indicating good retest reliability. The internal consistency scores and the very large intercorrelations between the personality variability measures, affect variability measures, and self-esteem variability measure again show that the substantive variability measures are associated with one another. Furthermore, the correlations between the Self-Concept Clarity Scale and the other variability measures were again rather low, with the exception of the small to moderate correlations with the neutral object variability measure at the first and second assessment. Consistent with the results for FLIP, we observed moderate to large correlations between the two response style variability measures measured at Time 1 and Time 2 and the variability measures for the substantive variables, indicating again that response styles may play an important role in intraindividual variability. Finally, only one of the four correlations of the two response style variability measures was significant and moderate in size.

Altogether the results show—see Table 3 for a meta-analytic aggregate of the corresponding correlations from Tables 1 and 2—that interindividual differences in response style variability could be reliably assessed and that these differences were moderately related to interindividual differences in intraindividual variability of different substantive variables. Furthermore, across the two studies, we found that the variability measures of the substantive variables were highly correlated. Finally, the correlation between the response style measures was only small to moderate.

### Table 1. Reliabilities of and correlations between different response style measures and different intraindividual variability measures in FLIP

| Measure  | 1.  | 2.  | 3.  | 4.  | 5.  | 6.  | 7.  | 8.  | 9.  | 10. |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| SCC      | (.90) |     |     |     |     |     |     |     |     |     |
| NO       | -.21 | (.73) |     |     |     |     |     |     |     |     |
| ST       | .00  | .19  | (.83) |     |     |     |     |     |     |     |
| Daily N  | .04  | .31  | .32  | (.88) |     |     |     |     |     |     |
| Daily E  | .17  | .28  | .48  | .80  | (.87) |     |     |     |     |     |
| Daily O  | .12  | .18  | .41  | .71  | .83  | (.90) |     |     |     |     |
| Daily A  | .05  | .25  | .46  | .83  | .86  | .83  | (.87) |     |     |     |
| Daily C  | .21  | .22  | .40  | .70  | .81  | .83  | .80  | (.89) |     |     |
| Daily Pers.| .12 | .27  | .46  | .88  | .94  | .91  | .95  | .90  | (.97) |     |
| Daily SE | .18  | .24  | .37  | .60  | .70  | .57  | .68  | .65  | .70  | (.81) |

Note. Reliability coefficients are reported in parentheses in the table’s diagonal. Correlations in bold are significant at p < .05; N = 96. SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures are intraindividual variability measures.
Table 2. Reliabilities of and correlations between different response style measures and different intraindividual variability measures in FLUX

| Measure          | 1.  | 2.  | 3.  | 4.  | 5.  | 6.  | 7.  | 8.  | 9.  | 10. | 11. | 12. | 13. | 14. | 15. |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| SCC T1           | .83 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SCC T2           |     | .83 |     |     |     |     |     |     |     |     |     |     |     |     |     |
| NO T1            | -.27| -.18|     |     |     |     |     |     |     |     |     |     |     |     |     |
| NO T2            | -.24| -.16| .64 |     |     |     |     |     |     |     |     |     |     |     |     |
| ST T1            | -.07| -.02| .02 | .15 |     |     |     |     |     |     |     |     |     |     |     |
| ST T2            | -.07| -.05| .18 | .39 | .66 |     |     |     |     |     |     |     |     |     |     |
| Daily N          | -.01| .00 | .54 | .34 | .27 | .34 |     |     |     |     |     |     |     |     |     |
| Daily E          | -.03| -.04| .36 | .21 | .40 | .41 | .68 |     |     |     |     |     |     |     |     |
| Daily O          | -.12| -.06| .39 | .17 | .36 | .35 | .72 | .80 |     |     |     |     |     |     |     |
| Daily A          | -.03| -.01| .35 | .22 | .40 | .42 | .75 | .77 | .77 |     |     |     |     |     |     |
| Daily C          | -.04| -.03| .34 | .20 | .38 | .39 | .63 | .72 | .67 | .71 |     |     |     |     |     |
| Daily Per.       | -.02| -.03| .42 | .25 | .42 | .44 | .82 | .93 | .87 | .93 | .81 |     |     |     |     |
| Daily PA         | .00 | .00 | .37 | .22 | .39 | .44 | .82 | .81 | .83 | .80 | .76 | .90 |     |     |     |
| Daily NA         | .06 | .02 | .36 | .26 | .37 | .35 | .66 | .73 | .62 | .73 | .58 | .78 | .73 |     |     |
| Daily SE         | -.03| .01 | .33 | .15 | .34 | .37 | .71 | .74 | .72 | .65 | .62 | .77 | .77 | .52 |     |

Note. Reliability coefficients are reported in parentheses in the table's diagonal. Correlations in bold are significant at $p < .05$; $N = 93$. SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures are intraindividual variability measures.
Variability and Outcome Measures

Correlations and Partial Correlations between Response Style Measures, Interindividual Variability Measures, and Several Outcome Measures Related to Well-Being

The correlations between the response style measures and several outcome measures related to well-being are presented in Table 4. These correlations were found to be significant for both FLIP and FLUX. The outcome measures included Life satisfaction (LS), Positive affect (PA), and Negative affect (NA). The correlations were moderate in size (ranging from -0.20 to 0.37) and significant at p < .05 levels.

