Check for updates

What Happens When You Add a 'Not Relevant' Response Option to the Unipolar Response Scales of Personality State Items?

Sarah Kritzler [1] ⒾⒹ , Peter Haehner [1] ⒾⒹ , Julia Krasko [1] ⒾⒹ , Susanne Buecker [1,2] ⒾⒹ

[1] *Department of Psychology, Ruhr University Bochum, Bochum, Germany.* **[2]** *Institute of Psychology, German Sport University Cologne, Cologne, Germany.*

Corresponding Author: Sarah Kritzler, Ruhr University Bochum, Department of Psychology, Universitätsstrasse 150, 44780 Bochum, Germany. E-mail: Sarah.Kritzler@rub.de

Abstract

What happens when you add a "not relevant" response option to the unipolar response scales of personality state items? In an experimental experience sampling study with a between-person design (total N = 248; n = 3,253 observations), we compared personality states measured with a unipolar response scale including or not including a "not relevant" response option. Overall, "not relevant" responses were quite prevalent but varied between items. Certain characteristics of the situation (particularly sociality) but not of the person predicted the use of the "not relevant" response option. Additionally, means and distributions of personality states significantly differed between the different response scales, but their associations with other relevant constructs did not. Overall, this study emphasizes the importance of systematically addressing how personality states should be measured and provides first evidence that a "not relevant" response option might be an important aspect to consider for the measurement of personality states.

Keywords

Relevance Statement

Anecdotal evidence suggests that participants sometimes perceive personality states as irrelevant in a given situation. This study is the first to empirically examine the utility of a "not relevant" response option for personality state scales.

Key Insights

- "Not relevant" responses to personality state items were quite prevalent
- The prevalence varied considerably between the different items
- Situation characteristics (but not personality traits) predicted "not relevant" responses
- Personality state mean levels differed between different response scales
- Convergent associations did not differ between different response scales

The conceptualization of personality as both trait-like and state-like has contributed tremendously to the understanding of the dynamics in people's thoughts, feelings, and behaviors (e.g., Hampson, 2012; Jayawickreme et al., 2019). *Personality states* are momentary manifestations of personality traits and describe how people think, feel, and behave at any given moment (Baumert et al., 2017). They provide important insights into the manifestation and variability of personality in everyday life (e.g., Geukes et al., 2017; McCabe & Fleeson, 2016).

However, there is little consensus on how personality states should be measured. Instead, personality state measures are usually created ad-hoc and have differed greatly regarding their items, instructions, and response scales (Horstmann & Ziegler, 2020). A commonality of almost all these measures is that participants are expected to be able to assess their personality states in any situation. However, participants may sometimes perceive certain states as irrelevant in certain situations and therefore have trouble responding to personality state items. Yet, this aspect of the measurement of personality states has received little attention.

The current study addresses this gap in the literature by exploring the usefulness and psychometric consequences for estimates of reliability and validity of including a "not relevant" response option in a unipolar response scale for personality state items. To do so, we conducted an experimental experience sampling study in which we randomly assigned participants to one of two conditions. In one condition, participants responded to the personality state measure using a unipolar response scale including a "not relevant" response option. In the other condition, this response option was not included (i.e., traditional personality state measure).

Measurement of Personality States

Although personality states became popular in personality psychology about 20 years ago (Fleeson, 2001; but see also Cattell, 1946), a validated questionnaire still does not

exist, nor are there many specific recommendations for measuring these personality states (Horstmann & Ziegler, 2020; Ringwald et al., 2022). Items measuring personality states are often created ad-hoc and have taken many different forms. For example, personality states have been measured concerning different time frames (e.g., the last hour, the current moment) and on different response scales such as unipolar Likert scales or bipolar scales using two adjectives as anchors.

Another important question regarding the measurement of personality states is the relevance of personality state items: Should personality state items be offered with an additional "not relevant" response option to represent situations in which participants may find it difficult to assess a given state? This issue could be particularly important in practice because it may affect participants' experiences with the scale. For example, when piloting personality state items with a typical Likert-type scale, several participants reported difficulties answering certain items in certain situations. These participants argued that it was not possible to be like this in the given situation (e.g., "I was just cooking by myself - how am I supposed to cook empathetically?"). Thus, a response option representing this remark might increase the ease of using the scale. Indeed, some studies already included "not applicable" response options in personality state measures (e.g., Fleeson & Gallagher, 2009; McCabe & Fleeson, 2016), but they did not address the consequences of including it (e.g., how frequently the option was used). Thus, a detailed examination of the "not relevant" response option for personality state items is still missing.

Current State of Research on "Not Relevant" and Similar Response Options

Non-substantial response options such as "don't know," "no opinion," or "not applicable" in survey questions have been the subject of considerable research (for detailed discussions, see Krosnick & Presser, 2010; Menold & Bogner, 2016). Typical arguments for using such response options are that they signal to participants that it is OK not to have a substantive answer and that they distinguish this response from the other scale points (Menold & Bogner, 2016). Arguments against using a "don't know" or "no opinion" response option are that they might encourage satisficing (i.e., offering responses that seem reasonable without any memory search or information integration) and that people with socially undesirable opinions, weak opinions, or poor knowledge of a topic might be tempted to (over-)use them (Krosnick et al., 2002; Krosnick & Presser, 2010). Although empirical research has produced mixed results, the trend in recent years has been to move away from such response options.

However, most of the research on such non-substantial response options comes from attitude research (e.g., political or environmental attitudes) and, therefore, may not be directly generalizable to personality research. Such response options have received far less attention in personality research. One notable exception is a study by Kulas et al.

