Published in final edited form as:

J Res Pers. 2017 August ; 69: 170–179. doi:10.1016/j.jrp.2016.06.015.

The Role of Response Styles in the Assessment of Intraindividual Personality Variability

Brendan M. Baird,

Louisville, KY

Richard E. Lucas, and

Michigan State University

M. Brent Donnellan

Texas A & M University

Abstract

Although personality psychologists often focus on between-person differences, understanding intraindividual variability is also a critical focus of the subdiscipline. Despite the fact that non-self-report techniques exist for assessing variability, questionnaire-based measures are still the norm. In two studies ($N = 149$ and $N = 202$) we examine the possibility that intraindividual variability measures derived from repeated self-report assessments are affected by certain response styles. These studies, which use a variety of techniques for assessing within-person variability, show that standard measures are moderately to strongly correlated with theoretically unrelated variability measures, including those based on ratings of satisfaction with neutral objects or the personality of cartoon characters. These results raise questions about the validity and utility of widely used measures for assessing intraindividual variability.

Keywords

subjective well-being; measurement; emotions; personality

Personality psychologists traditionally focus on individual differences in thoughts, feelings, and behaviors. They typically start with the assumption that there are certain relatively stable characteristics “inside” the person that influence enduring patterns of thoughts, feelings, and behaviors. Therefore, the goal of much personality research is to determine how individuals differ from one another, and this focus has resulted in the widespread use of a between-persons approach when studying personality. For instance, trait researchers may link individual differences in extraversion with individual differences in social behaviors, organizational personality psychologists may examine whether individual differences in conscientiousness are associated with work-related behaviors, and developmentally oriented

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

personality psychologists may focus on the extent to which individual differences in traits are maintained over time as people age.

Despite an emphasis on explaining between-person variance, it has also become clear that to understand how people differ from one another, it is necessary to look within the person to see how a person's thoughts, feelings, and behaviors fluctuate across short-term intervals such as a day, a week, or a month. In other words, within-person variability is becoming an increasingly important topic for personality psychologists. There are a number of reasons for this additional focus. First, although personality researchers may be interested in the stable manifestations of personality that lead to individual differences, the processes that underlie these trait-like differences are likely to be dynamic. For instance, some have suggested that to explain the between-person association between extraversion and positive affect, it is necessary to understand stable individual differences in the way that people respond to positive stimuli, and this issue can only be examined by assessing people's responsiveness (i.e., within-person change) to changing affective stimuli (Larsen & Ketelaar, 1991; Lucas & Baird, 2004; Smillie, Cooper, Wilt, & Revelle, 2012). In addition, researchers have noted that there are actually stable individual differences in patterns of within-person variability that can be assessed and incorporated into meaningful personality theories (Fleeson, 2001) and that specific patterns of within-person responses to specific stimuli may be a promising way of conceptualizing individual differences in personality (Mischel & Shoda, 1995). Thus, personality psychologists incorporate within-person techniques into their methodological toolkit, and an important component of this focus on within-person effects is the assessment of individual differences in within-person processes.

Indeed, there is a long tradition of theorizing about and actually assessing individual differences in the way that personality traits change across situations. For instance, much research has examined whether there are stable individual differences in the way that people vary and whether these individual differences correlate with outcomes like psychological health and well-being (Baird, Le, & Lucas, 2006; Donahue, Robins, Roberts, & John, 1993). Some theorists believe that variability reflects an ability to adapt to the environment and thus should be positively associated with well-being (e.g. Bem & Lewis, 1975; Paulhus & Martin, 1988; Snyder, 1974). In contrast, other theorists have argued that personality variability is a sign of psychological immaturity, or an inability to resist social pressures that push one to behave in ways that are inconsistent with his or her "true self" (e.g. Jourard, 1963; Maslow, 1968; Rogers, 1961). In other words, intraindividual variability may result from conflicts between one's dispositional tendencies and external pressures to behave in ways that are socially appropriate, but which may not come naturally to the person. According to this view, intraindividual variability should be associated with lower levels of psychological well-being (Block, 1961; Donahue et al., 1993). A critical test of these competing hypotheses requires reliable and valid measures of individual differences in within-person variability.

Shifting the focus to within-person variance does not eliminate the methodological issues that researchers face when attempting to assess personality constructs. For instance, although self-reports have known limitations, they are often the only feasible way to assess intrapsychic thoughts and feelings. And although behavior is observable, obtaining objective

assessments through outside observers or automated procedures is still difficult and resource intensive (though this is becoming easier over time with technological advances; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). The challenge of obtaining objective measures of behavior is even greater when research moves out of the laboratory, as it often does when within-person variability is the focus. Indeed, issues with self-report might be particularly relevant to within-person investigations.

The focus of this paper is on the possibility that certain response styles affect self-report measures of within-person variability. Specifically, we focus on potential methodological artifacts that may emerge when self-reported personality descriptors are administered repeatedly to obtain estimates of within-person variability. It is important to keep in mind, however, that the issue we address may also emerge when self-report methods are used to assess any construct repeatedly over time.

Measuring Between-Person Differences in Within-Person Variability

In the past, self-reports have been used to study variability in one of two ways. First, researchers have simply asked people direct questions about how variable they are. For instance, the Self-Pluralism Scale (McReynolds, Altrocchi, & House, 2000) includes items such as “I act and feel essentially the same whether at home, at work, or with friends” and “My personality is always the same regardless of whom I’m with or the situation I’m in.” Presumably, people who strongly endorse these statements should exhibit little variability in their behavior. Other measures focus explicitly on the underlying processes that might lead to observable variability in behavior. For example, the Self-Concept Clarity Scale (Campbell et al., 1996) includes items such as “My beliefs about myself often conflict with one another” and “Sometimes I feel that I am not really the person that I appear to be.” Similarly, Rosenberg, Schooler, and Schoenbach (1989) used a 5-item scale to capture variability in self-esteem that includes items such as “Does your opinion of yourself tend to change a good deal, or does it always continue to remain the same?” and “Do you ever find that on one day you have one opinion of yourself and on another day you have a different opinion?” These instruments may provide a way to test competing theories about the mechanisms underlying variability in observable behavior.

Alternatively, researchers have computed variability more directly by examining variation in self-reports across items, scales, contexts, or occasions of assessment. For instance, some researchers have attempted to assess constructs related to within-person variability by looking at the internal consistency of a person’s responses across items that measure the same trait (Lanning, 1988). If a measure contains multiple items that assess extraversion, for example, then each item in the scale can be treated as a separate indicator of the underlying trait. A person who is consistently extraverted should endorse the items “I am talkative” and “I am outgoing” equally (assuming they are equally good indicators of the underlying trait), whereas a person who is variably extraverted might be expected to endorse one item but not the other (Reise & Waller, 1993). In addition, if variability is a global trait, then the person who responds consistently across extraversion items should also respond consistently across items that measure other domains such as conscientiousness or neuroticism.