Correlations and Partial Correlations between Variability and Outcome Measures

In the next step, we examined the correlations between the substantive variability measures, the two response style variability measures, and several outcome measures related to well-being. Table 4 shows the results obtained for FLIP. The correlations between the response style and outcome measures were small to moderate and not significantly different from zero. Self-esteem was moderately negatively correlated with almost all the variability measures for the substantive variables, indicating that individuals who vary less (i.e., are more stable) have higher self-esteem values. These correlations were still significantly different from zero when we controlled for the two response style measures (see the partial correlation column in Table 4). Life satisfaction was significantly associated with three measures of variability. After controlling for response styles, however, only one of these correlations was significantly different from zero and moderate in size (i.e., variability in conscientiousness). Finally, no significant associations were found for positive and negative affect.

Table 3. Meta-analytic Average Correlations between Different Response Style Measures and Different Intraindividual Variability Measures in FLIP and FLUX

| Measure          | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|------------------|----|----|----|----|----|----|----|----|----|-----|
| SCC              | -  | -  | -  | -  | -  | -  | -  | -  | -   | -   |
| NO               | -.18| -  | -  | -  | -  | -  | -  | -  | -    | -    |
| ST               | -.03| .28| -  | -  | -  | -  | -  | -  | -    | -    |
| Daily N          | .02 | .32| .33| -  | -  | -  | -  | -  | -    | -    |
| Daily E          | .06 | .24| .45| .74| -  | -  | -  | -  | -    | -    |
| Daily O          | .03 | .18| .38| .71| .81| -  | -  | -  | -    | -    |
| Daily A          | .02 | .24| .44| .79| .82| .80| -  | -  | -    | -    |
| Daily C          | .09 | .21| .40| .67| .77| .75| .75| -  | -    | -    |
| Daily Pers.      | .05 | .26| .45| .85| .94| .89| .94| .86| -    | -    |
| Daily SE         | .10 | .20| .37| .65| .72| .64| .66| .64| .74  | -    |

Note. The average correlation was computed as a weighted average in which each of the two correlations was weighted with the inverse of the standard error of the correlation (i.e., a fixed-effects meta-analysis). Correlations in bold are significant at p < .05; significance was determined by computing the 95% confidence interval for the meta-analytic average. SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures are intraindividual variability measures. In case of FLUX, the integrated correlations for the response style measures refer to the measures assessed at the second time-point.

Table 4. Correlations and partial correlations between response style measures, interindividual variability measures and well-being outcome measures in FLIP

| Measure          | SE | LS | PA | NA |
|------------------|----|----|----|----|
| SCC              | -.13| -.18| -.24| -.32|
| NO               | .17 | .03 | .14 | -.18|
| ST               | -.12| -.12| .03 | .04|
| Daily N          | -.06| -.04| .01 | .02|
| Daily E          | .08 | -.01| .00 | -.02|
| Daily O          | -.08| -.01| .01 | .01|
| Daily A          | -.11| -.11| .02 | .02|
| Daily C          | -.27| -.20| .07 | .05|
| Daily Pers.      | -.25| -.13| .04 | .05|
| Daily SE         | -.26| -.26| .02 | .00|

| Measure          | SE | LS | PA | NA |
|------------------|----|----|----|----|
| SCC              | -.13| -.18| -.24| -.32|
| NO               | .17 | .03 | .14 | -.18|
| ST               | -.12| -.12| .03 | .04|
| Daily N          | -.06| -.04| .01 | .00|
| Daily E          | -.08| -.01| .01 | -.02|
| Daily O          | -.08| -.01| .01 | .01|
| Daily A          | -.11| -.11| .02 | .02|
| Daily C          | -.27| -.20| .07 | .05|
| Daily Pers.      | -.25| -.13| .04 | .05|
| Daily SE         | -.26| -.26| .02 | .00|

Note. Correlations and partial correlations in bold are significant at p < .05; N = 96 or N = 94 (in case of partial correlations). NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures in the rows are intraindividual variability measures.
life satisfaction and negative affect. In addition, we found that variability in neuroticism, positive affect, and self-esteem were moderately associated with negative affect. All other correlations were small and not significantly different from zero. When we controlled for the response style measures assessed at the first time point\(^2\), most of the significant correlations remained significantly different from zero.

To summarize, across the two studies we found small to moderate correlations between the response style measures and the outcome measures (see Table 6 for a meta-analytic aggregate of the corresponding correlations from Tables 4 and 5). With regard to the associations between the variability measures for the substantive variables and the outcome measures, a mixed picture emerged. For some outcome variables, the correlations in FLUX were smaller but in the same direction as in FLIP (e.g., self-esteem; this explains why the aggregated correlations are significantly different from zero in Table 6), but for other variables the associations disappeared. However, in both samples the correlations did not substantially change when we controlled for response styles.