(2008) on bipolar personality trait items. They found that people tended to use the middle category as a proxy for irrelevance in the absence of a corresponding response option. Such behavior could affect the level of measurement because an interval or even ordinal level of scaling may no longer be guaranteed when the middle category represents both moderate standings on the underlying scale and the irrelevance of the item. Although this ambiguity of the middle category did not negatively affect the reliability and validity of the measures (potentially because of the low frequency in this study), the authors explicitly recommended using a "not applicable" response option in personality measures (Kulas et al., 2008). Therefore, this option should also be considered for personality state measures.

Methodological Perspectives on a "Not Relevant" Response Option for Personality State Items

Assuming that participants perceive certain personality states as irrelevant in certain situations, several problems may arise from a methodological point of view if such a response option is not offered. Comparing personality trait measures to personality state measures, any problems of applicability (e.g., Kulas et al., 2008) are likely more pronounced for personality state measures: Whereas personality trait measures ask for average tendencies allowing irrelevant situations to be excluded when considering this average, personality state measures ask about one specific situation and thus irrelevant situations cannot be excluded.

Importantly, such problems are not limited to bipolar scales where the middle category might act as a stand-in for irrelevance. They also apply to unipolar scales where the lowest response option might be used to indicate irrelevance. For example, the common practice of including negatively worded items (e.g., "quiet" for extraversion), which need to be reverse-coded before analyses, can cause problems if the lowest response option is used to indicate irrelevance. Due to the reverse coding, a low score on such an item is interpreted as a high score on the state. For example, selecting "1 = not at all" on the "quiet" item would be reverse-coded to a 5 on the overall extraversion scale. Therefore, a "not relevant" response using the low end of the scale for such a negative item is then interpreted as high levels of the underlying state. This issue may bias scale scores and further statistics calculated using such items.

The Current Study

Although a "not relevant" response option may be methodologically and practically important for measuring personality states, an empirical investigation of the psychometric consequences and usefulness of providing such a response option is still needed. In the current study, we explored differences between personality states measured with and without a "not relevant" response option in unipolar response scales. To this end, we

addressed the following research questions: (1) Given a "not relevant" response option, how often and when do people use it? (2) How do people respond to personality state items if this option is unavailable? (3) Do psychometric properties of personality state items differ between scales with and without a "not relevant" response option?

Method

Study Design

We used the experience sampling method (ESM) to sample personality states in people's everyday lives. This study had a randomized between-person design with two experimental groups: Participants in the *not-relevant group* were offered the additional "not relevant" response option for the personality state items; participants in the *treatment-as-usual group* were not offered this response option. The two experimental conditions did not differ in any other way. Neither participants nor researchers in contact with participants were aware of the experimental groups assigned to each participant.

Participants

We aimed to recruit at least 100 participants per group. This number was determined by a trade-off between time constraints and the approximate sample size needed to reliably estimate multilevel models for longitudinal data. To ensure that this lower limit was met, we stopped recruiting on the day on which at least 100 participants had begun the ESM phase in each group.

A total of 277 participants started the study. Of these, 248 also filled out at least one ESM survey and were thus included in the final sample (N = 116 in the not-relevant group and N = 132 in the treatment-as-usual group). Participants were on average 26.40 years old (SD = 6.83), and 25 of them identified as male, 219 as female, and 4 as diverse. Overall, we collected 3,253 ESM surveys (M = 14.44 per participant, SD = 5.91, Min = 1, Max = 21).

Procedure

The data were collected in September and October 2021 using formr (Arslan et al., 2020). Participants were recruited online at German universities and were offered individual personality feedback and course credit as compensation. After providing informed consent, participants filled out the baseline survey assessing personality traits and demographic information. During the following three-day ESM phase, participants were invited via e-mail to the ESM surveys seven times each day (between 8am and 8pm). The ESM surveys included questions on the current activity, personality states, perceived situation characteristics, and state affect.

PsychOpen GOLD

Materials

Table 1 provides an overview of all the measures relevant to this study. In the following, only the measurement of personality states is discussed in more detail.

Questions Regarding Personality States

Big Five personality states were measured with an ad-hoc created German adjective scale adopting adjectives from Rüegger et al. (2020). Participants indicated on a scale from 1 (*not at all*) to 5 (*totally*) how they had perceived themselves in the previous situation. Participants in the "not relevant" group were additionally presented with a sixth response option labeled "not relevant". Each personality state was measured using one positively and one negatively worded adjective (e.g., "open-minded" and "uninterested" for state openness). To examine their relevance both on the item level and on the personality state level, personality state scores were calculated by averaging the two items belonging to the same state (given that both items were answered).

Afterwards, participants were asked to freely explain their response for one randomly selected personality state item and to indicate on a dichotomous scale whether it was possible to behave, think, and feel as described in the personality state items in the experienced situation for both items of a randomly selected personality state.

Data Analysis

Analyses were conducted with R (Version 4.1.1; R Core Team, 2021) and the R-packages *lme4* (Version 1.1.30; Bates et al., 2015), *papaja* (Version 0.1.1; Aust & Barth, 2020), *psych* (Version 2.2.5; Revelle, 2021), and *tidyverse* (Version 1.3.2; Wickham et al., 2019). We set the significance level to α = .01 for all analyses as preregistered.

Transparency, Openness, and Reproducibility

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The research questions and analytic strategy were preregistered on the *Open Science Framework* before analyzing the data. In a few cases, we had to deviate from the preregistered analyses, mostly due to convergence issues of the models, and explicitly explain them with the respective results.

Study materials, data, analysis scripts, and a reproducible manuscript are available on the Open Science Framework (https://osf.io/qjyb3/). We also provide comprehensive online Supplementary Materials detailing analytic strategies, deviations from the preregistration, additional analyses, and comprehensive results from all statistical models.