Although researchers have demonstrated that within-scale response consistency may be linked to some important individual differences (Berg & Collier, 1953; Goldberg, 1978; McFarland & Sparks, 1985; Siegrist, 1996), this approach to measuring personality variability suffers from two important limitations. First, indexes of cross-item variability can be confounded with mean scores across the same items. Specifically, in order for a person to get a high mean score on an extraversion scale, he or she must give responses that are consistently high across all of the items (Paunonen & Jackson, 1985). In contrast, a person can get a moderate score either by giving consistently moderate responses or by giving a combination of high and low responses. Therefore, individual differences in mean levels must be carefully separated from indexes of internal consistency. The second limitation concerns the potential impact of scale reliability and unidimensionality on response consistency. Specifically, a person may be more likely to give variable responses across items from an unreliable measure than from a reliable measure or across items that actually tap distinct constructs despite appearing on the same scale. Thus, there is some ambiguity about what within-scale variability indexes assess, and therefore, researchers have turned to alternative ways of estimating bottom-up consistency from self-reports.

Another approach to assessing within-person variability is by comparing a person's responses across repeated measurements designed to ask about behaviors in particular situations or roles. In other words, if an individual is asked to answer the same set of questions multiple times, then the extent to which his or her responses change across occasions can be used as an index of personality variability (Baird et al., 2006; Donahue et al., 1993; Eid & Diener, 1999; Fleeson, 2004). For example, researchers interested in the ways someone's extraversion changes across social roles might ask the person to describe what his or her personality is like around friends and around family members. Someone who is variably extraverted might say that he or she is highly talkative around friends but not at all talkative around family members, whereas someone who is consistently extraverted should report being equally talkative in both kinds of situations.

An advantage of this "bottom-up" approach over measures of internal-consistency is that it can be used to capture variability in specific behaviors, as opposed to variability across similar behaviors that are related to one another but are nonetheless distinct.

Furthermore, repeated-measures indexes may also be less sensitive to scale reliability than indexes of internal-consistency because responses are compared within each item, instead of across items. However, repeated-measures indexes still suffer from some important limitations. First, like indexes based on within-scale variability, estimates of consistency that are derived from repeated assessments can become confounded with mean scores in ways that alter the interpretation of findings (Baird et al., 2006; Eid & Diener, 1999; Locke, 2003). In order for a person to have an average that is extremely high or extremely low across occasions, he or she must give the same answer each time. Alternatively, people with moderate scores could get those scores by being consistently moderate or by giving a combination of high and low responses. Therefore, just as with other bottom-up indexes, individual differences in mean scores should be carefully separated from indexes of variability.

A second limitation of the repeated-measures approach is that it may be sensitive to unique response biases that are not typically found in global questionnaires. For instance, when people are asked to answer the same questions several times, norms that govern the exchange of information in everyday conversations may have an influence on the ways they respond (Schwarz, 1999). Baird and Lucas (2011) found that differences in self-reported personality across contexts were significantly larger if participants were asked the same questions multiple times within the same questionnaire than if they were asked each question only once. This suggests that a certain amount of inconsistency may be due to the measurement process, rather than to actual individual differences in personality.

Response Styles in Variability Measures

For many years, researchers have acknowledged that self-report measures are sensitive to certain biases that may threaten their validity (Cronbach, 1950; Jackson & Messick, 1958; Rundquist, 1950). Although these issues are inevitable whenever self-reports are used, increasing evidence suggests that unique problems may emerge when the same measures are administered repeatedly, as they are when indexes of within-person variability are obtained. For instance, in a slightly different context, Watson and Tellegen (2002) suggested that the tendency to blindly agree or disagree with items can create problems when multiple responses are combined into a single index. Specifically, compared to global ratings of affect, aggregated measures of momentary affect yield stronger positive correlations among affect scales of similar valence (e.g., joy and happiness) and weaker negative correlations among scales of opposite valence (e.g., joy and sadness). Watson and Tellegen called this the “aggregation paradox” (p. 596), and showed that it is directly related to acquiescent responding.

If response styles have an impact on the correlations among mean scores on an aggregate measure, then they may also have an impact on indexes of intraindividual variability across repeated measurements. In other words, some people may use response scales in ways that inflate the variability in their responses, whereas other people may respond in ways that are more consistent. This would mean that bottom-up indexes of intraindividual variability may be contaminated with stable individual differences in response styles. Of particular concern may be the tendency of some participants to use extreme numbers.

For instance, in one early study that examined this issue, participants were asked to report their reactions to several neutral words and abstract shapes (Berg, 1953). For each word or shape, participants indicated their initial impression of the stimulus on a 4-point scale. The number of times participants expressed an extreme response (“like much” “dislike much”) to neutral words was correlated with the number times they expressed an extreme response to abstract shapes. These consistent responses to unrelated stimuli suggested the existence of a response style. Subsequent research has shown that the tendency to give extreme responses is also quite stable (Austin, Deary, & Egan, 2006; Greenleaf, 1992; Wetzel, Carstensen, & Böhnke, 2013; Wetzel, Lüdtke, Zettler, & Böhnke, 2015). This is important for research on intraindividual personality variability because bottom-up indexes are likely to be inflated by extreme responding.

To date, few studies have directly investigated the role that extreme responding has on measures of personality variability (Diener & Larsen, 1984; Penner, Shiffman, Paty, & Fritzsche, 1994). However, some research indirectly suggests that response styles may play a role. For instance, in one line of research examining the correlates of within-person variability measures, Baird et al. (2006) tested the associations between different measures of variability and a variety of predictors and outcomes, including personality traits and global well-being. The goal of this research was to first test the psychometric properties of variability scores, and then to develop the nomological network surrounding the variability construct. However, as described above, scores on any measure of variability (for instance, a person's variability in extraversion across multiple social roles) are necessarily confounded with participants' mean levels across those roles. Thus, to test whether variability is related to predictors and outcomes, it is necessary to first correct or control for mean levels of the construct (see Baird et al., for details on procedures that can be used to make this correction).

Baird et al. found that once these corrections were made, measures of variability were reliable and stable over relatively long periods of time, and different measures of variability correlated with one another. However, these indexes no longer correlated substantially with any of the other substantive correlates that they examined. One possible explanation of this pattern of associations is that once mean-level variance is removed from measures of intraindividual variability, much of the remaining variance reflects individual differences in response styles. When combined with the fact that many response styles—including response styles for extreme responding—appear to be quite stable over time (Wetzel et al., 2013, 2015), understanding the role that these response-styles play in individual differences in within-person variability is an important goal.

