THE EFFECTS OF RE-FRAMING AND RESPONDENT PERSONALITY ON PSYCHOMETRIC OUTCOMES: A REPLICATION AND EXTENSION STUDY

by

MEGHAN JOANN BABCOCK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at
The University of Texas at Arlington
August, 2018

Arlington, Texas

Supervising Committee:

William Ickes, Supervising Professor
Angela Liegey Dougall
Jared B. Kenworthy
Daniel Levine
Kelsey Medeiros

ABSTRACT

THE EFFECTS OF RE-FRAMING AND RESPONDENT PERSONALITY

ON PSYCHOMETRIC OUTCOMES:

A REPLICATION AND EXTENSION STUDY

Meghan JoAnn Babcock, Ph.D.

The University of Texas at Arlington, 2018

Supervising Professor: William Ickes

The present study investigated the effects of four item characteristics – context, direction, factor, and trait aspect switching – on inter-item correlations and participant misresponse rates for adjacent item pairs within a large set of personality scales. The responses to an online survey from 300 individuals recruited from Amazon's Mechanical Turk allowed for an investigation of these effects. The results of this study suggested that, across 16 different personality scales, adjacent item pairs that: (1) required higher levels context switching, (2) represented different factors, or (3) differed in trait aspect predicted both lower inter-item correlations and a greater percentage of misresponders. These adjacent item characteristics create distinct interpretative problems for participants and make unique contributions to the impairment of inter-item correlation and a greater proportion of misresponders. In addition to these findings, the results of this study suggest that individuals who have difficulty shifting between tasks and who are more impulsive (both in terms of failing to pay attention or acting without thinking) are less consistent in their responding to personality scales. The results of the present study provide further

evidence for the detrimental effects of the switching that is required when responding to adjacent items that vary in terms of their context specificity, content domain, or trait aspect, and also provide evidence that individuals who are less able to switch their attention and are more impulsive are less successful at responding consistently to personality scale items. A number of strengths, limitations, and recommendations for future research are discussed.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Survey research is an essential form of data collection in the social sciences, especially in the field of psychology. In fact, Saris and Gallhofer (2007) reported that approximately 49% of the empirical studies published in the *Journal of Personality and Social Psychology* between 1994 and 1995 used the survey method, making it the most common form of data collection in these areas of psychology during this time. Given the widespread use of this methodology, it is important that scale developers, especially those developing personality or attitude measures, are aware of the effects that item characteristics, item ordering, and individual differences in respondent personality can have on the psychometric properties of scales.

The focus of the present study was to examine the effects of item characteristics and item ordering in *personality scales* that assess the extent to which individuals agree or disagree with scale items or indicate the frequency of which they engage in certain behaviors or experience specific thoughts or feelings. Within the field of personality psychology, scores on self-reported personality scales are the most commonly used form of personality assessment (Paulhus & Vazire, 2007).

In the present study, I attempted to replicate the effects of three distinct types of item characteristics on inter-item correlations and rates of participant misresponse that were reported in a previous study (see Ickes et al., 2018). In addition, I examined the role of a new item characteristic, trait aspect switching, on the same two outcome measures. Finally, I examined whether there were individual differences in response consistency, namely, whether relevant

personality traits could be used to accurately predict within-person response consistency across a wide range of personality scales.

Reverse-Stated Items

A considerable amount of research has been conducted in an effort to determine the effects of reverse-stated items on the psychometric properties of personality scales. Although some researchers suggest that including reverse-stated items can extend the coverage of the construct (Weijters & Baumgartner, 2012) and can cause respondents to slow down and pay more attention to the items (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), the inclusion of reverse-stated items might result in weaker item-total correlations and lower internal consistency reliability estimates (Hamby, Ickes, & Babcock, 2016; Roszkowski & Soven, 2010). In addition to these negative consequences, reverse-stated items often form their own factor dimension, separate from the non-reversed items, in scales that are intended to be unidimensional (e.g., Greenberger, Chen, Dmitrieva, & Farruggia, 2003; Motl, Conroy, & Horan, 2000). This pattern of findings suggests that although there are potential benefits to the inclusion of reverse-stated items, they can also negatively affect the psychometric properties of personality scales.

Frame of Reference

A second item characteristic that is relevant to the proposed study is the inclusion of a specific frame-of-reference (FOR) within the various scale items (Schmit, Ryan, Stierwalt, & Powell, 1995). A FOR can be added to scale items to identify a specific context that participants should refer to when responding to items (e.g., "I am talkative *at work*" or "I am talkative *at school*"). On the other hand, noncontextualized items (e.g., "I am talkative") lack a FOR and are assumed to be more "open to interpretation by respondents" (Bing, Whanger, Davison, & VanHook, 2004, p. 151).

Research suggests that using a FOR can affect the psychometric properties of a personality scale. For example, Lievens, De Corte, and Schollaert (2008) found that scale reliability was lower when a conscientiousness scale was modified so that approximately half of the items in a scale had one FOR (i.e., at school) and the other half of the items in the same scale had a different FOR (i.e., at work). This finding suggest that scale reliability might suffer if multiple frames-of-reference are used within one scale.

Item Ordering

In addition to examining item characteristics, researchers have also investigated the effects of item ordering on psychometric outcomes. Much of the research in this area concerns the effects of item randomization versus item grouping by factor or item grouping by scale. Some researchers suggest that scale reliability is not affected by differences in item ordering (e.g., Schell & Oswald, 2012), whereas others have found that scale reliabilities are significantly lower when the items are completely randomized rather than when they are grouped by factor (e.g., Solomon & Kopelman, 1984).

Cognitive Re-framing

A feature that appears to unite the characteristics of reverse-stated items, multiple frames-of-reference, and item randomization across factors is that they require a certain amount of *cognitive re-framing.* That is, scales that include items with different directional wordings, multiple frames of reference (i.e., context), or randomized item ordering require respondents to shift their interpretative frames as they move throughout the survey.

Recently, Ickes et al. (2018) identified three forms of re-framing and examined their effects on the psychometric properties of 18 personality scales. Specifically, they examined the effects of: (1) context switching, (2) direction switching, and (3) factor switching on inter-item

correlation and participant misresponse by determining whether the adjacent item pairs within a personality scale require respondents to switch between items that have different "directions", different levels of generality/specificity, or are part of different factors.

In their study, Ickes et al. (2018) conducted novel analyses using data that were obtained in three previous online studies (see Park, Ickes, & Robinson, 2014; Robinson & Ickes, 2016; Taylor & Hamby, 2014). A total of 18 different personality scales across these three studies were used for the analyses in their study, which resulted in a total of 269 non-redundant pairs of adjacent items (see Ickes et al., 2018 for more information about these 18 scales). The manner in which each of these three forms of switching (i.e., re-framing) in adjacent item pairs was calculated in the study by Ickes et al. (2018) and their effects on the psychometric outcomes of the scales are described in the following sections due to their relevance to the present study.

Direction switching. When respondents are presented first with a positively-stated items (e.g., "I feel that I have a number of good qualities") and then with a reverse-stated item (e.g., "All in all, I am inclined to feel that I am a failure"), they must shift their interpretive frame from one pole of the dimension (high self-esteem) to the opposite pole (low self-esteem). In order to respond consistently, participants must engage in a form of cognitive re-framing, that is, they must "switch their interpretative frame to one of negation (i.e., opposite pole)" (Hamby et al., 2016, p. 492).

The results of two recent studies (Hamby et al., 2016; Ickes et al., 2018) indicate that adjacent item pairs in a scale that are mismatched in their "direction" (i.e., one item is positively-stated and the other is negatively-stated), rather than matched in their "direction" (i.e., both items are positively-stated or both items are negatively-stated), are associated with weaker inter-item correlation and greater levels of participant misresponse. Ickes et al. (2018) define participant

misresponse as a case in which a participant responds to a personality item using one side of the neutral point on a Likert-scale (e.g., agree) and then responds to the next item using the other side of the neutral point (e.g., disagree), indicating disagreement or contradiction between the two responses. In the study by Ickes et al. (2018), the participant misresponse rate for an adjacent item pair was calculated by determining the number of participants who were classified as misresponders, out of the total number of participants who selected non-neutral alternatives for both adjacent items. This allowed for a calculation of the percent of misresponders for each adjacent item pair within the same scale.

Context switching. In addition to the cognitive re-framing that is necessitated by adjacent items that differ in their direction, the inclusion of multiple frames-of-reference in a single scale can also necessitate repeated re-framing. The concept of FOR shifting is similar to the process that Hamby, Ickes, and Babcock (2016) identified as *context switching.* Ickes et al. (2018) define context switching as the "mixing [of] items that imply a broad and general view of the trait with ones that imply more of a context-specific view of the trait" (p. 1).

When participants are presented with adjacent items that vary in terms of their context specificity, respondents must engage in re-framing in order to respond to the items consistently. For example, moving from a context-general item such as "I enjoy talking to people" to a more context-specific item "I enjoy talking to people in the workplace" requires contextual re-framing on the part of the respondent (Ickes et al., 2018).

To examine the effect of context switching on psychometric outcomes, Ickes et al. (2018) presented undergraduate research assistants with the 269 adjacent item pairs from the 18 personality scales that were previously described and asked them to rate the amount of re-framing (i.e., context switching) that was required to interpret and decide how to respond to the

later-appearing item in the pair. The research assistants rated the amount of re-framing on a 3-point scale: 1 (little or no re-framing was required), 2 (more, but not much, re-framing was required), or 3 (a lot of re-framing was required). Each of the judges assigned each adjacent item pair a value for the amount of context switching that was required, and the average of the judges' ratings was used to estimate the amount of context switching that was required in each case (see Appendix A for the context switching rating task directions).

The results of the study by Ickes et al. (2018) revealed that higher levels of context switching were associated with weaker inter-item correlations and higher levels of participant misresponse. Direction switching was also examined, and it remained a significant predictor of both outcome measures when the amount of context switching was statistically controlled. As Ickes et al. (2018) suggest, these two forms of re-framing create distinct interpretative problems for participants and make unique contributions to the impairment of inter-item correlation and a greater proportion of misresponders.

Factor switching. In addition to direction switching and context switching, Ickes et al. (2018) examined whether a third form of re-framing, *factor switching*, also affected inter-item correlation and participant misresponse. To do so, Ickes et al. (2018) used the results of separate exploratory factor analyses (EFAs) that were conducted for each of the 18 personality scales to determine whether adjacent item pairs belonged to the same or different factor. The adjacent items were then classified as either a *factor match*, meaning that the items in the pair had similar content and loaded onto the same factor, or a *factor mismatch*, meaning that the items in the pair did not load on the same factor.

In the study by Ickes et al. (2018), factor switching was a significant predictor of inter-item correlation (but not participant misresponse), such that adjacent items that were part of the

same factor had stronger inter-item correlations than adjacent items that were mismatched on factor. When the three forms of re-framing (i.e., direction, context, and factor) were simultaneously entered into the same multilevel model, factor switching remained a significant predictor of inter-item correlation.

In addition to providing evidence that these forms of cognitive re-framing were unique predictors of inter-item correlation and participant misresponse, Ickes et al. (2018) also tested a series of multilevel mediation models and determined that context switching acted as a partial mediator for the relationship between direction switching and both inter-item correlation and participant misresponse. In addition, context switching acted as a partial mediator for the relationship between factor switching and inter-item correlation.

Because the study by Ickes et al. (2018) was the first to examine all three of these forms of switching variables and their effects on both adjacent inter-item correlation and rates of participant misresponse, I sought to replicate these findings in a new sample of respondents for the present study. In addition to determining whether these three forms of cognitive re-framing would be significant predictors of inter-item correlation, I sought to extend the previous findings by examining the effects of a new form of cognitive re-framing that I have identified as "trait aspect switching".

Trait aspect switching. Scale items can differ in their trait aspect, that is, the item content can refer to respondents' *behaviors* or to their *subjective experience* (i.e., their thoughts or feelings). For example, the scale item in Goldberg et al.'s (2006) Dependence Scale "I need the approval of others" would be classified as an item relating to respondent's subjective experience, whereas the scale item "I show my sadness" would be classified as an item relating to the respondent's behavior. If these two items were presented consecutively within the

Dependence Scale, these adjacent items would be classified as a *trait aspect mismatch* because one item refers to a respondent's subjective experience and the other to his/her behavior, and thus require trait aspect switching. On the other hand, adjacent items that both concern behaviors or that both concern subjective experiences would be classified as *trait aspect matches.*

To my knowledge, the effect of trait aspect switching within scales has not been investigated in previous research.[1] In the present study, I examined the effects of trait aspect switching as well as those of context, direction and factor switching. Similar to the other forms of re-framing, trait aspect switching was examined as a predictor of both inter-item correlation and participant misresponse. Consistent with the previous findings for the other three cognitive re-framing predictors, I hypothesized that trait aspect switching would be a unique predictor of the same outcome variables even when the effects of the other three forms of re-framing were statistically controlled. Based on previous findings by Ickes et al. (2018), I did not predict any significant interaction effects among the four forms of re-framing, although these interaction effects were tested in the statistical models.

Respondent Confusion

Ickes et al. (2018) proposed that each of the three forms of re-framing (i.e., direction switching, context switching, and factor switching) lead to decreased inter-item consistency and greater levels of participant misresponse because of *respondent confusion*. They suggested that, for example, when respondents encounter adjacent items that are mismatched in direction, they become confused because they are forced to make a "metaphorical U-turn" and switch from using one pole of the trait dimension to using the opposite pole.