Table 1

Overview of the Relevant Measures

Measure	Instrument / Description	Sample Item	Rating Scale	Number of Items	Scale Scores	Additional Information
Baseline Survey						
Personality Traits	BFI-2-S (German version: Danner et al., 2019)	I am someone who is outgoing, sociable.	1 (*not at all*) to 5 (*totally*)	30 (6 per trait)	Mean of all items (if at least three were answered)	Internal Consistencies: α = .70 (O), α = .75 (C), α = .76 (E), α = .60 (A), α = .80 (ES)
Demographic Information	Age, gender, experience with psychological studies	How much experience do you have participating in online studies?	—	8	—	—
ESM Survey						
Situation Markers	Current activity, whereabouts, and interaction Partners	What were you just doing?	Multiple choice from 20 activities, 6 whereabouts, and 12 interaction partners	3	—	Response options were inspired by Richter and Schupp (2015) and Hudson et al. (2020)
Situation Perception	S8-I (Rauthmann & Sherman, 2018)	A job needs to be done.	1 (*not at all*) to 5 (*totally*)	8 (1 per characteristic)	—	—
Personality States	Ad-hoc adjective scale	Please indicate how you perceived yourself in the situation you just described: uninterested	1 (*not at all*) to 5 (*totally*), not relevant	10 (2 per state)	Mean of all items (if both were answered)	Adjectives were inspired by Rüegger et al. (2020)
Personality State Explanation	Ad-hoc item	Please briefly describe why you chose this answer.	open response format	1 out of 10 (randomly chosen) in each survey	—	—

Measure	Instrument / Description	Sample Item	Rating Scale	Number of Items	Scale Scores	Additional Information
Potential for Personality States	Ad-hoc dichotomous scale	Was it possible to be uninterested in the situation you just described?	0 (*yes*), 1 (*no*)	2 out of 10 (randomly chosen) in each survey	—	—
State Affect	Ad-hoc adjective scale	Right now, I feel good.	1 (*not at all*) to 5 (*totally*)	4	Mean of all items (if at least two were answered)	—
Diligence	Ad-hoc item	How thoroughly did you fill out this questionnaire?	1 (*not at all*) to 5 (*totally*)	1	—	—

Note. O, C, E, A, and ES represent the Big Five personality traits Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Sample items and response scales were translated from the original German wordings used in the study. The order of the measures in the table represents the actual order in the survey. Complete instructions, item texts, and response scales can be found in the codebook in the project repository on the Open Science Framework: https://osf.io/qjyb3/

Results

Descriptive statistics and intercorrelations of all study variables can be found in the Supplementary Materials. Group comparisons demonstrated that the two experimental groups did not significantly differ in their personality traits or demographic characteristics (see Supplementary Materials).

Given a "Not Relevant" Response Option, How Often Do People Use it?

Examining descriptive statistics in the not-relevant group, we found that "not relevant" responses were chosen 19% of the time across all items. Two-sample chi-squared tests of independence revealed that the prevalence of "not relevant" responses varied significantly between the different personality states and between the two items belonging to each state (Table 2).

Table 2

Frequency and Percentage of "Not Relevant" Responses for Each Item and Personality State in the Not-Relevant Group

Personality State	Item	Frequency	Relative Frequency	Chi-Squared Test
Overall		3,080	19%	
State Level				
Openness		694	21%	$\chi^2(4) = 811.01$, $p < .001$,
Conscientiousness		739	22%	Cohen's $\omega = 0.22$
Extraversion		346	11%	
Agreeableness		1,044	32%	
Emotional Stability		257	8%	
Item Level				
Openness	open-minded	483	29%	$\chi^2(1) = 134.80$, $p < .001$,
	uninterested	211	13%	Cohen's $\omega = 0.20$
Conscientiousness	imprudent	322	20%	$\chi^2(1) = 15.35$, $p < .001$,
	conscientious	417	25%	Cohen's $\omega = 0.07$
Extraversion	quiet	29	2%	$\chi^2(1) = 266.69$, $p < .001$,
	sociable	317	19%	Cohen's $\omega = 0.29$
Agreeableness	dismissive	437	27%	$\chi^2(1) = 39.80$, $p < .001$,
	empathetic	607	37%	Cohen's $\omega = 0.11$
Emotional Stability	insecure	179	11%	$\chi^2(1) = 42.34$, $p < .001$,
	even-tempered	78	5%	Cohen's $\omega = 0.11$

Note. Items were translated from the original German wordings used in the study.

PsychOpen GOLD

In a second step, we examined interindividual differences in the usage of the "not relevant" response option. The prevalence of "not relevant" responses varied considerably between persons (average within-person frequency M = 18%, SD = 13%). We estimated linear regression models to examine whether demographic variables (i.e., age, gender, education), experience with online studies, or Big Five personality traits predicted the within-person aggregated frequency of "not relevant" responses. However, we did not find any significant person-level predictors of choosing the "not relevant" response option. For example, neither experience with online studies, b = -1.03, 99% CI [-4.56, 2.49], t(101) = -0.77, p = .443, nor conscientiousness, b = -0.95, 99% CI [-6.65, 4.75], t(101) = -0.44, p = .663, were associated with the frequency of "not relevant" responses within persons (full model results can be found in the Supplementary Materials).

Given a "Not Relevant" Response Option, When Do People Use it?

For this research question, we first examined whether participants in the not-relevant group used the "not relevant" response option when they later indicated that it was impossible to behave, think, and feel this way. Two-sample chi-squared tests of independence revealed significant associations between choosing the "not relevant" response option and indicating that it was impossible to be this way for all personality states (all χ^2 ≥ 164.90, all p < .001) except emotional stability, χ^2(1) = 0.02, p = .899. For example, risk ratios indicated that the probability of responding with "not relevant" to state openness items increased by the factor of 3.73, 99% CI [2.67, 5.21], when participants had indicated that it was impossible to be this way. Moreover, when asked to freely explain their responses, participants often argued that it was not possible, not relevant, or not required to behave in this way when they had chosen the "not relevant" response option (see Supplementary Materials).