Given that so little research has investigated the effects of response styles on the measurement of intraindividual variability, it is somewhat unclear how to investigate these issues. One way to test the effects of extreme responding would be to correlate scores on measures of intraindividual variability in personality with an index of variability that is based on responses to different—and theoretically unrelated—attributes. For instance, variability in a person's repeated responses on a bottom-up measure of personality across contexts should not be associated with variability in his or her descriptions of daily weather conditions. Similarly, variability in a person's descriptions of other people should not be associated with variability in his or her moods across days of the week. Finally, variability in a person's descriptions of neutral objects should not be associated with variability in his or her daily reports of personality. Thus, correlations among these measures would provide particularly strong evidence that estimates of intraindividual personality variability are capturing irrelevant information.

The Current Studies

Few studies have examined the impact of response styles on indexes of intraindividual variability derived from self-report techniques. One way to conduct such a test would be to compare variability indexes derived from ratings of theoretically unrelated scales and see how strongly they relate to one another (Greenleaf, 1992). If within-person variability is

correlated with variability in responses to a wide variety of theoretically irrelevant constructs, then this would suggest that response styles such as extreme responding play an important role. In addition, we examine whether controlling for such response styles changes associations between within-person variability and theoretically relevant outcome measures.

Study 1

Study 1 included two measures of personality variability based in historical approaches to assessing this construct. The first is based on the within-scale variability of items assessing the same construct (Lanning, 1988). The second is based on the standard deviation of a person's response to the same items across different contexts or roles (Baird et al., 2006; Fleeson, 2001). In addition, two measures designed to assess response styles were included. The first of these is based on the variability in a person's ratings of inanimate objects, or mundane features of his or her environment; and the second is based on the variability of a person's ratings of different cartoon characters. If within-person-variability indexes are not confounded with response styles, then the two measures should cohere and be unrelated to the theoretically irrelevant measures of variability across cartoon characters or inanimate objects. Finally, participants completed questionnaires about their own perceptions of their variability, along with measures of well-being and relationship satisfaction, which were included as outcome measures whose association with intraindividual variability could be assessed.

Method

Participants—One hundred forty-nine university students were recruited from undergraduate psychology courses during the summer semester. A researcher visited each classroom and asked volunteers to complete a brief anonymous survey about personality and well-being. In exchange for completing the survey, participants received course credit.

Materials—A variety of self-report questionnaire methods were administered, including measures of intraindividual variability, measures designed to assess response style, and outcome measures to examine the correlates of the variability measures.

Intraindividual Variability: Participants first completed the 50-item International Personality Item Pool measure of the Big Five personality traits (Goldberg, 1999). This measure consists of several short statements that are each rated on a scale from 1 ("very inaccurate") to 5 ("very accurate"). Negatively worded items were reverse keyed so that each trait would be assessed with 10 items of the same valence. Within-scale variability indexes were constructed for each trait separately by calculating the within-person standard deviation across items assessing the same trait. Because of the inherent dependency between means and standard deviations for variability indexes derived from scales (see Baird et al., 2006), these scores were adjusted to remove the dependency. Specifically, for each person, a mean and cross-item standard deviation was calculated for each of the Big Five scales. The standard deviations for each trait were then regressed on the mean of the items from that trait, along with the square of the mean (as the forced dependence between means and

standard deviations are curvilinear; again, see Baird et al., 2006). The residuals from these regression analyses represent the extent to which a person with a given mean score has a high amount of cross-item variability relative to other individuals with the same mean. The residuals for the five traits were averaged to create an overall index of variability. As an index of reliability, Cronbach's alpha for the five trait-specific corrected variability scores were calculated.¹

Participants were also asked to complete a measure of personality across contexts (Fleeson, 2001). For each of 4 different social roles (Friend, Family Member, Romantic Partner, and Student), participants indicated how well 15 different words describe their behavior on a scale from 1 ("does not describe me") to 5 ("describes me very well"). From these ratings, variability scores can be created by first calculating the cross-role means and standard deviations for each of the 15 personality items for each person in the file. Each of these 15 columns of cross-role standard deviations were then regressed on the corresponding mean and the square of the mean, and the residuals were saved as corrected item-specific variability scores. These item-specific indexes can then be combined to create variability scores for each of the Big Five traits and for overall variability (Baird et al., 2006; Fleeson, 2001), though because the focus of this paper is on the total score, only these scores are reported. Reliability was assessed using Cronbach's alpha with adjusted, item-specific variability scores as items.

Finally, participants completed the Self-Concept Clarity scale (Campbell et al., 1996), which assesses participants' own perception of their intraindividual variability. This measure consists of 12 items, each rated on a scale from 1 ("strongly disagree") to 5 ("strongly agree"), each of which measures perceptions of variability). Scale scores were computed from the average across the items ($M = 3.32$, $SD = .92$).

Response Styles: Participants completed two measures of response styles. First, they were asked to indicate on a 5-point scale, how well each of ten adjectives describes four different characters from the television show *The Simpsons*. Because participants are rating the same stimuli, each person should report a similar amount of cross-character variability. To the extent that differences in variability exist (for instance, if participants have varying degrees of knowledge about the characters), these differences should be unrelated to actual variability in one's own personality traits across contexts. Thus, any correlations between indexes derived from ratings of cartoon characters should, theoretically, be unrelated to ratings of a person's own variability across contexts. As with the other measures of variability, cross-character means and standard deviations were first calculated for each person. Then, the 10 columns of cross-character standard deviations (one per item) were regressed onto the individual-level means and square of the means for each item. The 10 columns of residuals were saved and averaged to create an overall measure of variability in ratings of the four characters. Coefficient alpha was calculated using the adjusted variability scores for the individual items.

¹Previous work has shown that associations between variability scores and outcomes often change once these scores are adjusted for the mean. Although this issue is not the focus of this paper, the on-line supplement reports correlations between unadjusted variability scores and all other scores reported in this paper.

The second response style measure consisted of a survey of attitudes toward neutral “objects” that has been used in previous research to assess positive response biases (Judge & Bretz, 1992). Participants reported their feelings toward each of 25 relatively neutral targets (such as “The movies being produced today,” “8 1/2 by 11 inch paper,” “Your telephone number”) using a scale that ranged from 1 (“very dissatisfied”) to 5 (“very satisfied”). With this measure, an estimate of response styles in variability can be derived from the consistency of each person’s ratings across the objects. Because most of these objects are relatively neutral (and at the very least, mostly consistent across individuals), each person should report a similar amount of cross-object variability. To the extent that real differences in variability exist, there is little theoretical reason to expect that individual differences in this variability across targets should be related to cross-character variability in ratings of the Simpsons or cross-role and cross-item variability in self-reported personality. For this response-style measure, it is possible to calculate reliability using split-half tests. Because there is an extremely large number of potential split half tests that could be conducted, we took the average split-half reliability from 1000 random splits of the 25 items.

Outcome Measures: Finally, because some past work on outcomes of personality variability suggest that people’s relationships should suffer if they exhibit high levels of cross-situational variability, participants completed a measure of satisfaction with friendships and kinships. This measure was derived from a measure of perceived autonomy support (Deci & Ryan, 2006) and consists of 18 items that were each rated on a 5-point scale. Scale scores were computed by averaging together each participant’s responses on all the items ($M = 4.05$, $SD = .75$).