[1] The idea for investigating of trait aspect switching was originally suggested in an email communication from Lew Goldberg to William Ickes (September 19, 2016).

Similarly, Ickes et al. (2018) proposed that when respondents encounter adjacent items that vary in terms of their generality/specificity, they become confused about what construct is being measured. As Hamby et al. (2016, p. 2) suggested, if respondents are first presented with a broad, context-general item and then presented with a more context-specific item (or vice versa), it is likely that they will experience confusion about the generality/specificity of the construct that the items are measuring. Finally, when respondents encounter adjacent items that vary in their content domain, and thus load on different factors, they are, once again, likely to experience confusion about why the content differs (Ickes et al., 2018).

Evidence for this notion of respondent confusion can be seen in the weaker inter-item correlations and higher rates of misresponse for item-pairs that require these forms of switching. The results of the study by Ickes et al. (2018) suggest that each of these forms of switching account for independent variation in inter-item correlation and participant misresponse (with the exception of factor switching predicting participant misresponse). Trait aspect switching might be a fourth distinct form of re-framing that also causes confusion as to whether the same trait or different traits is/are being measured within the scale.

Individual Differences

Although we have knowledge about the effects of different item characteristics and forms of re-framing on psychometric outcomes, few researchers have investigated whether individual differences in personality and response styles also affect psychometric outcomes. One relevant variable to investigate in this regard is within-person response consistency, a measure that is operationalized as individual-level variability in the item responses within a given scale. Within-person response consistency is relevant to the present study because it has been shown to be related to scale reliability, such that the scale reliabilities of the Big Five personality traits were

significantly lower in a sample of inconsistent respondents compared to a sample of consistent respondents (Reddock, Biderman, & Nguyen, 2011).

Few studies to date have investigated the role of personality differences on survey responding. In one study, Couch and Keniston (1960) reported that higher levels of impulsiveness and extraversion were associated with acquiescent responding, but the researchers did not explore the effect of these personality traits on psychometric outcomes such as within-person response consistency or scale reliability. As a second example, researchers found that inconsistent responding was associated with lower cognitive ability, as measured by the Wonderlic Personnel Test (Biderman, 2007; Reddock et al., 2011).

In order to determine whether individual differences in personality affect relevant psychometric outcomes, I examined the role of personality predicting individual-level response consistency. As mentioned in a previous section, one way in which reliability has been measured at the individual level is by calculating a measure of "within-person response consistency", which is achieved by calculating the standard deviation of each participant's responses in a particular scale, such that a larger standard deviation is indicative of a greater level of inconsistent responding (Reddock et al., 2011). In the proposed study, I calculated an *overall* within-person response consistency score for each participant by first calculating the standard deviation of their responses for each of the 16 personality scales, and then averaging these deviations across the set of scales. This approach is identical to the one used by Reddock et al. (2011), who calculated overall within-person consistency by averaging the standard deviations of each individual's responses to each of the Big Five personality traits.

Although it is possible to calculate an individual-level of response consistency, there is not, on the other hand, an appropriate measure of individual-level misresponse that is analogous

to the one used in the study by Ickes et al. (2018). In the study by Ickes et al. (2018), misresponse rates for adjacent item pairs were calculated by determining the percentage of participants who responded to the first item in an adjacent item pair by using one side of the neutral point on a Likert-scale (e.g., agree) and then contradicted themselves by responding to the next item using the other side of the neutral point (e.g., disagree), out of all participants who selected non-neutral response alternatives for both of these adjacent items.

In order to calculate the rate of misresponse for *an individual respondent* across a set of adjacent item pairs, it would be necessary to calculate the percentage of times that he/she misresponded to adjacent item pairs in the survey, out of all instances where he/she provided two non-neutral responses to adjacent item pairs. Because the number of times that a participant responds to adjacent items with two non-neutral responses can (and will) differ across individuals, individual-level misresponse rates can provide an idiosyncratic measure of misresponse that cannot be compared across the individual participants within a sample.[2] For this reason, the proposed study will focus on the effects of personality traits only on within-person response consistency.

Individual differences in response consistency. To assess whether there are individual differences in response consistency, I examined a set of personality traits that I hypothesized had the potential to affect overall response consistency: attentional control, attentional and motor impulsivity, cognitive flexibility, conscientiousness, and need for cognition. These personality traits were chosen based on their emphasis on either individual differences in attention and competence (attentional control, conscientiousness, and impulsivity) or individual differences in

[2] For example, if Participant A misresponds 3 times out of 3 non-neutral responses, his/her misresponse rate would be 100%. Similarly, if Participant B misresponds 99 times out of 100 non-neutral responses in the same survey, his/her misresponse rate would be 99%. The two misresponse rates for Participant A and B, although nearly equivalent in terms of percentages, are qualitatively different and should not be treated as nearly equivalent misresponse rates.

the extent to which people approach and succeed in effortful cognitive tasks (cognitive flexibility and need for cognition).

The inclusion of these personality traits in a model predicting overall within-person response consistency will provide evidence as to whether individuals with certain personality traits can successfully navigate a scale that includes different forms of re-framing by responding in a consistent manner throughout a survey. In addition, based on the multiple regression coefficients, I can determine whether certain personality traits (i.e., ones focused on attention or cognition) explain more or less variance in within-person response consistency.

Research Objectives

In the proposed study, I had three main research objectives. The first objective was to conduct the first replication study following that of Ickes et al. (2018) to determine whether I could replicate their findings in new sample using participants obtained from Amazon's Mechanical Turk (MTurk; www.mturk.com). I hypothesized, based on the results of the study by Ickes et al. (2018), that context switching, direction switching, and factor switching would be significant predictors of inter-item correlation, and that both context and direction switching would be significant predictors of participant misresponse. Specifically, I hypothesized that inter-item correlation would be significantly weaker for adjacent item pairs that required more context switching, were mismatched in direction, and belonged to different factors. In addition, I hypothesized that participant misresponse rates would be significantly higher for adjacent item pairs that required more context switching and were mismatched in direction. I hypothesized that all of these effects would remain significant even when item-pair length and item length discrepancy (proxy variables for context generality/specificity and context switching, respectively; see Hamby and Ickes, 2015; Hamby et al., 2016) were included in the models. In

addition, I sought to determine whether context switching served as a significant mediator of the relationships between the other forms of re-framing and inter-item correlation and participant misresponse. Based on the results of the study by Ickes et al. (2018), I hypothesized that context switching would mediate the relationship between the other forms of re-framing and both outcome variables.

The second objective was to determine whether a new re-framing variable, trait aspect switching, was also a significant predictor of inter-item correlation and participant misresponse. I specifically predicted that inter-item correlation would be significantly weaker and participant misresponse rates would be significantly higher for adjacent item pairs that were mismatched on trait aspect compared to pairs that were matched on trait aspect.

The third objective was to determine whether there were individual differences in within-participant consistency of responding. I hypothesized that individuals who are cautious, competent, and tend to engage in effortful cognitive activities (especially activities that require attentional shifting) would be more consistent in their responding to personality scale items than individuals who are careless, inefficient, and do not prefer to engage in effortful cognitive activities.

To address these objectives, I collected data on the same 18 personality scales that were included in the study by Ickes et al. (2018) from a new sample of participants using Amazon's Mechanical Turk (MTurk). In addition to collecting the participants' responses to the 18 personality scales, I collected demographic data and additional personality data to measure the personality traits that I identified as having the potential to affect within person response consistency.[3]

[3] It is important to note that 2 of the 18 personality scales that were included in the analyses as part of the Ickes et al. (2018) study served as either a *personality predictor* (Need for Cognition) or *covariate* (Social Desirability) of

CHAPTER 2

METHOD

Participants

The participants in this study were 300 individuals who ranged in age from 18 to 70 years (M = 36.38, SD = 10.67) and resided in the United States.[4] Approximately 50.7% participants were male, 49% were female, and 1 participant (.3%) did not report their gender. The sample was 74.3% White, 8.7% Black, 5% Hispanic/Latino, 7% Asian, 1.7% Native American, and 3.3% Other or Multiracial. The majority of participants reported that they had completed some college or university education (38%) or had received a four-year degree from a college or university (39%). The remaining participants reported that they had earned no high school diploma or GED (.3%), a high school degree or GED (15%), a Master's degree (6%), or a Doctoral degree (1.7%).

Procedure

Participants were recruited from Amazon's Mechanical Turk (MTurk), an online service that allows "requesters" (i.e., researchers) to collect survey data from "workers" (i.e., participants) in exchange for financial compensation. Participants who signed up for this study received a website link that directed them to the online survey on Survey Monkey. Upon

---

response consistency in the present study. Therefore, only 16 of the 18 personality scales included in the study by Ickes et al. (2018) were used to assess the effects of re-framing on inter-item correlation and participant misresponse in the present study. This decision was made so that scores on these scales did not serve as both predictor variables (in the models tested in Research Objective 3) *and* as part of the outcome measures for calculations of inter-item correlation and misresponse (in the models tested in Research Objectives 1 and 2). To eliminate reader confusion, scores on the Need for Cognition and Social Desirability scales were only tested as predictors of response consistency.

[4] A total of 356 individuals completed the online survey, but only 300 of them correctly to all four of the attention-check items that were included in the survey. Each of the four attention-check items asked participants to mark the "Other" response option as a means to demonstrate their attention to the item content.

completion of the online survey, participants were compensated $2.00 in exchange for their participation. This amount is consistent with the amount that participants are paid per item in other personality studies using MTurk as a means for data collection (see, for example, Hamby & Taylor, 2016).

Measures

Participants in the study completed an online survey that consisted of three main parts: (1) demographic information, (2) individual difference measures to identify individuals with varying levels of attentional control, cognitive flexibility, conscientiousness, impulsivity, need for cognition, and social desirability, and (3) individual difference measures to extract item properties and psychometric information (i.e., inter-item correlation between adjacent items, rates of participant misresponse, context switching, direction switching, factor switching, trait aspect switching, and within-participant response consistency).

In total, the online survey consisted of 346 items: 4 demographic items; 79 items to assess individual differences in attention, competence, the tendency to approach and succeed in effortful cognitive tasks, and social desirability; 259 items that were used to extract item properties and psychometric information; and 4 attention-check items that were spaced throughout the survey.

All of the individual difference items included in the second and third parts of the online survey were measured on 5-point Likert-type scales, with the exception of the Affect Intensity for Anger and Frustration scale which was measured on a 6-point Likert-type scale, with response options ranging from (1) *never* to (6) *always*.

Part 1: Demographics. Participants completed four demographic items at the beginning of the survey to assess their: (1) age, (2) gender, (3) race/ethnicity, and (4) highest level of educational attainment.

Part 2: Individual difference measures. After responding to the demographic items, the participants responded to five personality scales to measure their level of attentional control, cognitive flexibility, conscientiousness, attentional and motor impulsivity (represented by separate scores on two subscales on an impulsiveness scale), and need for cognition. In addition to these personality scales, a measure of social desirability was included in the survey in order to model it as a potential covariate in the appropriate analyses. These six personality scales from Part Two of the survey are discussed in the sections below.

Attentional Control Scale (ACS; Derryberry & Rothbart, 1988). The ACS is a 20-item self-report scale that includes two subscales: (1) Attentional Focus (e.g., "It is hard for me to concentrate on a difficult task when there are noises around") and (2) Attentional Shifting ("I can quickly switch from one task to another"). The 10-item Attentional Shifting subscale measures the "the capacity to intentionally shift the attentional focus to desired channels, thereby avoiding unintentional focusing on particular channels" (Derryberry & Rothbart, 1988, p. 966) and was the subscale included in this study. Researchers have reported acceptable internal consistency reliability for scores on the Attentional Shifting subscale ($\alpha = .76$; Quigley, Wright, Dobson, & Sears, 2017).

Barratt Impulsiveness Scale (BIS-11; Patton, Stanford, & Barratt, 1995). The BIS-11 is a 30-item self-report scale that measures the personality and behavioral construct of impulsiveness and consists of three second-order subscales: (1) Attentional (e.g., "I don't pay attention"); (2) Motor (e.g., "I do things without thinking"); and (3) Non-planning (e.g., "I plan

tasks carefully"). According to Stanford et al. (2009), the Non-planning subscale measures individual differences in forethought when making decisions and is not theoretically relevant to response consistency on personality scales. Therefore, only the 8-item Attentional and the 11-item Motor subscales from the BIS-11 were used in the present study.

The BIS-11 has high internal consistency reliability for responses on the entire scale (α = .83) and acceptable reliability for the three second-order subscales (αs ranging from .59 to .74) (Stanford et al., 2009). In addition, Stanford et al. (2009) provided evidence of acceptable test-retest reliability for scores one month apart for the total (Spearman's Rho = .83) and second-order subscales (Spearman's Rho values ranging from .61 to .72).

Cognitive Flexibility Inventory (CFI; Dennis & Vander Wal, 2010). The CFI is a 20-item self-report scale that is composed of: (1) an Alternatives subscale that measures the ability to identify alternative explanations and generate multiple solutions; and (2) a Control subscale that measures the ability to view difficult situations as controllable (Dennis & Vander Wal, 2010). The 13-item Alternatives subscale was included in the proposed study given its relevance to responding to scale items that require respondents to switch between different viewpoints. Sample items from this subscale include: "I consider multiple options before making a decision" and "I often look at a difficult situation from different viewpoints."