Second, we examined whether aspects of the situation predicted usage of the "not relevant" response option. We used a stepwise multilevel logistic modeling approach by adding diligence as a potentially method-related predictor in a first step, categorical situation markers in a second step, and dimensional situation perceptions in a third step. First, we found that self-reported diligence was not associated with the probability of "not relevant" choices. Second, we included categorical situation markers (i.e., people's activities, whereabouts, and interactions partners; results of this intermediate step can be found in the Supplementary Materials). Finally, we included dimensional situation perceptions in a third step and found that perceived sociality was significantly associated with lower chances of "not relevant" responses for almost all personality state items (see Table 3 and Figure 1). More detailed model results can be found in the Supplementary Materials.

Table 3

Significant Predictors of "Not Relevant" Responses for Each Personality State Item From the Final Multilevel Logistic Regression Models

Predictor	Estimate	SE	z	p	Risk Ratio	99% CI for RR
Openness: open-minded						
Sociality	-0.67	0.09	-7.76	< .001	0.62	[0.50, 0.76]
Openness: uninterested						
Activity [Relaxing or eating]	2.79	0.72	3.88	< .001	4.54	[2.02, 5.65]
Activity [Housework or childcare]	2.14	0.69	3.10	.002	3.75	[1.34, 5.39]
Mating	0.29	0.10	2.77	.006	1.26	[1.02, 1.54]
Sociality	-0.54	0.11	-4.81	< .001	0.63	[0.48, 0.81]
Extraversion: sociable						
Sociality	-0.57	0.09	-5.98	< .001	0.62	[0.50, 0.76]
Agreeableness: dismissive						
Sociality	-0.69	0.09	-7.99	< .001	0.52	[0.42, 0.64]
Agreeableness: empathetic						
Diligence	0.63	0.23	2.73	.006	1.60	[1.03, 2.32]
Activity [Social activities]	-3.45	1.14	-3.02	.003	0.04	[0.00, 0.65]
Adversity	-0.81	0.23	-3.59	< .001	0.50	[0.29, 0.83]
Sociality	-0.51	0.08	-6.18	< .001	0.65	[0.54, 0.78]
Emotional Stability: insecure						
Positivity	-0.37	0.14	-2.69	.007	0.70	[0.49, 0.98]
Negativity	-0.44	0.16	-2.75	.006	0.66	[0.44, 0.97]
Sociality	-0.41	0.10	-4.03	< .001	0.67	[0.52, 0.87]

Note. Estimate represents logits. The final multilevel logistic regression models included all categorical situation markers, perceived situation characteristics (within-person centered), and diligence as predictors. The table only presents significant predictors from these models, full model results can be found in the Supplementary Materials. The personality state items imprudent, conscientious, quiet, and even-tempered are missing from this table because none of their models converged.

Figure 1

Perceived Sociality and Social Activities Significantly Predicted the Probability of "Not Relevant" Responses to the Agreeableness Item "Empathetic"
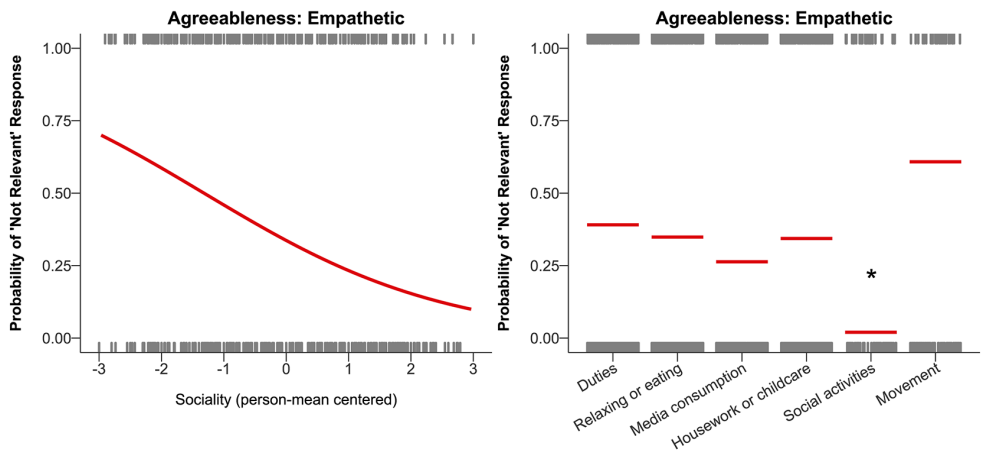

Note. Figures representing other significant predictors of "not relevant" responses in other items can be found in the Supplementary Materials.

How Do People Respond to Personality State Items When a Not Relevant Response Option is Not Available?

For this research question, we focused only on the treatment-as-usual group, which did not receive a "not relevant" response option. We used a two-sample chi-squared test of independence to explore whether there was a significant association between indicating that it was impossible to be a certain way and choosing certain response options from the personality state items. Indeed, the responses to personality state items significantly differed between possible and impossible cases, $\chi^2(4) = 263.35$, $p < .001$. This difference was due to people choosing lower response options more often when they indicated that it was not possible to be this way (Figure 2). Interestingly, the participants showed this behavior for both positively and negatively worded items (Figure 2B and 2C), which means that "it is not possible to be this way" biases scale scores towards smaller values for positive items and towards larger values for negative items.