Results

The primary goal of these analyses is to evaluate how strongly indexes of intraindividual variability are associated with indexes of variability derived from self-report ratings of theoretically unrelated constructs. However, a necessary first step in this process is to show that variability is a general construct and that variability in different traits cohere well enough to suggest that there is a unidimensional variability construct. Although Baird et al. (2006) examined this issue thoroughly, we report basic details about this coherence as preliminary analyses. For instance, for the first index, which is derived from the corrected cross-item standard deviation from each of the five Big Five traits, correlations between trait-specific variability indexes ranged from .22 to .42 (all p ’s < .05) with a mean r of .30. Similarly, for the 15 items rated across 4 roles, 15 individual corrected cross-role standard deviations were calculated. These residual standard deviations were first averaged to yield separate variability scores for each of the Big Five factors. The bivariate correlations among these trait-specific indexes ranged from .35 to .55 (all p ’s < .05), with a mean of .47. Thus, these components were combined into overall variability indexes (one for each of the two methods).

Table 1 shows the correlations among the various measures assessed in this study. First, it is important to note that the two measures of within-person variability—one derived from the within-scale, cross-items standard deviations and one derived from the cross-role standard deviation—correlated strongly with one another ($r = .54$). This suggests that the two

measures are tapping similar constructs, which would be predicted if both assessed true within-person variability.

Next, turning to the more explicit self-report of perceived intraindividual variability, a different picture emerges. Here the correlations between the standard-deviation-based measures and the explicit measure of perceived variability are essentially zero. This finding replicates similar results from Baird et al. (2006), suggesting that the two methods for assessing within-person variability are quite different from one another.

The correlations in Table 1 with the Simpsons and Neutral Objects measures are intended to assess the association between response styles and intraindividual variability measures. These correlations, which are moderate in size for both the contextual measure and the within-scale variability measure, suggest that response styles may play a substantial role in ratings of intraindividual variability. Specifically, these correlations show that people who report being variable across roles or who respond in variable ways to different items measuring the same trait, also report more variability across Simpsons characters or across their ratings of neutral objects. The fact that the correlations between these theoretically unrelated measures are almost as high as that between two theoretically related measures (the two standard-deviation-based measures of intraindividual variability) raises serious questions about the validity of these measures as indicators of true intraindividual variability.

The final row of Table 1 shows the correlations of all measures with self-reports of relationship satisfaction. Consistent with past research (Baird et al., 2006), once mean-level variance is removed from the intraindividual variability measures, correlations between these measures and measures of well-being (in this case relationship satisfaction) are very close to zero. Also consistent with past research, more explicit ratings of participants' perceptions of within-person variability (in the form of the Self-Concept Clarity Scale) do correlate moderately with satisfaction. Finally, somewhat surprisingly, the two measures intended to tap response styles also correlate significantly (though weakly) with satisfaction measures.

Overall, these results suggest that a substantial amount of variance in standard measures of within-person variability in personality constructs reflects individual differences in response styles that can be seen in patterns of correlations among variability indexes derived from theoretically unrelated constructs.

Study 2

Study 1 provided evidence that variability indexes could be associated with one another even when these indexes were derived from ratings of theoretically unrelated content. Specifically, indexes designed to capture cross-role or cross-item variability in personality were moderately strongly associated with theoretically unrelated sources of variability including variability in ratings of personality across cartoon characters and variability of desirability of relatively neutral object. Such results lead to concerns about what exactly these indexes of variability capture, and they raise the possibility that much of the reliable variance in these measures reflects individual differences in response styles, as opposed to substantive constructs.

However, one concern about the measures used in Study 1 was that they were all completed at a single point in time. In other words, they did not directly assess change in self-reported ratings as people experienced different situations. Instead, they assessed variability across items that ask about hypothetical roles or situations. It is possible that response style concerns will be mitigated when more ecologically valid methods, such as experience sampling methods or daily diary studies, are used. Previous research has shown that the measures used in Study 1 correlate strongly with intraindividual variability measures derived from experience sampling studies (Baird et al., 2006). However, it is still possible that variability indexes derived from repeated assessments correlate less strongly with the types of measures included in Study 1 to assess response styles.

Therefore, Study 2 was designed to include both exact and conceptual replications of results from Study 1 and to expand the analysis to measures of intraindividual variability that are derived from actual repeated assessments over time. Specifically, we included a daily diary component to assess variability in additional constructs over time and to incorporate an additional index of variability in theoretically unrelated constructs (in this case, the weather).

In addition, prior work on the psychometric properties of variability measures has primarily focused on within-person variability in personality. However, there is an independent tradition of examining variability in self-esteem over time (Kernis, 2005). If the issues discussed in this and previous papers (Baird et al., 2006) affect intensive repeated measures more broadly, then they should also play a role in measures of self-esteem. Thus, we also included daily measures of self-esteem. We first use these measures to assess whether the confounding influence of mean-level variance affects associations between variability in self-esteem and well-being related outcomes. Then we test whether corrected measures are associated with within-person variance in personality and with the measures designed to assess response style.

Method

Participants—Two hundred fifty-two undergraduate students participated in a three-week study of personality and emotions. The study was conducted using on-line questionnaires, and correspondence between researchers and participants was conducted via electronic mail. In exchange for completing the study, participants were given course credit. Individuals who dropped out before the end of the study received partial credit. Data from these participants had previously been analyzed in Donnellan, Oswald, Baird, and Lucas (2006), a paper focused on developing short measures of the Big Five Personality traits. The primary measures from this paper were not included in that earlier work.

Measures—Again, participants completed a number of self-report measures designed to assess personality variability. As in Study 1, these included measures of intraindividual variability, measures designed to tap response style, and outcome measures. In addition, in Study 2, a daily-diary component was added to examine variation in responses over multiple occasions.

Variability: Personality variability was assessed using some of the same measures administered in Study 1 and some slightly modified measures. Specifically, responses on the

50-item IPIP were used to calculate cross-item variability within a trait measure. Scores were again adjusted for mean levels using the same procedures recommended by Baird et al. (2006), and trait-specific variability scores were averaged to create a single index. This measure was administered at an initial session and again at a final session following the completion of the daily diary study (see below for details about the timing). In addition, a 20-item, 6-role measure of variability across roles was used to first create trait-specific cross-role variability scores (also corrected for mean levels), which were then averaged to create a single index of cross-role variability. In addition, two different measures of explicit perceptions of intraindividual variability were administered (one of which was retained from Study 1 and one of which was new). As in Study 1, the Self-Concept Clarity scale ($\alpha = .89$) was administered. In addition, the Stability of Self Scale (Rosenberg et al., 1989), which contains 5 items that capture consistency in one's opinions and ideas about the self (e.g., "Does your opinion of yourself tend to change a good deal, or does it always continue to remain the same?", "Do you ever find that on one day you have one opinion of yourself and on another day you have a different opinion?"), was administered ($\alpha = .86$).