Dennis and Vander Wal (2010) reported high levels of internal consistency reliability for total scores on the CFI (α = .91) and evidence of high test-retest reliability across seven weeks for scores on the Alternatives subscale (r = .75).

Conscientiousness (John, Donahue, & Kentle, 1991). The 9-item conscientiousness scale from the Big Five Inventory (BFI) was also included in the present study. This scale measures individual differences in competence, achievement striving, and orderliness (e.g., "I

see myself as someone who is a reliable worker" and "I see myself as someone who does things efficiently"). Researchers have reported internal consistency reliability estimates for the conscientiousness scale from the BFI (α = .82) that are comparable to the reliability estimates for other conscientiousness scales (e.g., Costa & McCrae's (1992) NEO-Five Factor Inventory; see John, Naumann, & Soto, 2008).

Need for Cognition Scale (NFC; Cacioppo & Petty, 1982). The short form of the NCS, which measures the tendency to engage in effortful cognitive tasks, was included in the present study. The shortened 18-item version has high internal consistency reliability (α = .90) and is strongly correlated with scores on the full 34-item version of the NFC (r = .95, p < .001) (Cacioppo, Petty, & Kao, 1984). Sample items from the NFC scale include: "I would prefer complex to simple problems" and "I only think as hard as I have to" (reverse-scored).

Shortened Version of the Social Desirability Scale (SD; Strahan & Gerbasi, 1972). The 10-item shortened version of the 33-item Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) was included in the proposed study as a control variable (i.e., covariate) in the appropriate data analyses. This scale measures individual differences in the tendency to answer questions in a socially desirable manner (e.g., "I have never deliberately said something to hurt someone's feelings."). The shortened 10-item version has high internal consistency reliability (α = .88) and is strongly correlated (r = .96) with the original 33-item version (Fischer & Fick, 1993).

Part 3: Measures for psychometric investigation. In the third part of the survey, participants responded to the 16 personality scales that were also included in the analyses of the 18 personality scales conducted by Ickes et al. (2018). The data for these 16 measures allowed for the calculation of the following predictors and outcome measures in the present sample: (1)

inter-item correlation between adjacent items, (2) levels of participant misresponse, (3) item-pair length, (4) item-length discrepancy, (5) context switching, (6) direction switching, (7) factor switching, (8) trait aspect switching, and (9) within-participant response consistency. See Table 1 for a list of these 16 personality scales and those included in Part 2 of the online survey, along with the reliability estimates from the present sample, and the original scale citations.[5]

Predictor Variables

To address the first research objective, that is, to replicate the findings reported by Ickes et al. (2018), five variables served as predictors of inter-item correlation and the percentage of misresponders: (1) item-pair length, (2) item-length discrepancy, (3) context switching, (4) direction switching, and (5) factor switching. To address the second research objective, (6) trait aspect switching served as a sixth predictor variable in two identical models predicting inter-item correlation and participant misresponse.

To address the first two research objectives, adjacent item pairs within each scale were first identified. Adjacent item pairs represent all nonredundant instances of two adjacent scale items within a particular scale. For example, the first pair of adjacent items in the Need to Belong scale (Leary et al., 2013) was composed of the first item in the scale ("If other people don't seem to accept me, I don't let it bother me") and the second item in the scale ("I try hard not to do things that will make other people avoid or reject me"), and the second pair of adjacent items was composed of the second item ("I try hard not to do things that will make other people avoid or reject me") and the third item in the scale ("I seldom worry about whether other people care about me").

[5] These 16 personality scales were originally chosen by Ickes et al. (2018) on the basis that they were: (1) rated with a Likert-type response format, and (2) had a minimum of 10 items in order to ensure a relatively large number of adjacent item pairs for each scale.

A total of 243 non-redundant pairs of adjacent items across the 16 personality scales were created.[6] Once the adjacent item pairs were created, the predictor variables measuring item-pair length, item-length discrepancy, and the four forms of re-framing were calculated. These six predictor variables are described in the sections below.

Item-pair length. The item-pair length (i.e., word count) of the adjacent item pairs was calculated by averaging the word count of the two adjacent items. For example, the first item of the Need to Belong scale ("If other people don't seem to accept me, I don't let it bother me") is composed of 14 words and the second item in the adjacent item ("I try hard not to do things that will make other people avoid or reject me") is composed of 16 items. The item-pair length for this adjacent item pair would be 15, that is, the average of the two adjacent item word counts.

Item-length discrepancy. The item-length discrepancy of the adjacent item pairs was determined by calculating the absolute difference between the item length (i.e., word count) of the two items in the adjacent item pairs. Using the same adjacent item pair described above as an example, the item-length discrepancy would be 2 (i.e., the absolute value of 14 words minus 16 words).

In two earlier investigations, Hamby and colleagues examined the effects of item-pair length on internal consistency reliability (Hamby & Ickes, 2015; Hamby et al., 2016) and average inter-item correlation (Hamby et al., 2016) across a large number of personality scales. In both studies, personality scales with fewer words per item resulted in higher internal consistency reliability estimates. In Hamby et al. (2016), item pairs within the same scale with less average words per item were associated with larger inter-item correlations.

[6] Because the first item in a given scale was not compared to a previous item, there was always one less adjacent item pair than the total number of items in scale. For example, the 10-item Need to Belong scale included 9 adjacent item pairs.

In regard to item-length discrepancy, Hamby et al. (2016) found that personality scales with greater variability in item length and item pairs within the same scale with greater variability in length were associated with lower internal consistency reliability estimates and smaller inter-item correlations, respectively.

To account for these findings, Hamby, Ickes, and colleagues (2015; 2016) proposed that shorter items imply a more *context-general* view of a given trait (e.g., I enjoy talking to people), whereas longer items imply a more *context-specific* view of the same trait (e.g., I enjoy talking to people in the workplace). Further, they proposed that when items of varying length are included in the same personality scale, participants are likely to become confused "about the generality versus specificity of the trait being measured" (Ickes et al., 2018, p. 1), and are required to engage in a process these same researchers later identified as *context switching*.

In these earlier studies, item-pair length served as a proxy measure of context generality or specificity and item-length discrepancy served as an indirect measure of context switching. In later research, Ickes et al. (2018) measured context switching more directly by having research assistants rate the amount of context switching that was required in order to interpret successive items within the same personality scale. In the latest study, Ickes et al. (2018) found that the more direct measure of context switching was a significant predictor of both inter-item correlation and participant misresponse even when the effects of item-pair length and item-length discrepancy were statistically controlled in a multilevel model.

In the present study, these two indirect measures of context generality/specificity and context switching were measured and included as potential predictor variables in the multilevel models that were tested. Based on previous research, I hypothesized that neither item-pair length nor item-length discrepancy would be significant predictors of inter-item correlation or

participant misresponse when entered in a model with the more direct measure of context switching (described in the following section).

Context switching. To measure the level of context switching between the items that composed adjacent item pairs within the same scale, a large group of undergraduate and graduate research assistants completed a rating task that was developed for the previous study in this area (see Ickes et al., 2018). The research assistants were asked to rate all pairs of adjacent items within a given scale for the amount of "re-framing" they had to do in order to interpret the later-appearing item within the item pair. The amount of re-framing was rated on a 3-point scale: (1) "little or no re-framing was required", (2) "more, but not much, re-framing was required, and (3) "a lot of re-framing was required".

The average interrater reliability for the context switching ratings across the 16 personality scales was high, considering the subjective nature of the task (interrater α values ranged from .85 to .89). The specific directions for the context switching task that were provided to the research assistants can be found in Appendix A, and a sample context switching rating task for the Need to Belong scale can be found in Appendix B.

Direction switching. To determine which adjacent item pairs required a re-framing in the directional perspective, each adjacent item pair was classified as a *directional match* (i.e., both items in the adjacent item pair were worded in the same direction) or as a *directional mismatch* (i.e., the items in the adjacent item pair were worded in opposite directions). Item pairs that were classified as directional mismatches required a re-framing or switch in directional perspective, whereas directional matches did not require this form of re-framing. This predictor was dummy-coded such that a directional matching was coded as "0" and directional mismatching was coded as "1".

To illustrate direction switching, consider again the first adjacent item pair in the Need to Belong scale. This adjacent item pair would be classified as a *directional mismatch*, because the first item was negatively-stated ("If other people don't seem to accept me, I don't let it bother me") and the second item in the adjacent item pair was positively-stated ("I try hard not to do things that will make other people avoid or reject me"). In this item pair, agreement with the first item indicates *lower* levels of needing to belong (and thus requires reverse-scoring) whereas agreement with the second item indicates *higher* levels of needing to belong (and does not require reverse-scoring).

Factor switching. To determine whether adjacent item pairs required a re-framing due to factor switching, item pairs were classified as belonging to the same or different factor based on a series of exploratory factor analyses (EFAs) using principal axis factoring with varimax rotation for each of the 16 scales. Adjacent items that belonged to the same factor were classified as a *factor match*, whereas adjacent items that belonged to different factors were classified as a *factor mismatch*. This predictor was dummy-coded such that a factor match was coded as "0" and factor mismatch was coded as "1".

The 16 scales in the present study were intended to be unidimensional, yet content-based differences were found for 5 of the 16 scales. For example, Lennox and Wolfe's (1984) 13-item Self-Monitoring Scale was marked as multidimensional and as having two separate factors with items that represented: (1) the ability to modify one's self-presentation (e.g., "In social situations, I have the ability to alter my behavior if I feel that something else is called for") and (2) sensitivity to the expressive behavior of others (e.g., "I am often able to read people's true emotions correctly through their eyes").

It is important to note that the content-based differences that led to the identification of different factors were not due to uniform directional wording or to item-response distributions that were skewed in the same direction (e.g., all responses positively skewed or negatively skewed). In cases where the EFA suggested a multidimensional scale but there were no content-based differences in items loading onto different factors, the scale was treated as unidimensional. Out of the 16 scales, a total of 5 were treated as multidimensional, with the number of factors ranging from two to three.[7] Of the 243 adjacent item pairs, 205 (84.4%) were factor matches and 38 (15.6%) were factor mismatches.

Trait aspect switching. To determine whether adjacent item pairs within the same scale were matched or mismatched on trait aspect, independent ratings of trait aspect for each adjacent item pairs were collected. A total of 11 undergraduate research assistants who were blind to the purpose of this study completed a trait aspect rating task that was developed specifically for this study. The research assistants were asked to rate the 259 individual difference items included in part three of the survey based on whether each item asked respondents to report on either their *behaviors* or their *subjective experience* (i.e., their thoughts or feelings). The directions for the rating task and a sample rating task for the Dependence Scale can be seen in Appendix C.

Because the required judgments were somewhat subjective, I identified a core group of raters whose ratings indicated a high level of interrater agreement using both factor analyses and reliability analyses. When rater was treated as a variable in an exploratory factor analysis with varimax rotation, nine of the eleven raters loaded on the first principal component. When the two raters who loaded on the second principal component were removed, the interrater reliability for trait aspect ratings was quite high ($\alpha = .89$). The ratings made by the nine individuals whose

[7] The five multidimensional scales were those measuring: (1) need for closure, (2) preference for consistency, (3) need to belong, (4) self-monitoring, and (5) emotional contagion. See Table 2 for more information about these five multidimensional scales.

ratings loaded highly on the first principal component were used in the analyses involving this predictor.

Items were classified as measuring respondents' behaviors when the majority of the raters (i.e., five or more raters) classified the item as a behavior, whereas items were classified as measuring respondents' subjective experiences when the majority of raters (i.e., five or more raters) classified the item as a thought or feeling. Next, adjacent item pairs were classified as *trait aspect matches* when both items in the pair were rated as measuring behaviors or subjective experiences or as *trait aspect mismatches* when one item in the pair was rated as measuring a behavior and the other a subjective experience. This predictor was dummy-coded such that a trait aspect match was coded as "0" and trait aspect mismatch was coded as "1".

Lastly, in order to address the third research objective, scores on the six personality traits from the second part of the online survey served as potential predictors of within-participant response consistency on the traits measured in the third part of the online survey, an outcome variable which is described in the following section.

Outcome Variables

To address the three research objectives, a total of three outcome variables were included in this study. To determine whether the effects reported in Ickes et al. (2018) could be replicated in a new sample and whether an additional form of re-framing was a significant predictor, both inter-item correlation and participant misresponse served as outcome variables. To address the third research objective, within-person response consistency served as the outcome variable.

Inter-item correlation. To measure the strength of the relationship between responses to adjacent item pairs within the same scale, Pearson's correlation coefficients were calculated. For example, inter-item correlation coefficients were calculated for Item 1 with Item 2 within a

particular scale, Item 2 with Item 3 within the same scale, Item 3 with Item 4, and so on. The inter-item correlations for each adjacent item pair were then transformed using Fisher's *r* to *z* transformation to make the inter-item correlations more normally distributed.

Participant misresponse. To measure participant misresponse rates to adjacent item pairs within the same scale, the percentage of individuals who responded to one item using one side of the neutral point on a 5-point Likert-type scale (e.g., agree) and then responded to the next item in the same scale using the opposite side of the neutral point (e.g., disagree), out of all individuals who selected non-neutral alternatives for both items in the pair, were calculated. Because the response options for the AIAF scale range from 1 ("never") to 6 ("always") with no true midpoint, scores of both "3" and "4" were treated as *neutral* responses, and a misresponse for this scale was recorded when a participant responded to one item in an adjacent item pair with a response of "1" or "2" and then responded to other item in this pair with a "5" or "6".[8]

Within-person response consistency. To measure the extent to which a participant responded in a consistent manner across scale items, the standard deviation of each participant's responses in a particular scale was calculated. To calculate a measure of *overall* within-person response consistency, the standard deviations from the 16 personality scales were averaged for each participant (*M* overall response consistency across scales = .85, *SD* = .18).