Figure 2

Response Distributions of Personality State Items Depending on Whether or Not Participants Indicated That the State Was Possible in the Treatment-As-Usual Group
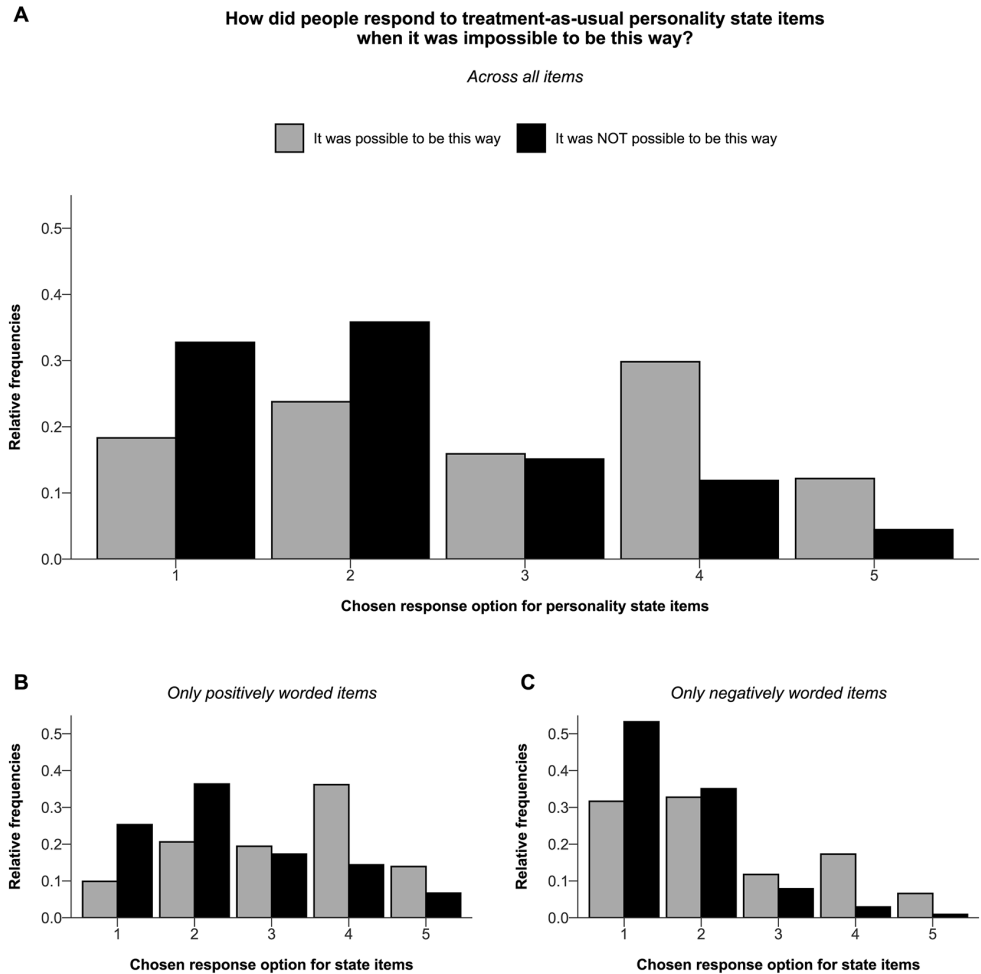

Do Psychometric Properties of Personality States Scales Differ Between the Two Response Format?

For this research question, we compared personality state scores between the two groups that did and did not receive a "not relevant" response option (Table 4). First, Welch's *t*-tests showed that participants in the "not relevant" group reported on average higher mean levels of state openness, state conscientiousness, and state agreeableness than participants in the treatment-as-usual group (Figure 3A). Second, Kolmogorov-Smirnov tests

showed that the distributions of personality state scores significantly differed between the two groups, such that personality state scores cannot be assumed to come from the same population (Figure 3B). Third, however, within-person variances compared using Wilcoxon rank-sum tests did not significantly differ between participants from the two groups.

Figure 3

Comparison of Aggregated Within-Person Means (A) and Distributions (B) of State Agreeableness Scores Between the Two Experimental Groups
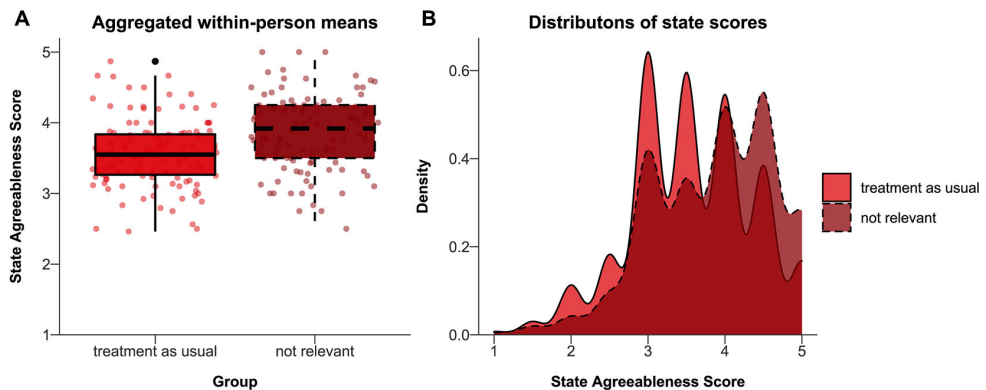
Note. Figures comparing means and distributions of the other personality states can be found in the SSupplementary Materials.

Next, we compared internal consistencies between the personality state measures (Table 4). Here, we had to deviate from the preregistered analysis plan to compare nested omega coefficients because of various estimation problems in the multilevel confirmatory factor analysis (see Supplementary Materials for details on these models). Instead, we report additional results comparing Cronbach's alpha coefficients. Cronbach's alpha coefficients did not systematically vary between the two groups, as neither group consistently had higher coefficients than the other. As one exception, Cronbach's alpha of state conscientiousness items was significantly higher in the not-relevant group (α = .50) than in the treatment-as-usual group (α = .33), $\chi^2(1)$ = 12.62, p < .001. However, note that Cronbach's alpha does not consider the nested data structure of personality state items.