**Discussion**

In a number of psychological disciplines, it is of great interest to investigate whether and why individuals fluctuate in their cognition, affect, personality, and behavior. To properly examine research questions concerning such interindividual variability, it is important to use reliable and valid measures of interindividual variability. Here, we provided a replication and extension of a study by Baird et al. (2017) and thereby investigated whether variability measures based on self-reports are affected by response styles, and in turn, whether it is important to control for response styles when studying interindividual variability phenomena.

Replicating results by Baird et al. (2017), we found that there are reliable and stable interindividual differences in response style variability and that these differences were moderately related to interindividual differences in personality, affect, and self-esteem variability. Extending results by Baird et al. (2017), response styles were not associated with the well-being outcome measures assessed here. This had the consequence that most bivariate correlations between substantive variables remained insignificant and unchanged when we controlled for these differences. However, a caveat with regard to these correlations between interindividual variability indicators and outcomes is that most variability measures were either uncorrelated with the adjustment measures or the associations could not be replicated across the two studies. As such, we recommend further thorough replication efforts to substantiate our findings. Nevertheless, our results indicate that one should consider response styles when examining interindividual variability, as is already the case when investigating aggregated responses.

There are several options for how to consider response styles. One possibility is to measure participants’ response style values and control for these values in the analyses. Our results show that both the Simpsons task and the neutral object rating task are suitable measures for this purpose: Both measures can be easily and quickly collected before and/or after the experience sampling phase of a study. Furthermore, both measures are internally consistent and retest-reliable. Another possibility is to use a model-based approach in which response styles are directly considered in the statistical model. Such a model is based on the notion that, in addition to the corrected standard deviation or variance approach used here, the determinants and consequences of interindividual variability can also be investigated with an extension of the multilevel growth model (with time points nested within individuals), in which the Level-1 residual variance is allowed to vary across individuals (see Geukes, Nestler, Hutteman, Dufner, et al., 2017; Hedeker et al., 2008, 2009; Nestler, 2020, 2021). In such models, it should be possible to consider response styles in the same way as in structural equation models or item-response models (see Henninger & Meiser, 2020). Up until now, however, we know of only one article that has proposed such a model (Deng et al., 2018). We consider the development of such approaches a fruitful and important avenue for future research, because they would allow to investigate whether our and Baird et al.’s (2017) findings would generalize beyond the statistical approach to model response styles. Furthermore, these models could also help to better understand an unexpected result of our study, namely that the two response style variability measures were only moderately correlated, although they are assumed to measure the same thing. In a model-based approach, one could determine the correlation of the two measures with the latent response style factor to examine, for example, whether the two variables capture different aspects of the latent response tendency variable.

\(^2\) We also computed partial correlations with the measures assessed at the second time point. The results were almost identical to the results obtained with the first time point measures.
Table 5. Correlations and partial correlations between response style measures, interindividual variability measures and well-being outcome measures in FLUX

| Measure    | SE  | LS  | PA  | NA  | JS  | LO  | SE  | LS  | PA  | NA  | JS  | LO  |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| NO T1      | .10 | .13 | .07 | -0.05 | .25 | .09 |     |     |     |     |     |     |
| NO T2      | .20 | .20 | .11 | -0.14 | .22 | .11 |     |     |     |     |     |     |
| ST T1      | -0.04 | .07 | .08 | .10 | -0.03 | .07 |     |     |     |     |     |     |
| ST T2      | .02 | .01 | .13 | .07 | -0.01 | .02 |     |     |     |     |     |     |
| Daily N    | -0.14 | .00 | -0.01 | .22 | .09 | -0.03 | -0.23 | -0.07 | -0.09 | .28 | -0.04 | -0.12 |
| Daily E    | -0.06 | -0.05 | .02 | .12 | -0.04 | .03 | -0.10 | -0.08 | -0.05 | .11 | -0.14 | -0.03 |
| Daily O    | -0.07 | -0.08 | -0.02 | .17 | -0.02 | -0.11 | -0.11 | -0.12 | -0.09 | .18 | -0.13 | -0.20 |
| Daily A    | -0.14 | -0.10 | -0.07 | .15 | .03 | -0.13 | -0.19 | -0.14 | -0.15 | .15 | -0.05 | -0.23 |
| Daily C    | -0.22 | -0.24 | -0.17 | .21 | -0.12 | -0.10 | -0.28 | -0.30 | -0.26 | .22 | -0.23 | -0.18 |
| Daily Pers.| -0.13 | -0.09 | -0.04 | .18 | -0.01 | -0.07 | -0.19 | -0.15 | -0.13 | .19 | -0.13 | -0.16 |
| Daily PA   | -0.17 | -0.11 | -0.11 | .25 | -0.08 | -0.11 | -0.23 | -0.16 | -0.20 | .27 | -0.19 | -0.21 |
| Daily NA   | -0.09 | -0.08 | -0.07 | .09 | -0.02 | -0.05 | -0.13 | -0.12 | -0.14 | .08 | -0.12 | -0.12 |
| Daily SE   | -0.14 | -0.09 | -0.10 | .25 | -0.05 | -0.09 | -0.18 | -0.13 | -0.17 | .27 | -0.14 | -0.16 |