[8] It is important to note that all of the reverse-keyed items in each scale were reverse-scored prior to determining the rates of participant misresponse. This step was taken to ensure that there was not a built-in correlation between direction switching and participant misresponse.

CHAPTER 3

RESULTS

Data Analytic Strategy

To determine whether the results reported by Ickes et al. (2018) replicated in the present sample (Research Objective 1) and to examine the utility of trait aspect switching predicting inter-item correlation or participant misresponse (Research Objective 2), a series of multilevel models were estimated.

In the present sample, adjacent item pairs (measured at Level 1) were clustered within 1 of 16 personality scales (measured at Level 2). Because ordinary least squares (OLS) regression assumes that all observations are independent of one another, OLS regression was not an appropriate statistical analysis to test the hypotheses outlined in the first two research objectives. In fact, using OLS regression with hierarchical data can lead to smaller standard error estimates and inflated Type 1 error rates (Krull & MacKinnon, 2001). Multilevel modeling addresses the limitations of OLS regression by considering the interdependence in the data, in this case, adjacent items nested within different personality scales. Therefore, to address Research Objectives 1 and 2, multilevel modeling with the restricted maximum likelihood (REML) estimation method was used.[9]

To determine whether there were individual differences in within-participant consistency of responding (Research Objective 3), I first calculated the zero-order correlations between the set of six personality traits and within-participant response consistency. Next, the significant

[9] These multilevel models were identical to those estimated in Ickes et al. (2018), with the addition of trait aspect switching as a potential predictor in the models testing the hypotheses in Research Objective 2.

personality correlates were entered into a multiple regression model predicting within-participant response consistency.

Data Screening

Prior to any analyses, all data were screened to ensure that all necessary assumptions of the proposed analyses were met. All of the 23 personality scales included in the present study were normally distributed, with skewness values less than West, Finch, and Curran's (1995) skewness cut-off value of +/-2. In the present sample, the skewness values for the 23 personality scales ranged from -.96 to .91 (SE_{skew} = .14). In addition to the personality scales, the continuous predictors of inter-item correlation and participant misresponse (i.e., item-pair length, item-length discrepancy, and context switching) were normally distributed, with skewness values ranging from .004 to 1.09 (SE_{skew} = .16). There was also evidence that the residuals for the outcome measures were normally distributed, as indicated by the normal probability plots of the residuals and the scatterplots that plot the standardized residuals against the standardized predicted values. See Table 3 for the descriptive statistics for each predictor and outcome variable, separated by scale. See Table 4 for the zero-order correlations between each predictor and outcome variable. It is important to note that there is a significant amount of overlap between direction switching and both context switching (r = .32, p < .005) and factor switching (r = -.26, p < .001).

There were a few instances of univariate outliers in the data set, but these data points remained in our analyses as legitimate cases. Inspection of the Mahalanobis distance values indicated that there were no instances of multivariate outliers. Finally, there was no evidence of multicollinearity among the predictors, as indicated by a number of collinearity statistics and

diagnostics (e.g., Pearson's *r* correlations, tolerance levels, variance inflation factors, and condition indices).

Research Objective 1

Two multilevel models predicting inter-item correlation and participant misresponse were estimated to explore the first research objective. The results of these two models are outlined in the following sections.

Predictors of inter-item correlation. The multilevel model predicting inter-item correlation was identical to the one used in Ickes et al. (2018), and examined the effects of five predictors: (1) context switching, (2) item-pair length, (3) item-length discrepancy, (4) direction switching, and (5) factor switching. The random intercept only model predicting inter-item correlation significantly improved on the fixed intercept-only model according to the likelihood ratio test, $\chi^2(1) = 132.13$, $p < .001$. The random intercept-only model yielded a significant intraclass correlation coefficient (ICC) = .52, indicating that differences among the scales accounted for a substantial proportion of variance in inter-item correlation between the adjacent items. This finding justified the decision to create a different intercept for each scale to control for any differences in inter-item correlation across scales.

In addition to the random intercept, the multilevel model included the five predictors that were estimated as fixed effects (i.e., the effects of each predictor on inter-item correlation was constant across scales). The following multilevel linear model was used to predict inter-item correlation:

$$z_{ij} = b_{0j} + b_1 CS_{ij} + b_2 Length_{ij} + b_3 Discrepancy_{ij} + b_4 Direction_{ij} + b_5 Factor_{ij} + e_{ij}$$

where $i = 1, \ldots, 243$ indexes the specific item-pair being considered; $j = 1, \ldots, 16$ indexes the specific scale being considered; z_{ij} is inter-item correlation transformed to z; b_{0j} is the random

intercept reflecting differences in inter-item correlation between the scales; b_1CS_{ij} is the amount of context switching; $b_2Length_{ij}$ is the item-pair length; $b_3Discrepancy_{ij}$ is the item-length discrepancy; $b_4Direction_{ij}$ is a dummy variable with 0 indicating that both items were reverse-scored or neither item was reverse-scored and 1 indicating that one item was reverse-scored and the other item was not reverse-scored; $b_5Factor_{ij}$ is a dummy variable with 0 indicating that both items loaded on the same factor and 1 indicating that they loaded on different factors; b_1, b_2, b_3, b_4 and b_5 are regression coefficients for the predictors; and e_{ij} is the residual error term.

The results of this multilevel model revealed that context switching and factor switching were the only significant predictors of inter-item correlation. As hypothesized, the average inter-item correlation was significantly lower for adjacent item pairs that required more context switching (B = -.18, *SE* = .02, β = -.37, $p < .001$) and for item pairs that loaded on different factors rather than the same factor (B = -.20, *SE* = .04, β = -.27, $p < .001$). Contrary to my hypothesis (and the results reported in the previous study), direction switching was not a significant predictor of inter-item correlation (B = -.03, *SE* = .02, β = -.06, $p = .19$).

Finally, the effects of item-pair length and item-length discrepancy were not significant as unique predictors of inter-item correlation (see Table 5 for all model parameters). As expected, the indirect measures of context generality/specificity (item-pair length) and context switching (item-length discrepancy) were not significant as predictors of inter-item correlation when the more direct measure of context switching (i.e., one measured by research assistants' ratings) was included in the multilevel model. The lack of significance of these two predictors replicates previous findings in this area (see Ickes et al., 2018).

After determining that context switching and factor switching were unique predictors of inter-item correlation, I tested to see if these two variables interacted to predict inter-item

correlation. The result of this analysis revealed no significant two-way interaction between context switching and factor switching predicting inter-item correlation (p = .13). All other possible two-way and three-way interactions were tested in additional multilevel models and none of the interaction effects were significant (ps > .05).

Predictors of participant misresponse. The multilevel model predicting participant misresponse was identical to the model used in Ickes et al. (2018) and examined the effects of the same five predictors in the model reported in the previous section. The random intercept only model predicting participant misresponse significantly improved on the fixed intercept-only model according to the likelihood ratio test, χ^2 (1) = 71.77, p < .001. The random intercept-only model yielded a significant intraclass correlation coefficient (ICC) = .31, indicating that differences among the scales accounted for a substantial proportion of variance in inter-item correlation between the adjacent items. This finding justified the decision to create a different intercept for each scale to control for any differences in participant misresponse across scales.

In addition to the random intercept, the multilevel model included the five predictors that were estimated as fixed effects (i.e., the effects of each predictor on participant misresponse was constant across scales). The following multilevel linear model was used to predict participant misresponse:

$$Misresponse_{ij} = b_{0j} + b_1 CS_{ij} + b_2 Length_{ij} + b_3 Discrepancy_{ij} + b_4 Direction_{ij} + b_5 Factor_{ij} + e_{ij.}$$

where $Misresponse_{ij}$ is the percentage of misresponders.

The results of the multilevel model revealed that context switching was the only significant predictor of participant misresponse. As hypothesized, participant misresponse rates were significantly higher for adjacent item pairs that required more context switching (B = .10, SE = .02, β = .37, p < .001). Contrary to my hypothesis, based on the results from the study by

Ickes et al. (2018), direction switching (B = -.01, *SE* = .02, β = -.04, *p* = .49) was not a significant predictor of participant misresponse. As found in the previous study by Ickes et al. (2018), factor switching (B = .05, *SE* = .03, β = .12, *p* =.10) was not a significant predictor of participant misresponse.

Finally, as predicted, the effects of item-length discrepancy and item-pair length were not significant as unique predictors of participant misresponse (see Table 6 for all model parameters).[10] Additional models with two-way and three-way interactions involving the three switching variables were tested and none of these interaction effects were significant (*p*s > .05).

Multilevel mediation models. In addition to examining the results of these two multilevel models, I sought to replicate the findings of the multilevel mediation models reported in Ickes et al. (2018). Specifically, I sought to determine whether context switching mediated (either partially or fully) the relationship between direction switching and factor switching with inter-item correlation.[11]

A series of 1-1-1 multilevel mediation models were estimated using the multilevel mediation (MLMed; Rockwood & Hayes, 2017) computational macro in SPSS. The 1-1-1 notation indicates that the independent, mediator, and outcome variables are all at the lowest level of measurement (i.e., at Level 1). However, because a grouping variable exists (i.e., scale), it was necessary to model this variable at a higher level of measurement (i.e., at Level 2).[12]

In each mediation model, the intercept was treated as a random coefficient (i.e., each scale had a unique intercept value) and all predictors were estimated as fixed effects; that is, they

[10] Two additional multilevel models (one for each of the two outcome variables) were tested without item-pair length and item-length discrepancy as predictors in the models, and the results were identical in terms of significance to those reported with all five predictors in the models.
[11] Although the mediation model examining the mediating effects of context switching on direction switching and participant misresponse was examined in the Ickes et al. (2018) study, this mediation model was not tested in the present study because direction switching was not significantly correlated with participant misresponse.
[12] See Krull and MacKinnon (1999, p. 258) for the 1-1-1 multilevel mediation model equations.

were estimated as constant across the Level 2 units. In the first mediation model, the predictor variable (*X*) was direction switching, the mediator (*M*) was context switching, and the outcome variable (*Y*) was inter-item correlation. The second mediation model was the same as the first, except factor switching served as the predictor variable (*X*) in this model. The final mediation model was the same as the second, except participant misresponse served as the outcome variable (*Y*) in this model. This final mediation model was not tested in the previous study by Ickes et al. (2018) because factor switching (*X*) was not a significant predictor of misresponse (*Y*) in those data.

Direction switching as predictor of inter-item correlation. As can be seen in Figure 1, items that were mismatched in item direction were associated with greater context switching, and greater context switching was associated with lower inter-item correlations. A 95% confidence interval for the indirect effect based on 10,000 Monte Carlo samples revealed a significant indirect effect of direction switching on inter-item correlation via context switching. The direct effect was not significant, suggesting that context switching fully mediates the relationship between direction switching and inter-item correlation. This finding, of apparent full mediation, helps to account for why direct switching was not a significant predictor of inter-item correlation in the multilevel mediation model that also included context switching as a predictor variable.

Factor switching as predictor of inter-item correlation. As can be seen in Figure 2, items that were mismatched on factor were associated with greater context switching, and greater context switching was associated with lower inter-item correlations. A 95% confidence interval for the indirect effect based on 10,000 Monte Carlo samples revealed a significant indirect effect of factor switching on inter-item correlation via context switching. However, the direct effect of factor switching predicting inter-item correlation remained significant in this mediational model

(B = -.20, $p < .001$), suggesting that context switching only partially mediates the relationship between factor switching and inter-item correlation. This finding replicates the result reported in Ickes et al. (2018) of partial mediation.

Factor switching as a predictor of participant misresponse. As can be seen in Figure 3, items that were mismatched on factor were associated with greater context switching, and greater context switching was associated with greater participant misresponse. A 95% confidence interval for the indirect effect based on 10,000 Monte Carlo samples revealed a significant indirect effect of factor switching on participant misresponse via context switching. The direct effect was not significant, suggesting that context switching fully mediates the relationship between factor switching and participant misresponse. Although this finding was not found in the previous study by Ickes et al. (2018), the results of this mediation model suggest that adjacent item pairs that differ in content (based on factor) lead to greater levels of context switching, which ultimately lead to higher levels of participant misresponse.

Research Objective 2

To determine whether the new predictor variable, trait aspect switching, was a significant predictor of inter-item correlation and participant misresponse, the same two multilevel models reported above were tested with this additional variable included as a sixth predictor in the models.

The following multilevel linear model was used to predict inter-item correlation:

$$z_{ij} = b_{0j} + b_1 CS_{ij} + b_2 Length_{ij} + b_3 Discrepancy_{ij} + b_4 Direction_{ij} + b_5 Factor_{ij} + b_6 Trait_{ij} + e_{ij}$$

The following multilevel linear model was used to predict participant misresponse:

$$Misresponse_{ij} = b_{0j} + b_1 CS_{ij} + b_2 Length_{ij} + b_3 Discrepancy_{ij} + b_4 Direction_{ij} + b_5 Factor_{ij} + b_6 Trait_{ij} + e_{ij}$$

Predictors of inter-item correlation. When the trait aspect switching variable was introduced as a predictor in to the model predicting inter-item correlation, both context switching (B = -.18, *SE* = .02, β = -.36, $p < .001$) and factor switching (B = -.21, *SE* = .04, β = -.29, $p < .001$) remained as significant predictors. As before, the average inter-item correlation was significantly lower for adjacent item pairs that required more context switching and for item pairs that loaded on different factors rather than the same factor.