Table 4

Comparisons of Psychometric Properties of Personality State Scores Between the Two Groups

Personality State	Comparing aggregated within-person means					Comparing within-person variances		Comparing distributions		Comparing internal consistencies				
	M_{NR}	M_{TAU}	t	df	p	Cohen's d	ω	p	D	p	α_{TAU}	α_{NR}	χ^2_α	df_α	p_α
Openness	3.98	3.65	-21.60	3,574	< .001	-0.72	7,903.00	.048	0.17	< .001	0.49	0.48	0.06	1.00	.913
Conscientiousness	4.05	3.81	-14.47	3,536	< .001	-0.49	7,271.00	.375	0.13	< .001	0.33	0.50	12.62	1.00	< .001
Extraversion	2.63	2.51	-6.47	3,574	< .001	-0.22	5,760.50	.014	0.08	< .001	0.19	0.19	0.00	1.00	> .999
Agreeableness	3.90	3.57	-19.73	3,570	< .001	-0.66	6,604.50	.968	0.18	< .001	0.27	0.31	0.47	1.00	.492
Emotional Stability	3.99	3.89	-5.26	3,575	< .001	-0.18	7,327.00	.535	0.05	.039	0.66	0.59	5.22	1.00	.022

Note. α_{TAU} represents Cronbach's alpha coefficients in the treatment-as-usual group and α_{NR} represents Cronbach's alpha coefficients in the not-relevant group.

Do Associations Between Personality States and Related Constructs Differ Between the Scales?

Finally, we examined whether associations between personality states and related constructs (i.e., state affect, situation characteristics, and personality traits) differed between personality states measured with and without the "not relevant" response option. For this purpose, we estimated multilevel regression models in which the group status moderated the association between personality states and the other constructs. For example, in the case of group-by-trait interactions, we examined whether the convergence between the participants' average personality state levels and their personality trait levels varied as a function of whether the personality state scale had a "not relevant" response option.

We again used a stepwise approach, first estimating a model with only main effects of group status, state affect, situation characteristics, and personality traits and then adding interactions with group status. Deviance tests for model comparisons showed that adding the interactions between group status and state affect, situation characteristics, and personality traits to the models generally never improved model fit over and above models that only included their main effects (all χ^2 < 19.25, all p > .156; full model results can be found in the Supplementary Materials). Therefore, we did not evaluate the significance of the individual interaction effects anymore.

Post-hoc simulations revealed that our sample likely lacked the power to reliably detect the statistical significance of these interactions: On average, we could only detect interaction effects down to a size of 0.18 (thresholds varied between 0.14 and 0.28 depending on the personality state and the specific interaction effect being tested). However, inspecting the standardized effect sizes of the interaction coefficients revealed that the interactions were overall very small (M_{eff} = 0.05, SD_{eff} = 0.04) by conventional standards (e.g., Funder & Ozer, 2019) which questions their practical relevance regardless of the power to detect statistical significance.

Discussion

Are there situations in which participants believe that certain personality states are irrelevant? What happens when a "not relevant" response option is offered on a unipolar response scale for personality state items? To approach these questions, we conducted an experimental experience sampling study. We compared personality states measured with and without a "not relevant" response option on a unipolar response scale in a between-person design. Our study has five key findings: First, "not relevant" responses were quite prevalent in our sample, and the prevalence differed considerably between personality states and the items used to measure the same state. Second, characteristics of the situation—often related to social aspects—but not characteristics of the person (i.e., self-reported Big Five personality traits, experience with online studies, age, gender, or

education) predicted the use of the "not relevant" response option. Participants often (but not exclusively) seemed to use this response option when they felt it was impossible to behave, think, or feel this way. Third, participants in the group without the "not relevant" response option tended to use the lower end of the scale when they felt it was impossible to behave, think, or feel this way. Fourth, personality states measured with and without a "not relevant" response option significantly differed in their means and distributions but not in their variances and Cronbach's alpha coefficients. Fifth, associations between personality states and related constructs did not differ significantly between personality states measured with and without a "not relevant" response option.

Should a "Not Relevant" Response Option Be Routinely Included in Unipolar Response Scales for Personality State Measures?

To the best of our knowledge, this was the first study to examine the consequences of including a "not relevant" response option in unipolar response scales in personality state measures. As the first study on this topic, it cannot provide a definitive answer to whether a "not relevant" response option should be used in personality state measures. Nevertheless, the findings of this study provide first arguments for and against using such a response option. The findings also point out directions for future research to further understand the consequences of including a "not relevant" response option.

Our results suggest in several respects that it might be advisable to offer a "not relevant" response option in unipolar response scales of personality state measures: First, usage of the "not relevant" response option was strongly associated with whether participants indicated that it was possible to behave, think, and feel in this way. For example, when participants thought it was impossible to be empathetic, they tended to indicate that being empathetic was irrelevant. This finding was also confirmed by participants' qualitative explanations for why they selected the "not relevant" response option (see Supplementary Materials). Moreover, the usage of the "not relevant" response option was predicted by characteristics of the situation but not by diligence, previous experiences with studies, or participants' personality. Thus, whether participants selected "not relevant" seemed to depend mostly on the situational circumstances. Together, these findings point to a high validity of the "not relevant" response option to indicate situations in which personality states might be irrelevant.

Second, our findings showed that participants seemed to use the lower end of the scale as a proxy for "not relevant" if they did not see a dedicated "not relevant" response option. This behavior could be problematic because the low end of the scale might then simultaneously represent low levels of a state and a state being irrelevant. Moreover, participants also showed this behavior for negatively worded items which would be interpreted as high scores on a personality state through reverse coding. Finally, the frequency with which this response option was used suggests that it is important and helps

participants answer the questions. Thus, participants seemed to agree that personality states are sometimes irrelevant in their lay understanding.