Note. Correlations and partial correlations in bold are significant at $p < 0.05$; $N = 93$ or $N = 91$ (in case of partial correlations). NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscienctiousness, SE = self-esteem, LS = Life satisfaction, PA = Positive affect, NA = Negative affect, JS = Job satisfaction, LO = Loneliness. SCC, NO, and ST are response style measures while all other measures in the rows are intraindividual variability measures.
A third possibility would be to use alternative response formats that are less affected by response styles. For example, a binary response format, with true/false response options, eliminates certain types of response style effects (Wetzel et al., 2016). Similarly, forced-choice response formats have been found to be less susceptible to response styles (see e.g., Maydeu-Olivares & Brown, 2010; but see also Schulte et al., 2021). However, a problem with alternative response formats is that it is at present unclear how to operationalize intraindividual variability when using such response formats: While the use of standard deviations could still be justified for a binary response format, standard deviations can no longer be used for paired comparisons. Nevertheless, we believe that this could be an interesting future research direction. Finally, another way to examine the consequences and determinants of intraindividual variability is to avoid the exclusive use of self-reports, but to complement them with intraindividual variability measures based on informant reports. We are aware of only a few experience sampling study (e.g., CONNECT; Geukes et al., 2019) in which other persons’ reports were collected in addition to self-reports. However, we believe that this should be done more often, because this allows to examine a number of interesting research questions, also with respect to the technical implementation.

Overall, our results raise the question of how reliable previous results on the phenomenon of intraindividual variability are, since most studies did not control for differences in response styles. They also point to the need for more research on the measurement-theoretic properties of self-report measures of intraindividual variability as well as self-reports in experience sampling studies (see Sun et al., 2020; Sun & Vazire, 2019, for interesting approaches). In addition, we believe that an interesting task for future research is to investigate whether one can also capture such constructs and their variation with measures that are not based on self-reports. For instance, it might be possible to reliably and validly capture variability with third-party reports (at least for observable constructs) or relevant mobile sensor data. However researchers proceed, we hope that the study of the measurement-theoretic properties of variability measures will be pursued more vigorously in the future.

---

**Data Accessibility Statement**

We embrace the values of openness and transparency in science (Schönbrodt et al., 2015; osf.io/4dvkw/). We therefore follow the 21-word solution (Simmons et al., 2012) or refer to the project documentation in the OSF. We furthermore publish all raw data necessary to reproduce the reported results and provide scripts for all data analyses reported in this manuscript (see https://osf.io/tajd9/).

**Competing Interests**

All authors certify that they have no conflict of interests in the subject matter or materials discussed in this manuscript.

**Contributions**

Steffen Nestler (SN) and Katharina Geukes (KG) made substantial contributions to the conception and design of the reported studies. SN was responsible for data collection. SN, Tobit Zaun (TZ), and Theresa Eckes (TE) analyzed the current data and jointly interpreted the results. SN drafted the article and KG, TZ, and TE revised it critically. SN, KG, TZ, and TE approved the final version of this manuscript to be published.

Submitted: June 28, 2021 PST, Accepted: October 26, 2021 PST

---

Table 6. Meta-analytic average correlations and partial correlations between response style measures, interindividual variability measures and well-being outcome measures in FLIP and FLUX

| Measure | Correlations | Partial correlations |
|---------|--------------|----------------------|
|         | SE | LS | PA | NA | SE | LS | PA | NA |
| NO      | .18 | .11 | .12 | -.16 | -.19 | .05 | -.02 | .15 |
| ST      | -.05 | .06 | .08 | .05 | -.22 | .12 | -.07 | .07 |
| Daily N | -.14 | -.03 | .03 | .10 | -.17 | -.14 | -.04 | .07 |
| Daily E | -.19 | -.12 | -.01 | .06 | -.19 | -.14 | -.10 | .08 |
| Daily O | -.16 | -.15 | .00 | .07 | -.31 | -.26 | -.20 | .13 |
| Daily A | -.16 | -.14 | .04 | .07 | -.23 | -.15 | -.09 | .10 |
| Daily C | -.27 | -.25 | -.14 | .12 | .22 | -.10 | -.08 | .16 |
|         | .20 | -.11 | .03 | .15 |     |     |     |     |
On the Role of Response Styles in the Study of Intraindividual Variability
REFERENCES

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90*(3), 512–527. https://doi.org/10.1037/0022-3514.90.3.512

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality, 69*, 170–179. https://doi.org/10.1016/j.jrp.2016.06.015

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 145–156. https://doi.org/10.1509/jmkr.38.2.145.18840