In support of my hypothesis specific to this research objective, trait aspect switching was also a significant predictor of inter-item correlation. Specifically, the average inter-item correlation was significantly lower for adjacent item pairs that were mismatched in trait aspect than for pairs that were matched in trait aspect (B = -.06, *SE* = .02, β = -.10, $p = .007$). See Table 7 for all model parameters.

Predictors of participant misresponse. When trait aspect switching was included as a predictor in the model predicting participant misresponse, context switching (B = .10, *SE* = .02, β = .35, $p < .001$) remained a significant predictor of participant misresponse. In contrast to the findings of the simpler multilevel model without trait aspect switching as a predictor, factor switching (B = .06, *SE* = .03, β = .14, $p = .049$) was now a significant predictor of participant misresponse in this model. Participant misresponse rates were significantly higher for adjacent item pairs that required more context switching or belonged to different factors.

In support of my hypothesis specific to this research objective, trait aspect switching was a significant predictor of participant misresponse (B = .04, *SE* = .02, β = .14, $p = .008$). Participant misresponse rates were significantly higher when adjacent item pairs were mismatched on trait aspect, compared to those matched on trait aspect. None of the other

predictors in the model were significant when predicting participant misresponse (see Table 8 for all model parameters).

Overall, I hypothesized that trait aspect switching would be a significant unique predictor of both inter-item correlation and participant misresponse, such that inter-item correlation would be significantly weaker and participant misresponse significantly greater for item pairs that were mismatched on trait aspect. Both of these hypotheses were supported in the present study.

Additional multilevel mediation models. Two multilevel mediation models with trait aspect switching as the predictor variable (*X*), context switching as the mediator (*M*), and both inter-item correlation and participant misresponse as the outcome variables (*Y*) were tested. In the first mediation model predicting inter-item correlation, there was no evidence of significant mediation (Indirect effect: B = -.01, *SE* = .01, *p* = .61). In addition, in the second mediation model predicting participant misresponse, there was no evidence of significant mediation (Indirect effect: B = .004, *SE* = .01, *p* = .60). The lack of significant mediation of context switching for the two outcome measures suggest that the type of re-framing that is required to respond to items that vary in trait aspect is qualitatively different than the type of re-framing that is required to respond to items that vary in context generality/specificity.

Effects of Cognitive Re-Framing

An examination of the standardized regression coefficients (i.e., effect sizes) of the three significant forms of cognitive re-framing variables are substantial when predicting inter-item correlation (βs ranging from -.36 to -.10) and participant misresponse (βs ranging from .14 to .35) in the full multilevel models with trait aspect switching included as a predictor variable.

To demonstrate this point, with regard to the effect of *context switching*, the average inter-item correlation for item pairs with *high ratings of context switching* were significantly

weaker (r = .27) than those for *low ratings of context switching* (r = .49), $t(233.96) = 9.67$, $p <$ .001. And, also as expected, the average misresponse rates were significantly lower for item pairs with low context switching ratings (20.60%) than for item pairs with high context switching ratings (30.88%), $t(41) = -5.64$, $p < .001$.[13]

With regard to direction switching, although not a significant predictor in the multilevel model predicting inter-item correlation, the average inter-item correlations for adjacent item pairs that were *directional matches* were significantly stronger (r = .41) than *directional mismatches* (r = .34), t (241) = 2.67, p = .01. However, there was no significant difference (p = .23) in rates of participant misresponse between directional matches (25.29%) and mismatches (26.26%).

With regard to factor switching, the average inter-item correlations for adjacent item pairs that were *factor matches* were significantly stronger (r = .39) than *factor mismatches* (r = .31), $t(94.29) = 3.67$, $p < .001$. In addition, the average misresponse rates were significantly lower for item pairs that were factor matches (24.85%) than for item pairs that were factor mismatches (30.14%), $t(41) = -2.00$, p = .047.

Finally, with regard to trait aspect switching, I found that the average inter-item correlation for items that were *mismatched in trait aspect* was significantly weaker (r = .34) than *trait aspect matches* (r = .41), $t(241) = 2.38$, p = .02. Moreover, the average misresponse rates were significantly lower, as expected, for item pairs that were matched in trait aspect (23.54%) than for item pairs that were mismatched in trait aspect (28.63%), $t(41) = -2.63$, p = .01.

[13] The item-pairs were split into two groups according to whether the mean context-switching ratings were greater than or equal to the midpoint on the rating scale (i.e., 2) (high; N = 120) or below the midpoint (low; N = 123).

Research Objective 3

To determine whether there were individual differences in response consistency (M = .85, SD = .18), the relationships between response consistency and the six relevant personality traits were examined. These six personality traits included: (1) attentional control and shifting, (2) cognitive flexibility, (3) conscientiousness, (4) attentional impulsivity, (5) motor impulsivity, and (6) need for cognition. I hypothesized that individuals who are more cautious, competent, and tend to engage in effortful cognitive activities (especially activities that require attentional shifting) would be more consistent in their responding to personality scale items.

To test this hypothesis, I first examined the zero-order correlations between the participants' scores on the six personality measures and participants' within-person response consistency across the 16 individual difference measures. As expected, the results of these zero-order correlations suggested that lower levels of response consistency (i.e., higher standard deviations) were associated with the inability to shift attention between tasks (Attentional Control and Shifting; $r(298) = -.15$, $p = .008$), a lack of concentration and attention (Attentional Impulsivity; $r(298) = .25$, $p < .001$), and behavioral impulsivity (Motor Impulsivity; $r(298) = .15$, $p = .009$). On the other hand, within-person response consistency was not significantly correlated with social desirability, cognitive flexibility, conscientiousness, or need for cognition (ps > .05). See Table 9 for the zero-order correlations between the personality variables and within-person response consistency.

In order to determine whether attentional control, attentional impulsivity, or motor impulsivity were unique predictors of within person response consistency, a multiple regression model with these three personality traits were entered as potential predictors of within-person response consistency. In addition to the three personality traits, social desirability was entered as

a control variable in the regression model. This multiple regression model, therefore, examined the role of personality predicting within-person response consistency while controlling for social desirability.

The results of the multiple regression analysis indicated that attentional impulsivity was the only unique predictor of within person response consistency (B = .01, *SE* = .003, *p* = .003). Specifically, individuals who reported a lack of concentration and attention on tasks were significantly less consistent in their responding to the personality scale items. None of the other personality variables in the model significantly predicted within-person response consistency (see Table 10 for all model parameters).[14]

Supplemental Analyses

In addition to investigating the three outlined research objectives, I also examined whether the attention check items had the effect of "refocusing" participants' attention at those points in the survey where they occurred. If the attention check items have a refocusing property, then the adjacent item pairs following the attention check items should have significantly stronger inter-item correlations and lower rates of participant misresponse compared to the adjacent item pairs that appear prior to the attention check items.

Three of the four attention check items appeared in Part 3 of the online survey, which allowed for a test of refocusing on inter-item correlation and participant misresponse in three personality scales: (1) Thin-Skinned Ego-Defensiveness (TSED), (2) Self-Monitoring (SM), and (3) Locus of Control (LOC). The three attention check items appeared roughly in the middle of each of the three personality scales, which resulted in a total of 27 adjacent item pairs that

[14] In addition, I conducted an exploratory factor analysis on the six personality traits and extracted two orthogonal factors, each with three personality scales per factor. When these two factors were entered into a multiple regression model with social desirability to predict within person response consistency, neither the factors nor social desirability were significant as predictors of within person response consistency.

appeared before the attention check item and 25 adjacent item pairs after the attention check item across the three scales. To test whether the attention checks affected inter-item correlation and participant misresponse, I examined whether the inter-item correlations and misresponse rates differed for adjacent item pairs that appeared *before* the attention check item and *after* the attention check item within a given scale.

The results of an independent samples *t*-test indicated that the average inter-item correlation for adjacent item pairs that appeared after the attention check item (M = .36, SD = .19) was significantly stronger than average inter-item correlation for adjacent item pairs that appeared before the attention check item (M = .24, SD = .17), $t(50) = -2.44$, $p = .02$. However, the results of a second independent samples *t*-test indicated that participant misresponse rates were not significantly different for adjacent item pairs that appeared before (M = .30, SE = .16) or after (M = .28, SE = .13) the attention check item, $t(50) = .58$, $p = .58$. Together, these findings suggest that attention items might have the desired effect of having participants refocus and pay greater attention to the subsequent scale items within a given scale, resulting in stronger inter-item correlations.

CHAPTER 4

DISCUSSION

The present study was the first attempt at replicating the findings reported in Ickes et al. (2018), which examined the effects of three forms of cognitive re-framing (context switching, direction switching, and factor switching) on inter-item correlation and participant misresponse. In addition, the present study was the first to measure a new form of cognitive re-framing, trait aspect switching, and examine its effects on the same psychometric outcome measures. Lastly, this was the first study to examine individual differences in within-person response consistency across six potentially relevant personality scales.

Research Objective 1

With regard to the first research objective, as predicted, context switching and factor switching were significant predictors of inter-item correlation. Specifically, average inter-item correlation was significantly lower for adjacent item pairs that required more context switching and for item pairs that loaded on different factors rather than the same factor. In addition, as predicted, participant misresponse rates were significantly higher for adjacent item pairs that required more context switching, but were unaffected by factor switching. However, contrary to my hypotheses and previous research by Ickes et al. (2018), direction switching was not a significant predictor of inter-item correlation or participant misresponse.

In addition to examining the effects of the three forms of re-framing on inter-item correlation and participant misresponse, a series of multilevel mediation models were examined in the present study. Specifically, as predicted, context switching acted as a significant mediator of the relationship between direction switching and inter-item correlation, and as a partial mediator for the relationship between factor switching and inter-item correlation. Context

switching also served as a significant mediator of the relationship between factor switching and participant misresponse. Because direction switching was not significantly associated with participant misresponse in the present sample, context switching did not serve as a mediator of this relationship as it did in the study conducted by Ickes et al. (2018).

Research Objective 2

With regard to the second research objective, three forms of re-framing were significant predictors of both inter-item correlation and participant misresponse: (1) context switching, (2) factor switching, and (3) trait aspect switching. As predicted, weaker inter-item correlations and higher rates of participant misresponse were associated with adjacent item pairs that required more context switching, belonged to different factors, and were mismatched on trait aspect. Contrary to my hypothesis based on the previous findings of Ickes et al. (2018) was the finding that direction switching was not a significant predictor of inter-item correlation or participant misresponse. Overall, however, the hypothesis that trait aspect switching was a unique form of cognitive re-framing that negatively affected both inter-item correlation and participant misresponse was supported, and there was also evidence that the effect of this variable on both outcome variables was not mediated by context switching.

Research Objective 3

Finally, with regard to the third research objective, three personality predictors were significantly associated with within person response consistency. As hypothesized, individuals who have difficulty paying and shifting attention and are more impulsive responded in a more inconsistent manner across the scales in the third part of the online survey.

The fact that some researchers (e.g., Baird, Kimdy, & Lucas., 2006; Fiske & Rice, 1955) have suggested that within-person variability in responding might be its own, stable personality

characteristic may make it difficult to determine what *types* of individuals are more or less variable in their responses. Baird et al. (2006), for example, found that the estimates of within person response consistency calculated for each of the Big 5 personality traits separately were significantly correlated with each other, suggesting that within-person variability in responding is at least somewhat consistent across traits. This pattern of results was generally replicated in the present study, as exhibited by the significant positive correlations of response consistency scores across scales.[15]

Overall Findings

Overall, then, the results of the present study supported my hypothesis that context switching and factor switching are two forms of cognitive re-framing that have a significant effect on both inter-item correlation and participant misresponse. In addition, the results of the present study supported my hypothesis that an additional form of re-framing, trait aspect switching, is another unique predictor of these same outcome variables. And finally, there were significant zero-order correlations between three relevant personality predictors (attentional control and shifting, attentional impulsivity, and motor impulsivity) and the global measure of within-person response consistency—findings which suggest that certain traits relevant to one's cognitive and emotional style might contribute to the extent to which they respond in a consistent manner to personality scales.

Although some researchers are proponents of requiring respondents to engage in re-framing through the inclusion of reverse-stated items or items that vary in terms of their generality/specificity, the current line of research in this area suggests that respondent re-framing can have negative consequences in the form of weaker inter-item correlations and greater levels

[15] A total of 16 of the 21 zero-order correlations between the seven response consistency values for the seven personality traits were significant, suggesting that individuals have a general tendency to respond with a similar level of consistency (or inconsistency) across scales.

of participant misresponse. Proponents of respondent re-framing assume, for example, that the inclusion of reverse-stated items can get respondents to pay closer attention to the items, can discourage response sets such as the acquiescence response set, and can extend the scale's coverage of the content (e.g., Paulhus, 1991; Weijters & Baumgartner, 2012). However, the results of the present study suggest that there is an alternative way to get respondents to pay closer attention to the content of items and to discourage response patterns without the negative consequences of constant re-framing: *the use of attention check items*. In fact, the results of the present study suggest that including attention checks do redirect participants' attention to the item content, as evidenced by the stronger inter-item correlations for items that appear after the attention check compared to those before the attention check.