However, two aspects of our findings might also argue against the "not relevant" response option. First, offering a "not relevant" response option means either that one has to accept more missing values by recoding "not relevant" responses into missing values or that the statistical modeling of these partially metric partially nominal data becomes more challenging (for examples on how to model such data, see Huggins-Manley et al., 2018; Loeys et al., 2012). Given the high prevalence of "not relevant" responses in this study, measuring personality states with a "not relevant" response option could thus lead to reduced power compared to traditional personality state measures. Second, we did not find any differences in associations between personality states and related constructs (e.g., personality traits, affect). One might argue that if these associations, which are often the focus of research on personality states, do not change when using a "not relevant" response option, it might not be important to use it.

Open Questions and Directions for Future Research

The findings from this study yield at least two important pathways for future research. First, the findings should be replicated and extended in further studies. For example, our results only apply to unipolar response scales. However, bipolar scales are also used similarly frequently to measure personality states (Horstmann & Ziegler, 2020)[1]. The usefulness and consequences of a "not relevant" response option might differ for bipolar scales. Therefore, it is important to replicate our study with a bipolar scale.

Moreover, future studies should investigate the response processes involved in choosing the "not relevant" response option and how participants interpret it. A rough inspection of qualitative explanations of responses to personality state items in our study showed that even though most participants used the category as we intended, there seemed to be some heterogeneity. Perhaps renaming the response option (e.g., "not applicable" or explicitly "not possible in this situation") or providing more detailed instructions would be helpful to standardize the usage and improve its usefulness. Finally, within-person designs or mixed between-within-person designs should be used to gain further insights into within-person associations when using a "not relevant" response option.

Second, further research on how personality states should generally be measured is needed. The findings from our study emphasize that carefully selecting personality state items is crucial and should therefore be made in an informed manner. Above all, this requires studies that systematically compare different personality state items and

---

1) Reviewing the personality states measures listed in Table 1 of Horstmann and Ziegler (2020)'s recent review on the measurement of personality states, we found that unipolar and bipolar scales were equally common in Big Five/HEXACO personality state measures.

PsychOpen GOLD

response scales and develop validated measures of personality states. Concerning the response scale, we found that the prevalence of "not relevant" responses varied considerably between the items used to measure each state. For example, the extraversion item "quiet" had the lowest prevalence of all items (2%), whereas "sociable" was much more frequently rated as not relevant (19%). Thus, item selection may also be important when discussing the relevance of personality states and the response format used to measure them. For example, our findings could indicate that it may be possible to develop items with a low prevalence of "not relevant" responses that therefore do not need this response option.

Theoretical Perspectives on the (Ir)Relevance of Personality States

Finding that participants frequently reported certain personality states as irrelevant in their everyday lives raises the question of theoretical perspectives on the relevance of personality states. From a theoretical perspective, one might argue that personality states are descriptions of people's momentary ways of thinking, feeling, and behaving using the contents of personality traits (Fleeson, 2001). Because people are always thinking, feeling, and behaving in some way, it should also always be possible to describe these thoughts, feelings, and behaviors with adjectives that characterize the personality domains. Whole Trait Theory, for example, proposes that different social-cognitive processes control the enactment of personality states and that the output of these processes are increases or decreases in certain personality states (Jayawickreme et al., 2019). Therefore, this theory assumes that personality states are always expressed to some level and that this level is only upregulated or downregulated but never shut off by these processes.

On the other hand, multiple theories stress that traits are contextualized and that their relevance for states of thinking, feeling, and behaving, therefore, depends on the situation (DeYoung, 2015; Tett & Guterman, 2000). For example, trait activation theory proposes that traits represent latent potentials that are only expressed in thoughts, feelings, and behaviors when activated by trait-relevant situational cues (Tett & Guterman, 2000). Similarly, Cybernetic Big Five Theory proposes that traits "require appropriate eliciting stimuli before they are manifested in behavior and experience" and "therefore, vary in their relevance across situations" (DeYoung, 2015, p. 35). Thus, when a trait is not activated by the situation, it will not be expressed in how people think, feel, and behave in this situation. Consequently, the respective personality state might be irrelevant.

Taken together, theoretical arguments can be made for the perspective that personality states can always be meaningfully expressed (and thus measured) and for the perspective that personality states may be irrelevant in certain situations. Our findings cannot differentiate between these perspectives because a personality state questionnaire including a "not relevant" response option could improve the measurement of personality states by improving clarity for participants or making the items more accurately reflective of reality (or both). Therefore, researchers concerned with empirical questions

involving personality states should further examine the theoretical aspects of the relevance of personality states.

Limitations

Some limitations of the present study must be considered. First, we used a convenience sample consisting largely of female psychology students, which is subject to the well-known problems of WEIRD samples (Henrich et al., 2010). Second, the data were collected during the Covid-19 pandemic but there were comparatively few restrictions at the time of data collection (e.g., large events allowed, no contact restrictions). Finally, although our sample size was quite typical for experience sampling studies, it may have lacked power for more complex analyses (e.g., multilevel regression models, multilevel structural equation models).

Conclusion

Our results provide first evidence that participants consider certain personality state items to be irrelevant in certain situations. We found sizable variability in the prevalence of "not relevant" responses between persons, between personality states, and between items—each raising new questions for future research. Additionally, we found that personality states measured with and without "not relevant" response options in unipolar response scales differed in some psychometric characteristics. Still, associations between personality states and related constructs did not differ between the two scales. Overall, this study emphasizes the importance of systematically addressing how personality states are measured. Researchers concerned with empirical questions involving personality states should clearly think about how personality states should be measured—including whether or not a "not relevant" response option should be provided—and aim to develop validated scales of personality states.