Response Style: Again, two different questionnaire measures designed to capture individual differences in response styles were administered. As in Study 1, the measure of satisfaction with various neutral objects was administered. Variability indexes were again created by taking the within-person standard deviation across items and correcting for mean scores. In addition, a second index was derived from participants' ratings of a close friend's personality across the same items and contexts used for the self-report of contextual variability. Again, variability ratings were created by taking the cross-role standard deviation for each item, correcting this standard deviation for its quadratic association with the item mean, and then averaging the residuals to create a single index.

Outcome Measures: Participants also completed several global measures of psychological well-being. Participants completed the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) as a measure of life satisfaction. The Positive Affect Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) was used to assess positive and negative affect. In addition, the Intensity and Time Affect Survey (Diener, Smith, & Fujita, 1995) was used to assess trait levels of positive affect ($\alpha = .82$) and negative affect ($\alpha = .90$). Participants indicated on a 5-point Likert-type scale from 1 ("very slightly or not at all") to 5 ("extremely"), the extent to which they experience each of the listed emotions. Global self-esteem was measured using the 10-item Rosenberg (1989) measure ($\alpha = .91$), with items rated on a 5-point scale ranging from 1 ("does not describe me at all") to 5 ("describes me very well"). Participants also completed Spielberger's (1983) 20-item Trait Anxiety Scale ($\alpha = .92$), the 10-item short form ($\alpha = .81$) of the Center for Epidemiological Studies Depression Scale (Cole, Rabin, Smith, & Kaufman, 2004), and a 10-item measure of dispositional optimism ($\alpha = .82$) (Scheier, Carver, & Bridges, 1994).

Daily Diary Measures: To assess personality variability across ongoing experiences, participants were asked to give daily reports of their behavior and affect. On this measure, each of the Big Five traits was assessed with two adjectives that also appeared on global self-reports of variability across contexts. In addition, participants answered 14 questions

about the emotions they experienced; these questions were taken from the Intensity and Time Affect Survey described above (Diener et al., 1995). A single item was used to assess variability in self-esteem: “Today I feel that I am a person of worth, at least on an equal basis with others.” Finally, to measure response style, participants were asked to report how sunny, windy, and cold it was each day on a scale that ranged from 1 (“not at all”) to 5 (“very”). Because all the participants were living in the same city and all participants took part in the study at the same time, they were exposed to the same weather conditions. Therefore, any individual differences in the variability of weather reports are assumed to reflect response style.

Procedure—To sign up for the study, participants logged on to the Psychology Department’s website and completed an online consent form. During this initial session, participants completed the 50-item IPIP, the measure of role variability, the affect questionnaire, the neutral objects questionnaire, and the Self-Concept Clarity scale. Then, for the remainder of the study, participants were instructed to log on to a password-protected website, where they could complete the daily reports and a follow-up questionnaire.

Participants completed the daily reports each evening during the same two-week period. Participants were informed that the website could be accessed between 5:00 PM and midnight each day, and that they would not be allowed to make up any days that were missed. On average, the daily report sessions lasted about 2 minutes. Participants who participated in the daily-diary portion of the study ($N = 239$) completed an average of 10.68 daily reports, and roughly 90% ($N = 217$) completed at least 7.

Finally, at the end of the two weeks, 229 participants completed a follow-up questionnaire on the same website. That questionnaire again included the 50-item IPIP (from which within-scale variabilities were derived), the measure of the participant’s friend’s variability across roles, the stability of self-measure, and the global measures of self-esteem, optimism, depression, and anxiety. For most analyses described below, the 203 participants with complete data across all measures are included; though results are quite similar when pairwise correlations are interpreted.

Results

Table 2 shows the correlations among the five measures of intraindividual variability (Time 1 within-scale variability, Time 2 within-scale variability, contextual variability, daily variability in personality, daily variability in positive affect, and daily variability in negative affect), the three measures included to assess the influence of response styles (contextual variability of friends, variability across neutral objects, and variability in ratings of weather), and the two explicit measures of perceptions of variability (the Self-Concept Clarity Scale and the Stability of Self Scale). These results replicate and extend those initially found in Study 1.

First, consistent with the results of Study 1, the various measures of personality variability cohere. Corrected within-scale variability is stable from Time 1 to Time 2, and it again correlates moderately strongly with the contextual measure (see the third row of Table 2). The addition of a daily-diary component allows for the examination of additional

intraindividual variability measures, and again, correlations among the various measures are moderate to strong. Specifically, correlations between corrected daily personality variability and the three questionnaire-based measures range from .44 to .60, whereas correlations between the affect variability measures and the questionnaire-based measures range from .25 to .39. Thus, even when relatively diverse methods for assessing variability (i.e., within scales, across contexts, or over time) are examined, these measures cohere moderately strongly.

The next question we can address using the correlations reported in Table 2 concerns the associations between measures designed to assess substantive differences in variability and those designed to capture response styles. First, Row 7 shows the associations between a person's cross-context ratings of a friend's personality and each of the other variability measures we administered. Although this index may capture real variability in the friend's personality, there is little reason to expect that it would correlate with the participant's own variability. Yet, the correlations in Row 7 are again relatively strong, ranging from a low of .38 (for variability in positive and negative affect) to a high of .64 (for the participant's own contextual measure).

Of course, these data alone do not rule out the possibility that there is a real correspondence between one's own intraindividual variability and one's friend's variability (though it would be surprising if this correspondence led to correlations that are this high). Alternatively, it is possible that participants have difficulty rating their friend's variability and therefore fall back on responses that mirror their own changes across situations. Thus, there is some ambiguity about the associations between the ratings of friends and the other measures of variability. Therefore, it is again useful to look at alternative measures that may tap response styles. Row 8 of Table 2 shows the correlations between variability in ratings of neutral objects and each of the other variability measures assessed. Consistent with the results from Study 1, these correlations are all moderate in size. Again, it is unlikely that there is a substantive explanation for these moderate correlations, which means that a methodological explanation is more likely.

As a final check on the possible role of response styles, we examined variability in ratings of weather across the 14-day daily-diary study. These correlations are reported in Row 9 of Table 2. Unlike the correlations among the other measures that were administered in this study, these correlations are all quite small (though some are significantly greater than zero). These results go against our expectation that variability indexes derived from a variety of content areas would cohere due to response styles, though as we describe in the Discussion, there were some features of the objective weather conditions during this two-week period that may have limited the utility of this measure in this particular period.

Finally, the last two rows of Table 2 show the associations between each of the two explicit measures of perceptions of variability and each of the other indexes. Although these two measures are very strongly correlated with one another, they are mostly uncorrelated with the other measures of variability assessed in this study. The primary exception is a moderate correlation between each measure and daily variability in positive affect (along with a small

but significant correlation between the Stability of Self Scale and daily variability in personality).