One potential way to reconcile the desire to increase respondent attention, discourage biased response patterns, and extend the scales' coverage of the construct, while also maintaining strong inter-item correlations (and thus, appropriate levels of internal consistency reliability), is to group items with similar characteristics together (e.g., reverse-stated items, broad/general items, context-specific items, behavioral items, subjective experience items) and include attention checks between these different sections of items. Doing so will still require participants to consider the construct from alternative perspectives and the opposite interpretive frame (i.e., through the inclusion of reverse-stated items) and also redirect participants' attention to the task at hand, and will likely have the positive effect of increasing inter-item correlations between scale items. This potential optimizing strategy deserves to be studied and evaluated empirically.

Comparison of Results to Ickes et al. (2018)

Apart from the lack of significant effects for direction switching, the results of the present study align well with the results reported in Ickes et al. (2018). The most consistent effects across

both studies were those of context switching predicting both inter-item correlation and participant misresponse. In both studies, the effects of context switching were large and highly significant. Together, the results of these two studies suggest that scales with adjacent items that differ in terms of their frame of reference or generality/specificity have the most robust and detrimental effects on inter-item correlation and participant misresponse. These findings are also in line with those reported by Lievens et al. (2008), who found that scale reliability was lower when scales were constructed such that half of the items had one frame of reference and the other half of the items had a different frame of reference.

In addition to context switching, the findings regarding factor switching predicting inter-item correlation were consistent across both studies and in every model that was tested. Although the effects were not as large those for as context switching, adjacent items that differ in content, and thus load on different factors, consistently result in weaker inter-item correlations.

Direction switching, on the other hand, was not a significant predictor in any model that was tested in the present study. Although inter-item correlations were significantly weaker and participant misresponse rates significantly higher for adjacent item pairs that were mismatched on direction when examined using independent samples *t*-tests, direction switching was *not* a significant unique predictor of these outcome measures in the multilevel models tested in the present study. The results of a series of zero-order correlations between the predictors of inter-item correlation and participant misresponse suggest that there is a significant amount of overlap between direction switching and both context switching ($r = .32$, $p < .005$) and factor switching ($r = -.26$, $p < .001$). [16] In fact, direction switching had the strongest correlations with the other

[16] Note that the collinearity diagnostics did not suggest that multicollinearity of predictors in the multilevel models was a potential problem. In addition, I employed a number of centering (i.e., grand-mean centering the continuous predictors and scale-mean centering the continuous predictors) and non-centering approaches, and direction switching was not a significant predictor of either outcome variable using any of these centering approaches.

forms of switching, which helps account for the finding that direction switching was not a unique predictor of inter-item correlation with the other forms of switching in the model.

When considered in terms of the mediation results, however, context switching acted as a full mediator of the relationship between direction switching and both outcome variables. Note that this mediation was not full, but instead only partial, in the previous study by Ickes et al. (2018). See Tables 11 and 12 for comparisons of regression coefficients obtained in the present study with those obtained in the study by Ickes et al. (2018).

Recommendations for Scale Developers

Based on the previous research in this area and the results of the present study, the following recommendations for personality scale developers can be made:

1. Write items that are relatively uniform in length (i.e., word count).
2. Write a small percentage of reverse-stated items.
3. Consider the extent to which participants will be required to re-frame their interpretation of item content from one item to the next, especially re-framings necessitated by adjacent items that vary in: context, direction, factor, and trait aspect. Stay within the same frame when possible.
4. Consider including attention checks to redirect participants' attention to the items and to identify individuals who are not paying sufficient attention to the task.

Strengths and Limitations

A number of strengths can be identified for the present study. First, the sample was more demographically diverse in terms of age, gender, and attained level of education compared to the study by Ickes et al. (2018). Second, the sample size in the present study was large in terms of the number of online survey respondents ($N = 300$) and the number of non-redundant adjacent

item pairs used in the reported analyses (N = 243). Third, this was the first attempt at a direct replication of the findings reported in Ickes et al. (2018) in a new sample of respondents. The replication piece of this study highlighted the importance of both context and factor switching as unique predictors of inter-item correlation and participant misresponse. Fourth, a new measure of trait aspect was developed and measured reliably by several research assistants (N = 9), which allowed for the reliable (α = .89) assessment of the trait aspect switching variable examined in this study. In addition, the present study was the first to examine the effects of trait aspect switching on psychometric outcomes. Fifth, this was the first study that attempted to examine the effects of several different personality traits on response consistency.

There are a few limitations, however, that should be identified. First, the same scales used in the study by Ickes et al. (2018) were used in the present study, which affects the generalizability of the reported findings. The same scales were used because of the nature of the direct replication study, but future studies that do not have the same self-imposed limitation should be able to test these hypotheses using additional personality scales.

Second, as mentioned by Ickes et al. (2018), we currently lack direct evidence that the effects of cognitive re-framing we have identified are a result of respondent confusion. At this point, we can hypothesize that when participants are required to switch between items that differ in their generality/specificity, direction, content, and/or trait aspect they become confused about what the item and/or scale is measuring and have difficulty responding in a consistent manner. Future studies should attempt to measure respondent confusion in a more direct manner to determine whether or not this is the process (though perhaps not the only process) underlying the effects of re-framing that we report. Third, the results that are reported here and in the previous studies in this area (Hamby et al., 2016; Ickes et al., 2018) concern reliability and not scale

validity, which means that their implications for scale validity, if any, is an important avenue for future research.

Future Directions

Researchers in this area should continue to investigate the effects of the different forms of cognitive re-framing on inter-item correlation and participant misresponse, and doing so with a new set of personality and/or attitude scales would address the generalizability issue that was previously mentioned. In addition, to examine the extent to which these forms of re-framing lead to respondent confusion, researchers could investigate either proxy measures of respondent confusion (e.g., by measuring the duration times for responding to adjacent items that differ in context, direction, factor, or trait aspect), or more direct measures of respondent confusion (e.g., asking participants to self-report the specific points within the survey when they felt confused about how to interpret an item's meaning). Investigating response times and self-reported levels of confusion among adjacent items with the same or different item characteristics would allow researchers to examine whether respondent confusion is one of the underlying processes for the reported effects.

In addition to conducting future studies using a new set of personality and/or attitude scales and attempting to more directly measure respondent confusion, researchers should seek to identify additional predictors (or potential moderators) of within-person response consistency. For example, researchers could examine whether certain demographic characteristics affect the extent to which participants respond in a consistent manner, such as age, IQ, and level of educational attainment. In addition, researchers that conduct future studies using Amazon's Mechanical Turk could investigate whether there are response consistency differences between

individuals identified as MTurk *Masters*, workers who have "demonstrated excellence across a wide range of HITs" [Human Intelligence Task; www.mturk.com], and MTurk *Non-Masters*.[17]

Finally, researchers should investigate the extent to which each of the forms of re-framing affect scale validity. One could speculate that scales that require high levels of cognitive re-framing would have negative effects on scale validity because re-framing requires careful attention and thought, and participants who have to do a lot of re-framing should be more likely to get confused about the item content and how to respond consistently, ultimately affecting the scale's validity. On the other hand, it is also possible that scales involving high levels of cognitive re-framing could expand the content domain of a particular construct by requiring participants to consider the scale from multiple perspectives and across multiple domains, thus enhancing a scale's validity. Research conducted by Lievens et al. (2008) on the effect of frame of references in scale items on criterion-related validity suggests that the former hypothesis (i.e., lower scale validity with more re-framings) is likely. In this study, Lievens et al. (2008) found that the validity of conscientiousness predicting grade point average was strongest when all of the scale items had an "at school" frame of reference compared to scales with 50% of the items having an "at school" frame of reference and the other 50% having an "at work" FOR. Future research should investigate the effects of context switching and the other forms of switching on scale validity.

Conclusion

The results of the present study provide further evidence for the detrimental effects of cognitive re-framing on the psychometric outcomes of inter-item correlation and participant misresponse, especially those re-framings necessitated by context, factor, and trait aspect switching. This study and future studies of its kind will inform psychometricians and personality

[17] MTurk *Non-Masters* were included in the present study.

psychologists about the effects that certain item characteristics and their ordering within a scale can have on relevant psychometric outcomes so that they can be considered when evaluating existing personality scales or when developing new ones of their own.

FIGURES AND TABLES

Table 1

Scale Information

Scale	Items	Scale Reliability	IIC	Reference
Social Desirability	10	.80	.29	Strahan & Gerbasi (1972)
Attentional Control and Shifting	10	.90	.47	Derryberry & Rothbart (1988)
Attentional Impulsivity	8	.82	.37	Patton, Stanford, & Barratt (1995)
Motor Impulsivity	11	.65	.16	Patton, Stanford, & Barratt (1995)
Cognitive Flexibility	13	.92	.47	Martin & Rubin (1995)
Conscientiousness	9	.90	.52	John, Donahue, & Kentle, 1991
Need for Cognition	18	.96	.57	Cacioppo, Petty, & Kao (1984)
Self-Esteem	10	.94	.64	Rosenberg (1965)
Conventional Morality	45	.91	.18	Tooke & Ickes (1988)
Thin-Skinned Ego-Defensiveness	17	.90	.35	Ickes et al. (2012)
Affect Intensity for Anger/Frustration	15	.94	.49	Ickes et al. (2012)
Need for Closure	15	.90	.38	Roets & Van Hiel (2011)
Indecisiveness	11	.92	.51	Frost & Shows (1993)
Self-Oriented Perfectionism	15	.94	.52	Hewitt & Flett (1990)
Cognitive Flexibility	12	.81	.27	Martin & Rubin (1995)
Preference for Consistency	18	.91	.36	Cialdini, Trost, & Newsom (1995)
Need to Belong	10	.89	.45	Leary et al. (2013)
Self-Concept Clarity	12	.95	.60	Campbell et al. (1996)
Self-Monitoring	13	.89	.39	Lennox & Wolfe (1984)
Conformity	11	.82	.30	Mehrabian & Stefl (1995)
Emotional Contagion	15	.90	.39	Doherty (1997)
Locus of Control	28	.88	.22	Duttweiler (1984)
Sense of Self	12	.92	.48	Flury & Ickes (2007)
Total	**338**	**.89**	**.41**	

Notes. The scale reliability values represent each scale's internal consistency reliability measured using Cronbach's alpha using the data collected for the present study. IIC represents the mean inter-item correlation for all scale items (i.e., not just for adjacent item pairs within the same scale).

Table 2

Supplemental Information for the Five Multidimensional Factors

Scale	Number of Factors	*M* Factor Loading	Number of Items	Factor Description	Sample Item
Need for Closure	3				
Factor 1		.69	7	Preference for unpredictable circumstances	I don't like situations that are uncertain.
Factor 2		.84	3	Preference for order and consistency	I find that establishing a consistent routine enables me to enjoy life more.
Factor 3		.53	5	Ease at which one makes decisions and seeks the opinions of others to make decisions	I do not usually consult many different opinions before forming my own view.
Preference for Consistency	3				
Factor 1		.62	8	Desire to be seen by others as consistent/predictable	I make an effort to appear consistent to others.
Factor 2		.64	6	Desire for others to be consistent/predictable	I want my close friends to be predictable.
Factor 3		.61	4	Self-desire to have consistent attitudes and behaviors	It is important to me that my actions are consistent with my beliefs.
Need to Belong	2				
Factor 1		.68	7	Need to belong, be accepted, and included by others	I try hard not to do things that will make other people avoid or reject me.
Factor 2		.63	3	Need to be surrounded by other people or know they are available	I need to feel that there are people I can turn to in times of need.
Self-Monitoring	2				
Factor 1		.71	7	Ability to modify self-presentation	In social situations, I have the ability to alter my behavior if I feel that something else is called for.
Factor 2		.67	6	Sensitivity to the expression of others	I am often able to read people's true emotions correctly through their eyes.
Emotional Contagion	3				
Factor 1		.63	9	Contagion for negative emotions	I get filled with sorrow when people talk about the death of their loved ones.
Factor 2		.77	3	Contagion for general positive emotions	Being with a happy person picks me up when I'm feeling down.
Factor 3		.69	3	Contagion for romance-related emotions	When I look into the eyes of the one I love, my mind is filled with thoughts of romance.

Note. M Factor Loading represents the mean factor loading for each factor.