PsychOpen GOLD

Author Contributions: *Sarah Kritzler*—Idea, conceptualization | Design planning | Resource provision (materials, participants, etc.) | Research implementation (software, hardware, etc.) | Data collection | Data management (storage, curation, processing, etc.) | Visualization (data presentation, figures, etc.) | Data analysis | Writing | Project coordination, administration. *Peter Haehner*—Idea, conceptualization | Design planning | Resource provision (materials, participants, etc.) | Research implementation (software, hardware, etc.) | Data collection | Visualization (data presentation, figures, etc.) | Data analysis | Feedback, revisions. *Julia Krasko*—Idea, conceptualization | Visualization (data presentation, figures, etc.) | Data analysis | Feedback, revisions. *Susanne Buecker*—Idea, conceptualization | Visualization (data presentation, figures, etc.) | Data analysis | Feedback, revisions.

Other Manuscript Versions: The manuscript has been uploaded to a preprint server (https://doi.org/10.31234/osf.io/v3e7n). Moreover, the initially submitted version of manuscript has been used within the first author's dissertation (https://doi.org/10.13154/294-9612).

Data Availability: For this article, data is freely available (see Kritzler et al., 2023).

Supplementary Materials

For this article, the following Supplementary Materials are available:

- the pre-registration (see Kritzler et al., 2021)
- raw data, processed data, codebook, R script, R-Markdown, analyses, files and instructions for reproducing (see Kritzler et al., 2023)
- open peer-reviews (see Personality Science, 2023)

Index of Supplementary Materials

Kritzler, S., Haehner, P., Krasko, J., & Buecker, S. (2021). *How do you cook empathetically? The (ir)relevance of personality states in everyday situations* [Pre-registration]. OSF Registries. https://doi.org/10.17605/osf.io/3phk8

Kritzler, S., Haehner, P., Krasko, J., & Buecker, S. (2023). *Supplementary materials to "What happens when you add a 'not relevant' response option to the unipolar response scales of personality state items?"* [Data, codebook, script, analyses]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.13462

Personality Science. (Ed.). (2023). *Supplementary materials to "What happens when you add a 'not relevant' response option to the unipolar response scales of personality state items?"* [Open peer-review]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.13461

References

Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using r. *Behavior Research Methods, 52*, 376–387. https://doi.org/10.3758/s13428-019-01236-y

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* https://github.com/crsh/papaja

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., MacLeod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality, 31*(5), 503–528. https://doi.org/10.1002/per.2115

Cattell, R. B. (1946). Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology, 36*(2), 88–103.

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2. *Diagnostica, 65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

DeYoung, C. G. (2015). Cybernetic Big Five theory. *Journal of Research in Personality, 56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

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

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology, 97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

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.2016.06.003

PsychOpen GOLD

Hampson, S. E. (2012). Personality processes: Mechanisms by which personality traits get outside the skin. *Annual Review of Psychology, 63*, 315–339. https://doi.org/10.1146/annurev-psych-120710-100419

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences, 33*(2-3), 111–135. https://doi.org/10.1017/S0140525X10000725

Horstmann, K. T., & Ziegler, M. (2020). Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality, 34*(6), 1037–1059. https://doi.org/10.1002/per.2266

Hudson, N. W., Lucas, R. E., & Donnellan, M. B. (2020). Are we happier with others? An investigation of the links between spending time with others and subjective well-being. *Journal of Personality and Social Psychology, 119*(3), 672–694. https://doi.org/10.1037/pspp0000290

Huggins-Manley, A. C., Algina, J., & Zhou, S. (2018). Models for semiordered data to address not applicable responses in scale measurement. *Structural Equation Modeling, 25*(2), 230–243. https://doi.org/10.1080/10705511.2017.1376586

Jayawickreme, E., Zachry, C. E., & Fleeson, W. (2019). Whole trait theory: An integrative approach to examining personality structure and process. *Personality and Individual Differences, 136*, 2–11. https://doi.org/10.1016/j.paid.2018.06.045

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W., Kopp, R. J., Mitchell, R., Presser, S., Ruud, P. A., Smith, V., Moody, W. R., Green, M. C., Conaway, M. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66*(3), 371–403. https://doi.org/10.1086/341394

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In J. D. Wright & P. V. Marsden (Ed.), *Handbook of survey research* (pp. 263–313). Emerald Group Publishing.

Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology, 22*(3), 251–259. https://doi.org/10.1007/s10869-008-9064-2

Loeys, T., Moerkerke, B., De Smet, O., & Buysse, A. (2012). The analysis of zero-inflated count data: Beyond zero-inflated poisson regression. *British Journal of Mathematical & Statistical Psychology, 65*(1), 163–180. https://doi.org/10.1111/j.2044-8317.2011.02031.x

McCabe, K. O., & Fleeson, W. (2016). Are traits useful? Explaining trait manifestations as tools in the pursuit of goals. *Journal of Personality and Social Psychology, 110*(2), 287–301. https://doi.org/10.1037/a0039490

Menold, N., & Bogner, K. (2016). *Design of rating scales in questionnaires: GESIS survey guidelines.* GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_015

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

Rauthmann, J. F., & Sherman, R. A. (2018). *S8-I - Situational Eight-I - Deutsche Fassung.* ZPID (Leibniz Institute for Psychology Information) – Testarchiv. https://doi.org/10.23668/psycharchives.823

Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research.* Northwestern University. https://CRAN.R-project.org/package=psych

Richter, D., & Schupp, J. (2015). The SOEP Innovation Sample (SOEP IS). *Journal of Contextual Economics – Schmollers Jahrbuch, 135*(3), 389–399. https://doi.org/10.3790/schm.135.3.389

Ringwald, W. R., Manuck, S. B., Marsland, A. L., & Wright, A. G. C. (2022). Psychometric evaluation of a big five personality state scale for intensive longitudinal studies. *Assessment, 29*(6), 1301–1319. https://doi.org/10.1177/10731911211008254

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality, 34*(5), 687–713. https://doi.org/10.1002/per.2309

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*(4), 397–423. https://doi.org/10.1006/jrpe.2000.2292

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . .Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

PsychOpen GOLD