Blanke, E. S., Brose, A., Kalokerinos, E. K., Erbas, Y., Riediger, M., & Kuppens, P. (2020). Mix it to fix it: Emotion regulation variability in daily life. *Emotion, 20*(3), 475–485. https://doi.org/10.1037/emo0000566

Borkenau, P., & Ostendorf, F. (1998). The Big Five as an integrated view of personality: Traits as density of personality evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124–138. https://doi.org/10.1016/j.jrp.2016.06.005

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90*(3), 512–527. http://doi.org/10.1037/0022-3514.90.3.512

Borkenau, P., & Ostendorf, F. (1998). The Big Five as an integrated view of personality: Traits as density of personality evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124–138. https://doi.org/10.1016/j.jrp.2016.06.005

Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology, 35*(5), 307–311. https://doi.org/10.1037/h0055617

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*(1), 141–156. https://doi.org/10.1037/022-3514.70.1.141

Deng, S., McCarthy, D. E., Piper, M. E., Baker, T. B., & Bolt, D. M. (2018). Extreme response style and the measurement of intra-individual variability in affect. *Multivariate Behavioral Research, 53*(2), 199–218. http://doi.org/10.1080/00273171.2017.1415366

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Geukes, K., Breil, S. M., Hutteman, R., Nestler, S., Küfner, A. C. P., & Back, M. D. (2019). Explaining the longitudinal interplay of personality and social relationships in the laboratory and in the field: The PILS and the CONNECT study. *Plos One, 14*(1), e0210424. https://doi.org/10.1371/journal.pone.0210424

Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C. P., Egloff, B., Denissen, J. J. A., & Back, M. D. (2017). Puffed up but shaky selves: State self-esteem level and variability in narcissists. *Journal of Personality and Social Psychology, 112*(5), 769–786. https://doi.org/10.1037/osp0000093

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124–138. https://doi.org/10.1016/j.jrp.2016.06.015

Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the satisfaction with life scale (SWLS). *European Journal of Psychological Assessment, 27*(2), 127–152. https://doi.org/10.1027/1015-5759/a000058

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction, 104*(2), 297–307. https://doi.org/10.1111/j.1600-0443.2009.02435.x

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64*(2), 627–634. https://doi.org/10.1111/j.1541-0420.2007.00924.x

Henning, M., & Meiser, T. (2020). Different approaches to modeling response styles in divide-by-total item response theory (part 1): A model integration. *Psychological Methods, 25*(5), 560–576. https://doi.org/10.1037/met0000249

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL). *Diagnostica, 51*(3), 145–155. https://doi.org/10.1026/0012-1924.51.3.145

Judge, T. A., & Bretz, R. D. (1992). *Reports on an alternative measure of affective disposition (CAHRS Working Paper #2–39.*

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Alternative zur Skala zum Ausdruck von positivem und negativem affekt (PANAS). *Diagnostica, 42*, 159–156. https://doi.org/10.1007/t0107/t49650-000

Magee, C., Buechel, E. E., Human, L. J., Murray, D. R., & Biesanz, J. C. (2018). Is personality variability associated with adjustment? *Journal of Research in Personality, 72*, 22–43. https://doi.org/10.1016/j.jrp.2016.08.005

Maydeu-Olivares, A., & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research, 45*(6), 935–974. https://doi.org/10.1080/00273171.2010.531251
Schulte, N., Holling, H., & Bürkner, P.-C. (2021). Can high-dimensional questionnaires resolve the ipsativity issue of forced-choice response formats. *Educational and Psychological Measurement, 81*(2), 262–289. [https://doi.org/10.1177/0013164420954861](https://doi.org/10.1177/0013164420954861)

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 Word Solution*. [https://doi.org/10.2139/ssrn.2160588](https://doi.org/10.2139/ssrn.2160588)

Stucke, T. S. (2002). Überprüfung einer deutschen Version der Selbstkonzeptklarheits-Skalen von Campbell. *Zeitschrift für Differentielle und Diagnostische Psychologie, 23*(4), 475–484. [https://doi.org/10.1024/0170-1789.23.4.475](https://doi.org/10.1024/0170-1789.23.4.475)

Sun, J., Rhemtulla, M., & Vazire, S. (2020). Eavesdropping on missing data: What are university students doing when they miss experience sampling reports? *Personality and Social Psychology Bulletin, 47*(11), 1535–1549. [https://doi.org/10.1177/0146167220964639](https://doi.org/10.1177/0146167220964639)

Sun, J., & Vazire, S. (2019). Do people know what they’re like in the moment? *Psychological Science, 30*(3), 405–414. [https://doi.org/10.1177/0956797618818476](https://doi.org/10.1177/0956797618818476)

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift für Differentielle und Diagnostische Psychologie, 24*(1), 3–7. [https://doi.org/10.1024/0170-1789.24.1.3](https://doi.org/10.1024/0170-1789.24.1.3)

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. [https://doi.org/10.1037/0022-3514.54.6.1063](https://doi.org/10.1037/0022-3514.54.6.1063)

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-reported questionnaires. *Applied Psychological Measurement, 34*(2), 105–121. [https://doi.org/10.1177/0146621609358593](https://doi.org/10.1177/0146621609358593)

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment, 23*(3), 279–291. [https://doi.org/10.1177/1073191115583714](https://doi.org/10.1177/1073191115583714)
deviations and the variability scores based on the mean-corrected standard deviation of the average were significantly correlated, neuroticism: \( r = 0.71, p < 0.01 \); extraversion: \( r = 0.64, p < 0.01 \); openness: \( r = 0.93, p < 0.01 \); agreeableness: \( r = 0.80, p < 0.01 \); conscientiousness: \( r = 0.75, p < 0.01 \). Table A1a presents the correlations (and partial correlations) we obtained for the five measures.