Table 3

Descriptive Statistics for Predictors and Outcome Variables

Scale	*r* to *z*	Misresponse	Length	Discrepancy	CS	DS	FS	TS
Self-Esteem	.91 (.28)	.09 (.05)	9.94 (2.31)	2.78 (2.05)	1.59 (.51)	.56	.00	.22
Conventional Morality	.20 (.14)	.39 (.15)	11.98 (2.59)	4.55 (3.43)	2.16 (.45)	.41	.00	.45
Ego-Defensiveness	.38 (.15)	.36 (.14)	13.38 (2.32)	2.75 (2.46)	1.49 (.31)	.38	.00	.38
Affect Intensity for A/F	.63 (.26)	.11 (.09)	12.39 (2.77)	3.07 (2.23)	1.51 (.52)	.43	.00	.43
Need for Closure	.43 (.29)	.31 (.17)	12.82 (2.42)	3.50 (2.47)	2.08 (.72)	.00	.64	.29
Indecisiveness	.59 (.15)	.20 (.06)	9.65 (3.04)	3.90 (3.00)	1.44 (.40)	.70	.00	.40
Self-Oriented Perfectionism	.66 (.17)	.24 (.10)	9.71 (1.46)	2.43 (1.16)	1.43 (.32)	.29	.00	.64
Cognitive Flexibility	.27 (.09)	.17 (.11)	10.09 (1.83)	1.27 (1.27)	2.40 (.54)	.45	.00	.64
Preference for Consistency	.46 (.20)	.20 (.11)	11.74 (3.90)	4.18 (4.45)	1.68 (.38)	.06	.47	.24
Need to Belong	.58 (.21)	.25 (.12)	11.67 (2.60)	5.33 (3.16)	1.85 (.34)	.56	.44	.22
Self-Concept Clarity	.79 (.21)	.12 (.05)	15.55 (2.68)	5.64 (4.27)	1.83 (.55)	.36	.00	.36
Self-Monitoring	.41 (.20)	.20 (.06)	17.71 (1.74)	3.75 (2.70)	2.24 (.45)	.33	.67	.50
Conformity	.32 (.12)	.27 (.10)	13.20 (2.61)	4.80 (3.52)	2.14 (.42)	.70	.00	.30
Emotional Contagion	.42 (.19)	.31 (.14)	12.54 (2.89)	4.93 (4.43)	2.04 (.37)	.00	.64	.79
Locus of Control	.22 (.16)	.27 (.15)	15.09 (2.58)	3.96 (2.65)	2.31 (.43)	.70	.00	.52
Sense of Self	.58 (.19)	.17 (.09)	13.14 (2.61)	4.64 (1.69)	1.89 (.53)	.45	.00	.00

Notes. Means for the outcome and predictor variables are listed in the table with their standard deviations in parentheses. The symbol *r* to *z* represents the mean adjacent inter-item correlation (*r*) by scale transformed to *z*; Misresponse is the percent of misresponders; Length is the length (i.e., word count) of the item-pair; Discrepancy is the absolute difference in length between the item-pair; CS (context-switching) is the amount of context-switching; DS (direction-switching) is the percent of item pairs that mismatched on direction; FS (factor-switching) is the percent of item pairs that mismatched on factor; TS (trait-switching) is the percent of item pairs that mismatched on trait aspect.

Table 4

Zero-Order Correlations Between Predictors and Outcome Variables (*N* = 243 adjacent item pairs)

	IIC_z	Misresponse	Length	Discrepancy	CS	DS	FS	TS
IIC_z								
Misresponse	-.70**	-						
Length	-.11	-.02	-					
Discrepancy	-.04	.05	.18**	-				
CS	-.61**	.38**	.15*	.01	-			
DS	-.18**	.03	.06	-.05	.32**	-		
FS	-.17**	.13*	.11	-.01	.23**	-.26**	-	
TS	-.16*	.17*	.02	-.09	.08	-.02	-.05	-

Notes. IIC_z represents adjacent inter-item correlation (*r*) by scale transformed to *z*; Misresponse is the percent of misresponders; Length is the length (i.e., word count) of the item-pair; Discrepancy is the absolute difference in length between the item-pair; CS (context-switching) is the amount of context-switching; DS (direction-switching) is the percent of item pairs that mismatched on direction; FS (factor-switching) is the percent of item pairs that mismatched on factor; TS (trait-switching) is the percent of item pairs that mismatched on trait aspect.

*$p < .05$. **$p < .001$.

Table 5

Multilevel Linear Model Predicting Inter-Item Correlation Using Restricted Maximum Likelihood Estimation Method (*N* = 243 adjacent item pairs)

	B	*SE* B	*t*	*df*	95% CI
Fixed intercept	.89	.08	11.65**	100.33	[.74, 1.05]
Context switching	-.18	.02	-7.34**	232.65	[-.23, -.13]
Item-pair length	< .001	.004	.02	233.64	[-.01, .01]
Item-length discrepancy	-.004	.003	-1.36	224.90	[-.01, .002]
Direction switching	-.03	.02	-1.32	227.41	[-.08, .02]
Factor switching	-.20	.04	-5.04**	236.40	[-.28, -.12]

Variance Components	Estimate	*SE*	Wald *Z*	95% CI
Residual	.02	.002	10.54**	[.02, .03]
Intercept	.03	.01	2.56*	[.01, .06]

*$p < .05$. **$p < .001$.

Table 6

Multilevel Linear Model Predicting Participant Misresponse Using Restricted Maximum Likelihood Estimation Method (*N* = 243 adjacent item pairs)

	B	*SE* B	*t*	*df*	95% CI
Fixed intercept	.05	.05	1.04	149.51	[-.05, .15]
Context switching	.10	.02	5.54**	236.99	[.07, .14]
Item-pair length	-.002	.003	-.67	236.72	[-.01, .004]
Item-length discrepancy	.002	.002	.68	228.62	[-.003, .01]
Direction switching	-.01	.02	-.69	232.46	[-.05, .02]
Factor switching	.05	.03	1.67	229.95	[-.01, .11]

Variance Components	Estimate	*SE*	Wald *Z*	95% CI
Residual	.01	.001	10.56**	[.01, .02]
Intercept	.006	.003	2.42*	[.003, .01]

Notes. The direction switching variable was modeled as a dummy variable such that 0 = *both items were reverse-scored or neither items were reverse-scored* and 1 = *one item was reversed-scored and the other item was not reverse-scored.* B = unstandardized coefficient.

*$p < .05$. **$p < .001$.

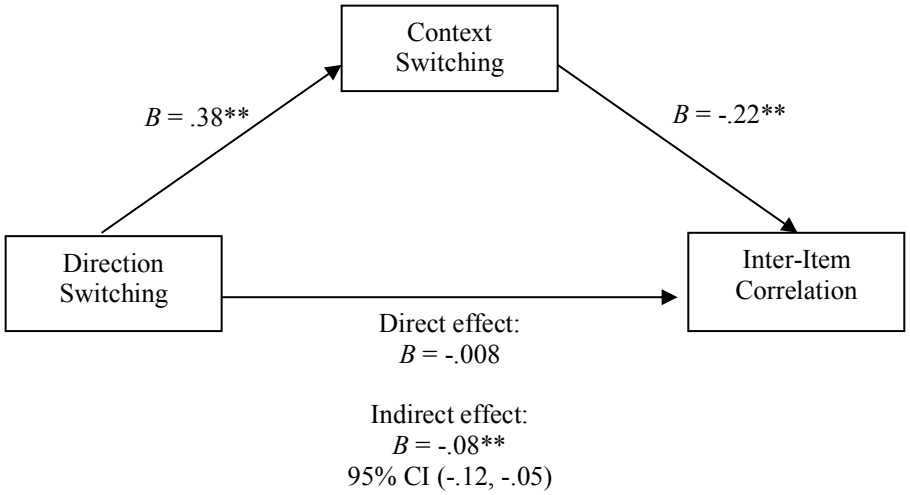


Figure 1. Effect of direction switching on inter-item correlation via context switching in a random intercept, fixed slope multilevel mediation model. B = unstandardized coefficient and 95% CI represents the 95% Monte Carlo confidence intervals based on 10,000 samples.

*$p < .05$. **$p < .001$.

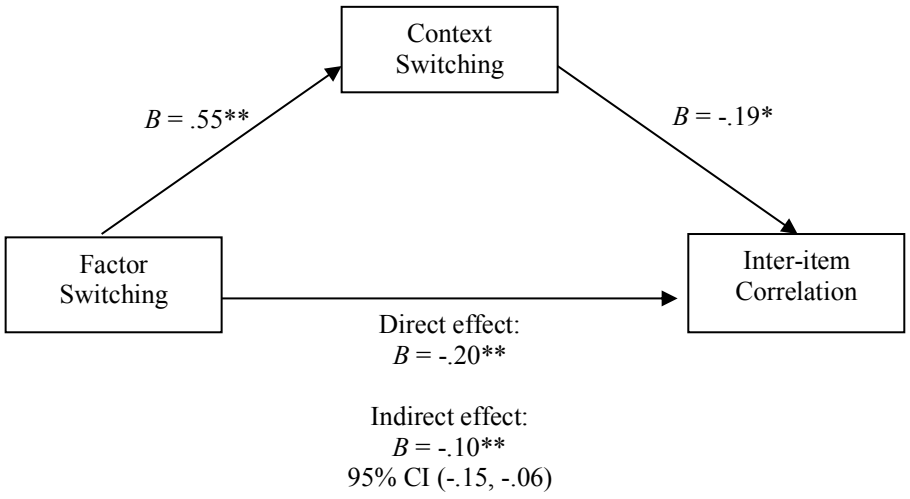


Figure 2. Effect of factor switching on inter-item correlation via context switching in a random intercept, fixed slope multilevel mediation model. *B* = unstandardized coefficient and 95% CI represents the 95% Monte Carlo confidence intervals based on 10,000 samples.

*$p < .05$. **$p < .001$.

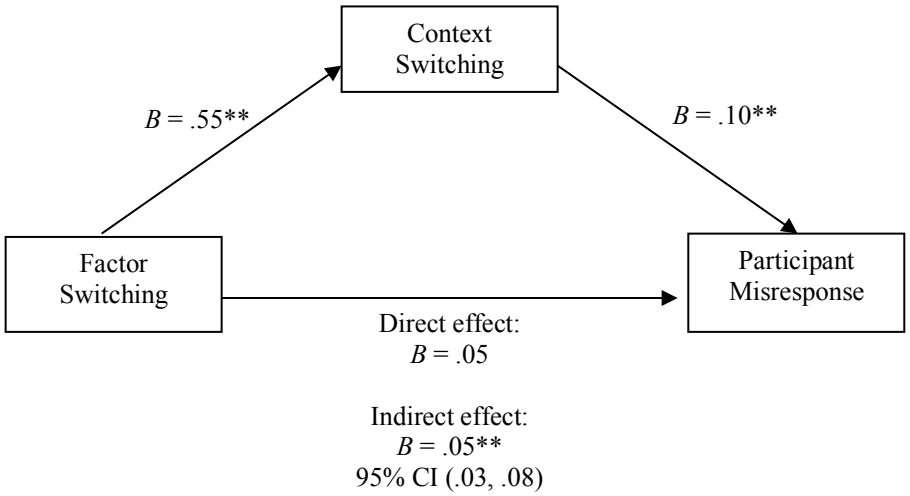


Figure 3. Effect of factor switching on participant misresponse via context switching in a random intercept, fixed slope multilevel mediation model. *B* = unstandardized coefficient and 95% CI represents the 95% Monte Carlo confidence intervals based on 10,000 samples.

*$p < .05$. **$p < .001$.

Table 7

Multilevel Linear Model Predicting Inter-Item Correlation Using Restricted Maximum Likelihood Estimation Method (*N* = 243 adjacent item pairs)

	B	*SE* B	*t*	*df*	95% CI
Fixed intercept	.91	.08	11.92**	99.78	[.75, 1.06]
Context switching	-.18	.02	-7.19**	231.46	[-.22, -.13]
Item-pair length	.001	.004	.17	232.53	[-.01, .01]
Item-length discrepancy	-.01	.003	-1.62	223.79	[-.01, .001]
Direction switching	-.04	.02	-1.48	226.21	[-.08, .01]
Factor switching	-.21	.04	-5.38**	235.39	[-.29, -.13]
Trait aspect switching	-.06	.02	-2.72*	224.49	[-.10, -.02]

Variance Components	Estimate	*SE*	Wald *Z*	95% CI
Residual	.02	.002	10.51**	[.02, .03]
Intercept	.03	.01	2.56*	[.01, .06]

*$p < .05$. **$p < .001$.

Table 8

Multilevel Linear Model Predicting Participant Misresponse Using Restricted Maximum Likelihood Estimation Method (*N* = 243 adjacent item pairs)

	B	*SE* B	*t*	*df*	95% CI
Fixed intercept	.04	.05	.87	149.76	[-.06, .15]
Context switching	.10	.02	5.37**	236.00	[.06, .13]
Item-pair length	-.002	.003	-.82	235.75	[-.01, .003]
Item-length discrepancy	.002	.002	.94	227.45	[-.003, .01]
Direction switching	-.01	.02	-.54	231.21	[-.04, .03]
Factor switching	.06	.03	1.98*	228.85	[.0002, .11]
Trait aspect switching	.04	.02	2.69*	228.71	[.01, .07]

Variance Components	Estimate	*SE*	Wald *Z*	95% CI
Residual	.01	.001	10.54**	[.01, .015]
Intercept	.006	.003	2.42*	[.003, .01]

Notes. The direction switching variable was modeled as a dummy variable such that 0 = *both items were reverse-scored or neither items were reverse-scored* and 1 = *one item was reversed-scored and the other item was not reverse-scored.* B = unstandardized coefficient.

*$p < .05$. **$p < .001$.

Table 9

Zero-Order Correlations Between Personality Variables and Within-Person Response Consistency (N = 300)

Variables	Response Consistency
Attentional Control and Shifting	-.15*
Attentional Impulsivity	.25**
Motor Impulsivity	.15*
Cognitive Flexibility	-.01
Conscientiousness	-.10
Need for Cognition	-.09
Social Desirability	-.01

*p < .05. **p < .001.