As with Study 1, we also examined the associations between the substantive intraindividual variability measures and a variety of outcomes related to well-being. These correlations are shown in Table 3. As can be seen in this table, with the exception of variability in daily positive affect, correlations with these variables are generally very small and not statistically significant. These are consistent with previous results suggesting that intraindividual variability in personality characteristics is mostly unrelated to well-being after controlling for mean levels. The difference in results for variability in positive affect was not predicted and should be replicated before it is interpreted.

As a final step in our analyses, we extended our investigation to self-esteem variability, first replicating analyses from our prior study (Baird et al., 2006) and then testing whether measures of self-esteem variability were also associated with measures designed to capture response style variance. Results from these analyses are presented in Table 4. First, these results replicate past findings showing that self-esteem variability is significantly associated with all the measures of well-being that we included in this study. However, once variability scores are corrected for the dependence with the mean, all associations except for the association with depression are nonsignificant and very small. Thus, at least in this sample, once mean-level variance is removed, self-esteem variability is mostly uncorrelated with well-being measures.

Next, Table 4 shows that both indexes, but especially the corrected index, correlate weakly to moderately with the measures of personality variability. For instance, corrected daily variability in self-esteem correlates .54 with corrected daily variability in personality. This again suggests that variability measures capture something that is consistent across constructs.

Third, these results show that although the uncorrected measure is correlated with the two questionnaire-based measures of perceived variability (the SCC and SSS), these correlations drop considerably (and in one case to nonsignificance) when appropriate corrections are made. Thus, this suggests that people who are generally low on self-esteem perceive themselves as more variable, but once this effect of mean-level self-esteem is partialled out, there are only relatively weak associations between actual variability and perceptions of variability as measured by questionnaires.

Finally, Table 4 shows that the corrected self-esteem variability measure correlated weakly to moderately with the measures designed to assess response styles, including significant correlations with variability across the ratings of neutral objects and contextual ratings of friends' personalities. However, consistent with the results for the personality variability measures, this measure of self-esteem variability was not significantly associated with variability in ratings of weather.

Discussion

Personality psychologists seek to understand intraindividual variability in thoughts, feelings, and behaviors, both because this source of variance gives insight into the processes that underlie stable individual differences, and because these characteristic patterns of cross-situational change may reflect important individual differences in their own right. Of course, to study intraindividual variability, one must measure it; and to do this, personality researchers have often turned to self-report methods. The question we addressed in these studies was whether the pattern of associations among various indexes of variability—some of which were selected specifically because they were theoretically unrelated to the others—suggested the influence of response styles that may be uniquely relevant when studying intraindividual variability.

Our results showed that measures of intraindividual personality variability do cohere, which at first glance might suggest a broad and pervasive underlying factor. In other words, using the logic of convergent validity, the fact that people who are variable in their levels of extraversion across roles are also variable in their levels of conscientiousness would seem to indicate that the two measures are tapping a common factor. The fact that measures of cross-role variability (assessed in a single setting) also correlate moderately to strongly with measures of cross-situational variability derived from repeated measures administered over a two-week period should provide further support for this conclusion (also see Baird et al., 2006). However, these indexes of within-person variability were also moderately to strongly correlated with indexes of variability in ratings of theoretically unrelated constructs. For instance, those participants who reported personality scores that varied across contexts also reported that their friends' personalities varied across contexts. Even more disconcerting was the fact that people who said that their own personalities varied considerably across contexts or situations also reported greater variability in their satisfaction with a variety of neutral objects and more variability in the personalities of a set of cartoon characters. These latter two findings are especially difficult to explain with existing theories of personality variability.

Importantly, the breadth of the measures we used suggests that there is no easy way to avoid this problem. The measures that were included to assess response styles were not only moderately correlated with questionnaire-based measures of variability administered in a single setting; they also predicted actual changes in both self-report affect and self-reported personality over time in a daily-diary study. These results suggest that the impact of response styles is pervasive across a wide variety of methods for assessing variability. This fact, when combined with the result that corrected indexes of intraindividual variability rarely correlate strongly with any other construct that is typically assessed by personality psychologists—including mean-level personality traits or a broad range of well-being measures (see Baird et al., 2006, in addition to the results from the current studies)—should give pause to researchers who use these measures in their work. In short, our results suggest that a programmatic effort to evaluate the validity of measures of intraindividual variability is needed.

It is also important to note that explicit measures of people's perceptions of their own personality variability cannot yet solve this problem. Although these measures are

consistently and robustly correlated with outcome measures such as measures of subjective well-being (and with other personality trait measures; see Baird et al., 2006), the correlations between these perception measures and the other variability indexes we calculated tend to be very close to zero. This finding also replicates prior work (Baird et al., 2006). Thus, future attempts to validate measures of intraindividual variability should also carefully consider what these perception measures actually assess.

Of course, the two studies presented here do not definitively prove that self-report measures of intraindividual variability are substantially affected by response styles. For instance, the associations with at least one of the measures that was designed to examine the role of response styles—participants' ratings of their friends' personality variability—has an important alternative explanation. Although the correlations are too high to be plausibly explained by actual similarity across friends, it may be that participants lack knowledge about their friends and therefore rely on their own changes across situations when making judgments about their friends. And although it is more difficult to come up with similar explanations for the associations with the ratings of neutral objects or cartoon characters, it is possible that such an explanation does exist.

Furthermore, at least one of the measures that we developed to assess response styles showed weaker associations with the variability measures (though even with this index, the correlations were frequently significantly greater than zero and occasionally moderate in size). Specifically, participants' ratings of the weather across the 14-day period were less strongly associated with the other variability measures included in the study. However, an examination of the objective weather conditions during those days suggests that this measure may not have been ideal for capturing response styles, at least during this period. For instance, during the two-week period, there was very little rain, with 10 days with no precipitation at all and very little rain on all the others (which meant that scores for "rainy" were very close to the floor of the scale). In contrast, the temperature varied quite dramatically through the period, with high temperatures ranging from 38 degrees Fahrenheit (where "cold" scores were close to the ceiling of the scale) to 81 degrees (where "cold" scores were very close to the floor of the scale). Thus, the objective and observable features of the weather during these weeks may have overwhelmed the impact of response styles on the measures. Of course, this is a post hoc explanation of the discrepant results for this particular measure, and future studies should either re-examine the use of weather ratings for this purpose or include alternative measures of response style in the design.

A number of theorists have argued that intraindividual variability is an important component of personality and some have even developed complex models to account for the processes underlying it (Block, 1961; Donahue et al., 1993). Many of these models make clear predictions regarding the relationships among multiple indicators of consistency. For instance, theories that link intraindividual variability to identity conflicts or to unstable self-concepts predict that multiple aspects of a person's behavior should be similarly unstable. Furthermore, these theories predict that the internal conflicts that lead people to behave inconsistently can be assessed by asking variable people direct questions about their self-concepts. Finally, conflict theories also predict that personality variability should be associated with lower levels of psychological well-being.