In FLUX, multiple indicators were only available for the Big five dimensions of extraversion and agreeableness and for positive and negative affect. Again, the variability scores based on the average of the mean-corrected standard deviations and the variability scores based on the mean-corrected standard deviation of the average were significantly correlated, extraversion: \( r = 0.71, p < 0.01 \); agreeableness: \( r = 0.78, p < 0.01 \); conscientiousness: \( r = 0.93, p < 0.01 \); negative affect: \( r = 0.93, p < 0.01 \). Table A1b presents the correlations (and partial correlations) we obtained for the five measures.

### Table A1a. Correlations and partial correlations between response style measures, variability measures and well-being outcome measures in FLIP

| Measure | D.-N | D.-E | D.-O | D.-A | D.-C |
|---------|------|------|------|------|------|
| SCC T1  | .12  | .23  | .06  | .11  |      |
| SCC T2  | .13  | .25  | .03  | .04  |      |
| NO T1   | .24  | .20  | .35  | .22  |      |
| NO T2   | .23  | .10  | .17  | .16  |      |
| ST T1   | .31  | .38  | .28  | .38  |      |
| ST T2   | .25  | .30  | .37  | .31  |      |
| SE      | -.26 | -.36 | -.18 | -.09 | -.30 |
| LS      | -.21 | -.32 | -.11 | -.09 | -.14 |
| PA      | -.12 | -.32 | -.12 | -.03 | -.18 |
| NA      | -.22 | -.35 | -.23 | .11  | -.22 |
| JS      | -.06 | -.17 | -.06 | -.06 | -.13 |
| LO      | -.15 | -.28 | -.12 | -.04 | -.21 |

Note. Correlations and partial correlations in bold are significant at \( p < .05 \); \( N = 96 \) or \( N = 94 \) (in case of partial correlations). SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, SE = self-esteem, LS = Life satisfaction, PA = Positive affect, NA = Negative affect, D.- N = Daily neuroticism, D.- E = Daily extraversion, D.- O = Daily openness, D.- A = Daily agreeableness, D.- C = Daily conscientiousness.

### Table A1b. Correlations and partial correlations between response style measures, variability measures and well-being outcome measures in FLIP

| Measure | D.-E | D.-A | D.-PA | D.-NA |
|---------|------|------|-------|-------|
| SCC     |      |      |       |       |
| NO      |      |      |       |       |
| ST      |      |      |       |       |
| SE      | -.26 | -.36 | -.18  | -.09  |
| LS      | -.21 | -.32 | -.11  | -.09  |
| PA      | -.12 | -.32 | -.12  | -.03  |
| NA      | -.22 | -.35 | -.23  | .11   |
| JS      | -.06 | -.17 | -.06  | -.06  |
| LO      | -.15 | -.28 | -.12  | -.04  |

Note. Correlations and partial correlations in bold are significant at \( p < .05 \); \( N = 96 \) or \( N = 94 \) (in case of partial correlations). SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, SE = self-esteem, LS = Life satisfaction, PA = Positive affect, NA = Negative affect, D.- N = Daily neuroticism, D.- E = Daily extraversion, D.- A = Daily agreeableness, D.- PA = Daily positive affect, D.- NA = Daily negative affect.

### Appendix 1

In this appendix, we report the results for all analyses when using another approach to obtain the intradividual variability scores for the substantive variables. This approach consisted of (a) aggregating the daily diary items per substantive dimension (e.g., the extraversion items) and (b) computing a (mean-corrected) standard deviation using this aggregate. Note that this approach can only be used when multiple indicators for a substantive dimension have been assessed; otherwise, it is identical to the approach whose results are reported in the main text.

In FLIP, multiple indicators were available for each of the Big five dimensions. For each dimension, the variability scores based on the average of the mean-corrected standard deviations and the variability scores based on the mean-corrected standard deviation of the average were significantly correlated, neuroticism: \( r = 0.71, p < 0.01 \); extraversion: \( r = 0.64, p < 0.01 \); openness: \( r = 0.93, p < 0.01 \); agreeableness: \( r = 0.80, p < 0.01 \); conscientiousness: \( r = 0.75, p < 0.01 \). Table A1a presents the correlations (and partial correlations) we obtained for the five measures.
Appendix 2

In this appendix, we report the results for all analyses when we use the relative variability index of Mestdagh et al. (2018) to obtain the intraindividual variability scores for the substantive variables.