Table 10

Multiple Regression Analysis Predicting Within Person Response Consistency

Predictor Variables	B	*SE* B	*t*
Attentional Control and Shifting	-.001	.002	-.73
Attentional Impulsivity	.009	.003	3.04*
Motor Impulsivity	.002	.003	.76
Social Desirability	.001	.001	.91

*$p < .05$. **$p < .001$.

Table 11

Comparison of Multilevel Model Results from Ickes et al. (2018; N = 269 adjacent item pairs) and the Present Study with Trait Aspect Switching (N = 243 adjacent item pairs) Predicting Inter-Item Correlation

	Present Study			**Ickes et al. (2018)**		
Predictors	B	β	*t*	B	β	*t*
Context switching	-.18	-.36	-7.19**	-.15	-.42	-8.54**
Item-pair length	.001	.01	.17	.001	.02	.50
Item-length discrepancy	-.01	-.06	-1.62	.001	.02	.53
Direction switching	-.04	-.06	-1.48	-.13	-.33	-7.47**
Factor switching	-.21	-.29	-5.38**	-.11	-.22	-3.46**
Trait aspect switching	-.06	-.10	-2.72*	-	-	-

*$p < .05$. **$p < .001$.

Table 12

Comparison of Multilevel Model Results from Ickes et al. (2018; N = 269 adjacent item pairs) and the Present Study with Trait Aspect Switching (N = 243 adjacent item pairs) Predicting Participant Misresponse

	Present Study			**Ickes et al. (2018)**		
Predictors	B	β	*t*	B	β	*t*
Context switching	.10	.35	5.37**	.08	.32	4.94**
Item-pair length	-.002	-.05	-.82	.0001	.001	.03
Item-length discrepancy	.002	.05	.94	-.002	-.05	-.97
Direction switching	-.01	-.03	-.54	.07	.25	4.23**
Factor switching	.06	.14	1.98*	.04	.12	1.53
Trait aspect switching	.04	.14	2.69*	-	-	-

*$p < .05$. **$p < .001$.

References

Baird, B. M., Kimdy, L., & Lucas, R. E. (2006). On the nature of intra-individual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90,* 512-527.

Biderman, M. D. (2007, April). Variability indicators in structural equation models. Part of symposium: Examining old problems with new tools: Statistically modeling applicant faking. Paper presented at the 22nd Annual Conference of The Society for Industrial and Organizational Psychology, New York, NY.

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89,* 150-157.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

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

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology, 69*, 318-328.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R professional manual.* Odessa, FL: Psychological Assessment Resources.

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Abnormal and Social Psychology, 60*, 151-174.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349-354.

Dennis, J. P., & Vander Wal, J. S. (2010). The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34*, 241-253.

Derryberry, D., & Rothbart, M. K. (1988). Affect, arousal, and attention as components of temperament. *Journal of Personality and Social Psychology, 55,* 958-966.

Doherty, R. W. (1997). The emotional contagion scale: A measure of individual differences. *Journal of Nonverbal Behavior, 21*, 131-154.

Duttweiler, P. C. (1984). The internal control index: A newly developed measure of locus of control. *Educational and Psychological Measurement, 44*, 209-221.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and Psychology Measurement, 53*, 417-424.

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin, 52,* 217-250.

Flury, J. M., & Ickes, W. (2007). Having a weak versus strong sense of self: The Sense of Self Scale (SOSS). *Self and Identity, 6*, 281-303.

Frost, R. O., & Shows, D. L. (1993). The nature and measurement of compulsive indecisiveness. *Behaviour Research and Therapy, 31*, 683-692.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter? *Personality and Individual Differences, 35,* 1241-1254.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloniger, Cr. R., & Gough, H. C. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Hamby, T., & Ickes, W. (2015). Do the readability and average item length of personality scales affect their reliability? Some meta-analytic answers. *Journal of Individual Differences, 36*, 54-63.

Hamby, T., Ickes, W., & Babcock, M. (2016). Evidence for context switching in the effects of average item length and item-length variability on internal consistency. *Journal of Personality Assessment, 98*, 491-502.

Hamby, T., & Taylor, W. (2016). Survey satisficing inflates reliability and validity measures: An experimental comparison of college and amazon mechanical turk samples. *Educational and Psychological Measurement, 76*, 912-932.

Hewitt, P. L., & Flett, G. L. (1990). Perfectionism and depression: A multidimensional analysis. *Journal of Social Behavior and Personality, 5*, 423-438.

Ickes, W., Babcock, M. J., Hamby, T., Park, A., Robinson, R. L., & Taylor, W. (2018). Side streets and U-turns: Effects of context switching, direction switching, and factor switching on inter-item correlations and misresponse rates. *Journal of Personality Assessment*.

Ickes, W., Park, A., & Robinson, R. L. (2012). F#!%ing rudeness: Predicting the propensity to verbally abuse strangers. *Journal of Language and Social Psychology, 31*, 75-94.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory – Versions 4a and 5a.* Berkeley, CA: University of California at Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and research* (3rd ed., pp. 114-158). New York, NY: Guilford Press.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23*, 418-444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277.

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the Need to Belong Scale: Mapping the nomological network. *Journal of Personality Assessment, 95*, 610-624.

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology, 46*, 1349-1364.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268-279.

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports, 76*(2), 623–626.

Mehrabian, A., & Stefl, C. A. (1995). Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality: An International Journal, 23,* 253-263.

Motl, R. W., Conroy, D. E., & Horan, P. M. (2000). The social physique anxiety scale: An example of the potential consequence of negatively worded items in factorial validity studies. *Journal of Applied Measurement, 1*, 327-345.

Park, A., Ickes, W., & Robinson, R. L. (2014). More f#!%ing rudeness: Reliable personality predictors of verbal rudeness and other ugly confrontational behaviors. *Journal of Aggression, Conflict and Peace Research*, *6*, 26-43.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51,* 768-774.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York: Guilford.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Quigley, L., Wright, C. A., Dobson, K. S., & Sears, C. R. (2017). Measuring attentional control ability or beliefs? Evaluation of the factor structure and convergent validity of the attentional control scale. *Journal of Psychopathology and Behavioral Assessment,* 1-13.

Reddock, C. M., Biderman, M. D., Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment, 19*, 119-131.

Robinson, R. L., & Ickes, W. (2016). *Different personality traits are associated with different forms of influenceability.* Manuscript in preparation.

Rockwood, N. J., & Hayes, A. F. (2017, May). *MLmed: An SPSS macro for multilevel mediation and conditional process analysis.* Poster session presented at the meeting of the Association of Psychological Science, Boston, MA.

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences, 50*, 90-94.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education, 35*, 117-134.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research.* Hoboken, NJ: John Wiley & Sons, Inc.

Schell, K. L., & Oswald, F. L. (2013). Item grouping and item randomization in personality measurement. *Personality and Individual Differences, 55*, 317-321.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607-620.

Solomon, E., & Kopelman, R. E. (1984). Questionnaire format and scale reliability: An examination of three modes of item presentation. *Psychological Reports, 54*, 447-452.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt impulsiveness scale: An update and review. *Personality and Individual Differences, 47*, 385-395.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 28,* 191-193.

Taylor, W., & Hamby, T. (2014). *Relationships between maximizing, satisficing, and related decision-making tendencies.* Unpublished dataset, University of Texas at Arlington.

Tooke, W. S., & Ickes, W. (1988). A measure of adherence to conventional morality. *Journal of Social and Clinical Psychology, 6*, 310-334.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*, 737-747.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA, US: Sage Publications, Inc.

APPENDIX A

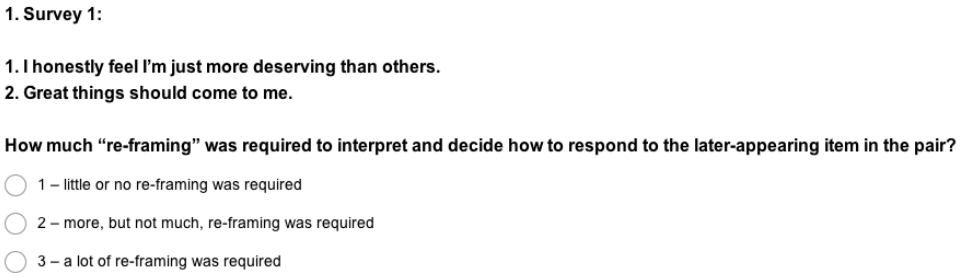
DIRECTIONS FOR CONTEXT SWITCHING RATING TASK

You have been given a set of personality scales. You will not be answering the items on any of the personality scales. Instead, your task is to rate how much "re-framing" you have to do to interpret and decide how to respond to each of the items in the personality scales you are assigned to rate. With the exception of the first item in a scale, the rating for each item should always be made with reference to the preceding item, that is, the item that appears directly *before* the new item you are rating.

Each "question" will refer you to a pair of items from the same scale. Your job is to answer the question: "How much "re-framing" was required to interpret and decide how to respond to the later-appearing item in the pair?" Your rating about the later-appearing item for that "question" should be made with reference to the first item of that pair.

It is important that you make the ratings for each item *in the exact order* that it is presented in the survey. In addition, you should complete the ratings for the entire set of personality scales *one* sitting, that is, do not take a break in the middle of the rating task.

An example "question" with a pair of items can be seen below. This example is Question 1 of Scale 1. There are two items listed as Item "1." and "2." Your job is to determine how much "re-framing" relative to Item 1 you had to do in order to interpret and decide how to respond to Item 2: "Great things should come to me."


1. Survey 1:

1. I honestly feel I'm just more deserving than others.
2. Great things should come to me.

How much "re-framing" was required to interpret and decide how to respond to the later-appearing item in the pair?

- ◯ 1 – little or no re-framing was required
- ◯ 2 – more, but not much, re-framing was required
- ◯ 3 – a lot of re-framing was required

If the new item (i.e., the later-appearing item in the item pair) requires **little or no re-framing** (applying a *new perspective* or a *new context* or a *new frame of reference*) in order to understand the item and decide how to respond to it, you should rate the item as **1** to indicate that little or no "re-framing" was required.

On the other hand, if the **new item requires more, but not much, re-framing** (applying *a new perspective*, a *new context*, or a *new frame of reference*) in order to understand the item and how to respond to it, you should rate the item as **2** to indicate that more, but not much, re-framing was required.

Or, if the **new item requires a lot of re-framing** (applying *a new perspective*, a *new context*, or a *new frame of reference*) in order to understand the item and how to respond to it, you should rate the item as **3** to indicate that a lot of re-framing was required.

These are subjective judgments, and the way you make them might differ from the way that other raters make them. What is really important, however, is that you use the same judgmental criteria *consistently* throughout the rating task. Once you have developed your own criteria for making these judgments, continue to use these criteria without any change.

APPENDIX B

SAMPLE RATING TASK FOR THE NEED FOR COGNITION SCALE

For each question, rate how much re-framing was required to interpret and decide how to respond to the later-appearing item in the pair.

1. little or no re-framing was required
2. more, but not much, re-framing was required
3. a lot of re-framing was required

1. Item 1: If other people don't seem to accept me, I don't let it bother me.
   Item 2: I try hard not to do things that will make other people avoid or reject me.

2. Item 1: I try hard not to do things that will make other people avoid or reject me.
   Item 2: I seldom worry about whether other people care about me.

3. Item 1: I seldom worry about whether other people care about me.
   Item 2: I need to feel that there are people I can turn to in times of need.

4. Item 1: I need to feel that there are people I can turn to in times of need.
   Item 2: I want other people to accept me.

5. Item 1: I want other people to accept me.
   Item 2: I do not like being alone.

6. Item 1: I do not like being alone.
   Item 2: Being apart from my friends for long periods of time does not bother me.

7. Item 1: Being apart from my friends for long periods of time does not bother me.
   Item 2: I have a strong need to belong.

8. Item 1: I have a strong need to belong.
   Item 2: It bothers me a great deal when I am not included in other people's plans.

9. Item 1: It bothers me a great deal when I am not included in other people's plans.
   Item 2: My feelings are easily hurt when I feel that others do not accept me.

APPENDIX C

DIRECTIONS FOR TRAIT ASPECT RATING TASK

Psychological constructs represent tendencies to think, feel, or act in certain ways. These constructs do not "reduce to" any particular thought, feeling, act. Instead, each is a summary of a complex set of behaviors and internal processes. One way to measure psychological constructs, such as personality traits, is through self-report. Self-report scales are those in which participants report on their own thoughts, feelings, and behaviors.

Your task is to read items from a number of self-report scales to determine whether the item is asking respondents to report on either their ***thoughts or feelings*** (how you think or how you feel) *or* ***behaviors*** (how you behave). You can ask yourself: "Do I need to consider how I think, feel, or behave to answer this item?"

- **Thoughts** = convey what we are thinking; typically precede our feelings
    - I think that I am a failure.

- **Feelings** = convey our emotional and/or physical states
    - I feel scared when my husband yells at me.

- **Behaviors =** convey how we typically behave in certain situations
    - I rarely laugh at comedies.

Let's work through some examples to familiarize you with this task:

Psychological construct: Dependence
Rating scale: 1 (strongly disagree) to 5 (strongly agree)

I …

1. Need reassurance.
2. Let myself be influenced by others.
3. Need the approval of others.
4. Need protection.
5. Often need help.
6. Show my sadness.
7. Suspect that my facial expressions give me away when I feel sad.
8. Seek support.
9. Can't do without the company of others.
10. Want to be liked.

Please do not stop ratings mid scale for any of the scales - complete your ratings for an entire scale without any breaks or pauses.