The results reported in these studies confirm prior findings that when appropriate adjustments for confounded mean-level variance are applied, associations between variability in self-reported ratings of personality characteristics and well-being outcomes are attenuated and generally close to zero. However, the current studies also raise new concerns that these self-report-based measures of variability may be substantially influenced by response styles and thus, may not provide strong tests about true variability in personality characteristics. Thus, future work on personality variability should, for the time being, treat measures based on self-reports somewhat skeptically. In addition, a concerted effort to examine the validity of these measures is needed. Finally, any such effort should ideally include non-self-report criteria as an integral component of the validation process, so that concerns about shared method variance and the influence of response styles can be eliminated. Hopefully, as technological improvements in psychologists' ability to assess thoughts, feelings, and behaviors over time improves, these serious concerns about existing measures of intraindividual personality variability can be addressed in a strong way.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by funding from the National Institute on Aging (R01AG040715).

References

- Austin EJ, Deary IJ, Egan V. Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*. 2006; 40(6):1235–1245.
- Baird BM, Le K, Lucas RE. On the nature of intraindividual personality variability: reliability, validity, and associations with well-being. *Journal of personality and social psychology*. 2006; 90(3):512–527. [PubMed: 16594835]
- Baird BM, Lucas RE. "... And How About Now?": Effects of Item Redundancy on Contextualized Self-Reports of Personality. *Journal of personality*. 2011; 79(5):1081–1112. [PubMed: 21241305]
- Bem SL, Lewis SA. Sex role adaptability: One consequence of psychological androgyny. *Journal of personality and Social Psychology*. 1975; 31(4):634.
- Berg IA. The reliability of extreme position response sets in two tests. *The Journal of Psychology*. 1953; 36(1):3–9.
- Berg IA, Collier JS. Personality and group differences in extreme response sets. *Educational and Psychological Measurement*. 1953; 13:164–169.
- Block J. Ego identity, role variability, and adjustment. *Journal of Consulting Psychology*. 1961; 25(5):392–397. [PubMed: 13870048]
- Campbell JD, Trapnell PD, Heine SJ, Katz IM, Lavalley LF, Lehman DR. Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology*. 1996; 70(1):141–156.
- Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. *Psychological assessment*. 2004; 16(4):360–372. [PubMed: 15584795]
- Cronbach LJ. Further evidence on response sets and test design. *Educational and psychological measurement*. 1950; 10:3–31.
- Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *Journal of Personality Assessment*. 1985; 49:71–75. [PubMed: 16367493]

- Diener E, Larsen RJ. Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of personality and social psychology*. 1984; 47(4):871–883. [PubMed: 6512686]
- Diener E, Smith H, Fujita F. The personality structure of affect. *Journal of personality and social psychology*. 1995; 69(1):130–141.
- Donahue EM, Robins RW, Roberts BW, John OP. The divided self: concurrent and longitudinal effects of psychological adjustment and social roles on self-concept differentiation. *Journal of Personality and Social Psychology*. 1993; 64(5):834. [PubMed: 8505712]
- Donnellan MB, Oswald FL, Baird BM, Lucas RE. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment*. 2006; 18(2):192–203. <http://doi.org/10.1037/1040-3590.18.2.192>. [PubMed: 16768595]
- Eid M, Diener E. Intraindividual Variability in Affect: Reliability, Validity, and Personality Correlates. *Journal of Personality and Social Psychology*. 1999 Apr; 76(4):662–676.
- Fleeson W. Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States. *Journal of Personality and Social Psychology*. 2001 Jun; 80(6):1011–1027. [PubMed: 11414368]
- Fleeson W. Moving personality beyond the person-situation debate the challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*. 2004; 13(2):83–87.
- Goldberg LR. The reliability of reliability: The generality and correlates of intra-individual consistency in responses to structured personality inventories. *Applied Psychological Measurement*. 1978; 2:269–291.
- Goldberg LR. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*. 1999; 7:7–28.
- Greenleaf EA. Measuring extreme response style. *Public Opinion Quarterly*. 1992; 56(3):328–351.
- Jackson DN, Messick S. Content and style in personality assessment. *Psychological Bulletin*. 1958; 55(4):243–252. [PubMed: 13567967]
- Jourard, SM. *Personality adjustment: An approach through the study of healthy personality*. New York, NY: Macmillan; 1963.
- Judge TA, Bretz RD. Reports on an alternative measure of affective disposition (CAHRS Working Paper #92-39). 1992
- Kernis MH. Measuring Self-Esteem in Context: The Importance of Stability of Self-Esteem in Psychological Functioning. *Journal of Personality*. 2005 Dec; 73(6):1569–1605. [PubMed: 16274446]
- Lanning K. Individual differences in scalability: An alternative conception of consistency for personality theory and measurement. *Journal of Personality and Social Psychology*. 1988; 55(1): 142–148.
- Larsen RJ, Ketelaar T. Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology*. 1991; 61(1):132–140. [PubMed: 1890584]
- Locke KD. H as a measure of complexity of social information processing. *Personality and Social Psychology Review*. 2003; 7(3):268–280. [PubMed: 12788691]
- Lucas RE, Baird BM. Extraversion and emotional reactivity. *Journal of Personality and Social Psychology*. 2004; 86(3):473–485. [PubMed: 15008650]
- Maslow, AH. *Toward a Psychology of Being*. Princeton, NJ: Van Nostrand; 1968.
- McFarland SG, Sparks CM. Age, education, and the internal consistency of personality scales. *Journal of Personality and Social Psychology*. 1985; 49(6):1692–1702.
- McReynolds P, Altrocchi J, House C. Self-Pluralism: Assessment and Relations to Adjustment, Life Changes, and Age. *Journal of personality*. 2000; 68(2):347–381. [PubMed: 10820690]
- Mehl MR, Pennebaker JW, Crow DM, Dabbs J, Price JH. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*. 2001 Nov; 33(4):517–523.
- Mischel W, Shoda Y. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*. 1995; 102(2):246–268. [PubMed: 7740090]