In FLIP we found that the relative variability index for a variable was highly correlated with the mean-corrected standard deviation of the variable: neutral object rating task: \( r = 0.98, p < .01 \); Simpson task: \( r = 0.98, p < .01 \); neuroticism: \( r = 0.96, p < .01 \); extraversion: \( r = 0.97, p < .01 \); openness: \( r = 0.92, p < .01 \); agreeableness: \( r = 0.90, p < .01 \); conscientiousness: \( r = 0.97, p < .01 \), overall personality: \( r = 0.97, p < .01 \), and self-esteem: \( r = 0.91, p < .01 \). Table A2a presents the reliabilities and the correlations between the different response style and intraindividual variability measures.

Furthermore, Table A2b presents the correlations (and partial correlations) we obtained for the associations between the response style and intraindividual variability measures, respectively, and the outcome measures.

The relative variability indices were also highly correlated with the mean-corrected standard deviations in FLUX: neutral object rating task T1: \( r = 0.97, p < .01 \); neutral object rating task T2: \( r = 0.99, p < .01 \); Simpson task T1: \( r = 0.98, p < .01 \); Simpson task T2: \( r = 0.98, p < .01 \); neuroticism: \( r = 0.90, p < .01 \); extraversion: \( r = 0.94, p < .01 \); openness: \( r = 0.95, p < .01 \); agreeableness: \( r = 0.84, p < .01 \); conscientiousness: \( r = 0.94, p < .01 \), overall personality: \( r = 0.96, p < .01 \), and self-esteem: \( r = 0.97, p < .01 \). Table A2c presents the reliabilities and the correlations between the different response style and intraindividual variability measures.

Furthermore, Table A2d presents the correlations (and partial correlations) we obtained for the associations between the response style and intraindividual variability measures, respectively, and the outcome measures.
Table A2a. Reliabilities of and correlations between different response style measures and different intraindividual variability measures in FLIP

| Measures | 1.   | 2.   | 3.   | 4.   | 5.   | 6.   | 7.   | 8.   | 9.   | 10.  |
|----------|------|------|------|------|------|------|------|------|------|------|
| SCC      | (0.90) |      |      |      |      |      |      |      |      |      |
| NO       | -.21 | (0.73) |      |      |      |      |      |      |      |      |
| ST       | .03  | .26  | (0.83) |      |      |      |      |      |      |      |
| Daily N  | -.04 | .33  | .31  | (0.87) |      |      |      |      |      |      |
| Daily E  | .06  | .27  | .44  | .83  | (0.88) |      |      |      |      |      |
| Daily O  | .09  | .13  | .42  | .66  | .83  | (0.92) |      |      |      |      |
| Daily A  | -.02 | .23  | .31  | .81  | .81  | .74  | (0.80) |      |      |      |
| Daily C  | .16  | .18  | .37  | .72  | .80  | .82  | .76  | (0.91) |      |      |
| Daily Pers. | .05 | .25  | .39  | .88  | .94  | .88  | .93  | .89  | (0.96) |      |
| Daily SE | .14  | .20  | .30  | .64  | .67  | .58  | .66  | .67  | .72  | (0.82) |

Note. Reliability coefficients are reported in parentheses in the table's diagonal. Correlations in bold are significant at \( p < .05 \); \( N = 96 \). SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures are intraindividual variability measures.

Table A2b. Correlations and partial correlations between response style measures, interindividual variability measures and well-being outcome measures in FLIP

| Measure       | SE   | LS   | PA   | NA   | SE   | LS   | PA   | NA   |
|---------------|------|------|------|------|------|------|------|------|
| NO            | .15  | -.01 | .09  | -.13 |      |      |      |      |
| ST            | -.08 | -.13 | .01  | .03  |      |      |      |      |
| Daily N       | -.07 | -.03 | .12  | -.06 |      |      |      |      |
| Daily E       | -.22 | -.13 | -.01 | -.09 |      |      |      |      |
| Daily O       | -.26 | -.21 | -.09 | -.02 |      |      |      |      |
| Daily A       | -.17 | -.10 | -.01 | -.04 |      |      |      |      |
| Daily C       | -.29 | -.24 | -.09 | .00  |      |      |      |      |
| Daily Pers.   | -.22 | -.14 | -.02 | -.05 |      |      |      |      |
| Daily SE      | -.31 | -.19 | -.07 | .08  |      |      |      |      |