- Paulhus DL, Martin CL. Functional flexibility: A new conception of interpersonal flexibility. *Journal of personality and social psychology*. 1988; 55(1):88.
- Paunonen SV, Jackson DN. Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review*. 1985; 92(4):486–511.
- Penner LA, Shiffman S, Paty JA, Fritzsche BA. Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology*. 1994; 66(4):712–721. [PubMed: 8189348]
- Reise SP, Waller NG. Traitendness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology*. 1993; 65(1):143–151.
- Rogers, CR. On becoming a person. Oxford, England: Houghton-Mifflin; 1961.
- Rosenberg M, Schooler C, Schoenbach C. Self-esteem and adolescent problems: Modeling reciprocal effects. *American sociological review*. 1989:1004–1018.
- Rundquist EA. Response sets: a note on consistency in taking extreme positions. *Educational and Psychological Measurement*. 1950; 10:97–99.
- Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of personality and social psychology*. 1994; 67(6):1063–1078. [PubMed: 7815302]
- Schwarz N. Self-Reports: How the Questions Shape the Answers. *American Psychologist*. 1999; 54(2): 93–105.
- Siegrist M. The influence of self-consciousness on the internal consistency of different scales. *Personality and individual differences*. 1996; 20(1):115–117.
- Smillie LD, Cooper AJ, Wilt J, Revelle W. Do extraverts get more bang for the buck? Refining the affective-reactivity hypothesis of extraversion. *Journal of Personality and Social Psychology*. 2012; 103(2):306–326. [PubMed: 22582899]
- Snyder M. Self-monitoring of expressive behavior. *Journal of personality and social psychology*. 1974; 30(4):526–537.
- Spielberger, CD. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
- Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*. 1988; 54(6):1063. [PubMed: 3397865]
- Watson D, Tellegen A. Aggregation, acquiescence, and the assessment of trait affectivity. *Journal of Research in Personality*. 2002; 36(6):589–597.
- Wetzel E, Carstensen CH, Böhnke JR. Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*. 2013 Apr; 47(2):178–189.
- Wetzel E, Lüdtke O, Zettler I, Böhnke JR. The Stability of Extreme Response Style and Acquiescence Over 8 Years. *Assessment*. 2015 May. 1073191115583714.

Table 1

Correlations Between Study 1 Variables.

Measure	1	2	3	4	5	6
1. Within-Scale	0.68					
2. Contextual	0.54	0.83				
3. Simpsons	0.43	0.45	0.88			
4. Neutral Objects	0.40	0.47	0.47	0.64		
5. SCC Scale	−0.01	−0.11	0.00	0.01	0.90	
6. Relationship Satisfaction	0.04	0.01	0.16	0.17	0.45	0.93

Note. Correlations greater than .16 are significant at .05; correlations greater than .21 are significant at .01; correlations greater than .26 are significant at .001; $N=149$. “Within-Scale” indexes reflect the corrected standard deviation across items within the same scale; “Contextual” indexes reflect the corrected standard deviation across different roles; SCC = Self-Concept Clarity; Values on the diagonal are scale reliabilities (split-half for the Neutral Object scale; Cronbach’s alpha for all others). Correlations with unadjusted variability scores are available in the on-line supplement.

Table 2

Correlations Between Study 2 Variability Measures.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Within-Scale T1	0.75										
2. Within-Scale T2	0.67	0.70									
3. Contextual Self	0.49	0.46	0.92								
4. Daily Personality	0.44	0.46	0.60	0.89							
5. Daily Pos. Affect	0.25	0.26	0.39	0.64	0.87						
6. Daily Neg. Affect	0.29	0.27	0.32	0.55	0.52	0.85					
7. Contextual Friend	0.44	0.43	0.64	0.59	0.38	0.38	0.94				
8. Neutral Objects	0.46	0.41	0.44	0.34	0.26	0.25	0.40	0.51			
9. Weather	0.16	0.08	0.22	0.22	0.12	0.15	0.24	0.11	0.60		
10. SCC Scale	0.09	0.06	-0.06	-0.10	-0.29	-0.04	0.01	0.05	0.03	0.89	
11. SSS Scale	0.04	-0.02	-0.07	-0.15	-0.29	-0.12	-0.02	0.06	0.03	0.65	0.86

Note. Correlations greater than .13 are significant at .05; correlations greater than .18 are significant at .01; correlations greater than .22 are significant at .001; $N = 203$. "Within-Scale" indexes reflect the corrected standard deviation across items within the same scale; "Contextual" indexes reflect the corrected standard deviation across different roles; SCC = Self-Concept Clarity; SSS = Stability of Self Scale. Values on the diagonal are scale reliabilities (split-half for the Neutral Object scale; Cronbach's alpha for all others). Correlations with unadjusted variability scores are available in the on-line supplement.

Table 3

Correlations Between Variability and Outcome Measures

	Within-Scale T1	Within-Scale T2	Contextual Personality	Daily Personality	Daily PA	Daily NA
Life Satisfaction	0.06	0.07	-0.04	-0.01	0.19	0.01
PANAS PA	0.07	0.01	-0.02	-0.05	-0.22	0.04
PANAS NA	-0.08	-0.03	0.06	-0.05	0.24	-0.13
ITAS PA	0.11	0.11	-0.03	0.08	-0.15	0.15
ITAS NA	-0.03	0.02	0.16	0.05	0.29	-0.05
Self-Esteem	0.14	0.07	0.02	-0.05	-0.28	0.01
Optimism	0.07	0.05	-0.04	-0.10	-0.28	0.05
Depression	-0.03	0.02	0.07	0.15	0.37	0.05
Anxiety	-0.03	0.04	0.11	0.15	0.37	0.02

Note. Correlations greater than .13 are significant at .05; correlations greater than .18 are significant at .01; correlations greater than .22 are significant at .001; $N = 203$. "Within-Scale" indexes reflect the corrected standard deviation across items within the same scale; "Contextual" indexes reflect the corrected standard deviation across different roles; PANAS = Positive Affect Negative Affect Schedule; ITAS = Intensity and Time Affect Survey.

Table 4

Correlations Between Self-Esteem Variability Measures and Personality Variability and Outcomes

	Self-Esteem Variability	
	Uncorrected SD	Corrected SD
Within-Scale T1	0.07	0.25
Within-Scale T2	0.07	0.21
Contextual Self	0.18	0.31
Daily Personality	0.42	0.54
Daily Pos. Affect	0.49	0.43
Daily Neg. Affect	0.30	0.38
Contextual Friend	0.23	0.42
Neutral Objects	0.05	0.18
Weather	0.04	0.13
SCC Scale	−0.35	−0.10
SSS Scale	−0.37	−0.14
Life Satisfaction	−0.21	−0.06
PANAS PA	−0.14	0.06
PANAS NA	0.19	−0.04
ITAS PA	−0.14	0.05
ITAS NA	0.24	0.04
Self-Esteem	−0.37	−0.07
Optimism	−0.27	0.00
Depression	0.38	0.14
Anxiety	0.34	0.08

Note. Correlations greater than .13 are significant at .05; correlations greater than .18 are significant at .01; correlations greater than .22 are significant at .001; $N = 203$. “Within-Scale” indexes reflect the corrected standard deviation across items within the same scale; “Contextual” indexes reflect the corrected standard deviation across different roles; SCC = Self-Concept Clarity; SSS = Stability of Self Scale; PANAS = Positive Affect Negative Affect Schedule; ITAS = Intensity and Time Affect Survey.