Note. Correlations and partial correlations in bold are significant at \( p < .05 \); \( N = 96 \) or \( N = 94 \) (in case of partial correlations). NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem, LS = Life satisfaction, PA = Positive affect, NA = Negative affect. SCC, NO, and ST are response style measures while all other measures in the rows are intraindividual variability measures.
Table A2c. Reliabilities of and correlations between different response style measures and different intraindividual variability measures in FLUX

| Measure       | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. | 15. |
|---------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| 1. SCC T1    |    |    |    |    |    |    |    |    | .92 |     |     |     |     |     |     |
| 2. SCC T2    | .83 |    |    |    |    |    |    |    |     | (.91) |     |     |     |     |     |
| 3. NO T1     | -.19 | -.12 |    |    |    |    |    |    |     |     | (.65) |     |     |     |     |
| 4. NO T2     | -.22 | -.14 |    |    | .63 |    |    |    |     |     |     | (.58) |     |     |     |     |
| 5. ST T1     | -.10 | -.06 | .01 | .19 |    | .83 |    |    |     |     |     |     | (.88) |     |     |     |
| 6. ST T2     | -.03 | .00 |    | .17 | .39 | .65 |    |    |     |     |     |     |     | (.79) |     |     |
| 7. Daily N   | -.13 | -.12 | .43 | .32 | .27 | .31 |    |    |     |     |     |     |     |     | (.79) |     |
| 8. Daily E   | -.05 | -.07 | .31 | .21 | .38 | .42 | .62 |    |     |     |     |     |     |     | (.84) |     |
| 9. Daily O   | -.13 | -.08 | .35 | .18 | .40 | .37 | .70 | .74 |     |     |     |     |     |     | (.63) |     |
| 10. Daily A  | -.07 | -.09 | .34 | .26 | .36 | .41 | .68 | .75 | .72 |     |     |     |     |     |     | (.69) |
| 11. Daily C  | -.11 | -.09 | .33 | .25 | .42 | .45 | .65 | .75 | .74 | .68 |     |     |     |     |     | (.72) |
| 12. Daily Per.| -.06 | -.06 | .44 | .30 | .39 | .44 | .79 | .93 | .85 | .91 | .84 |     |     |     |     | (.91) |
| 13. Daily PA | -.05 | -.07 | .38 | .27 | .39 | .45 | .75 | .79 | .85 | .76 | .79 | .89 |     |     |     | (.91) |
| 14. Daily NA | -.01 | -.04 | .30 | .32 | .38 | .43 | .52 | .74 | .54 | .80 | .57 | .80 | .65 |     |     | (.79) |
| 15. Daily SE | -.05 | -.02 | .31 | .15 | .34 | .35 | .61 | .68 | .76 | .57 | .65 | .72 | .76 | .39 |     | (.70) |

Note. Reliability coefficients are reported in parentheses in the table's diagonal. Correlations in bold are significant at \( p < .05 \); \( N = 93 \). SCC = Self-concept clarity, NO = Neutral objects, ST = Simpsons task, N = neuroticism, E = extraversion, O = openness, A = agreeableness, C = conscientiousness, SE = self-esteem. SCC, NO, and ST are response style measures while all other measures are intraindividual variability measures.
Table A2d. Correlations and partial correlations between response style measures, interindividual variability measures and well-being outcome measures in FLIP

| Measure     | SE  | LS  | PA  | NA  | JS  | LO  | SE  | LS  | PA  | NA  | JS  | LO  |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| NO T1       | .00 | .02 | .02 | .04 | .20 | .01 |     |     |     |     |     |     |
| NO T2       | .15 | .16 | .09 | -.10| .20 | .08 |     |     |     |     |     |     |
| ST T1       | -.02| -.08| .07 | .07 | .03 | .06 |     |     |     |     |     |     |
| ST T2       | -.07| -.08| .07 | .10 | -.05| -.05|     |     |     |     |     |     |
| Daily N     | -.03| .10 | .10 | .10 | .11 | .05 | -.03| .13 | .11 | .08 | .04 | .04 |
| Daily E     | -.04| -.03| .06 | .04 | .00 | .08 | -.03| -.01| .05 | .01 | -.06| .06 |
| Daily O     | -.06| -.10| .00 | .18 | -.08| -.12| -.06| -.08| -.03| .17 | -.16| -.18|
| Daily A     | -.04| -.01| .02 | .02 | .06 | -.06| -.04| .01 | .00 | -.02| .00 | -.10|
| Daily C     | -.11| -.14| .03 | .14 | -.02| -.04| -.11| -.13| .01 | .11 | -.08| -.08|
| Daily Pers. | -.09| -.05| -.04| .10 | .01 | -.01| -.09| -.04| -.06| .07 | -.08| -.05|
| Daily PA    | -.09| -.05| .03 | .17 | -.02| -.06| -.09| -.03| .01 | .16 | -.10| -.10|
| Daily NA    | -.01| -.04| .06 | -.05| .04 | .00 | .00 | -.02| .05 | -.10| -.01| -.03|
| Daily SE    | -.13| -.09| -.06| .24 | -.05| -.11| -.14| -.08| -.08| .23 | -.12| -.15|

On the Role of Response Styles in the Study of Intraindividual Variability

Collabra: Psychology
SUPPLEMENTARY MATERIALS

Peer Review History
Download: https://collabra.scholasticahq.com/article/29929-on-the-role-of-response-styles-in-the-study-of-intraindividual-variability/attachment/75657.docx?auth_token=4xUuGy9bEEmoyxL_BPTx