The Concept of Reliability:
Putting the “Psyche” into Psychometric

BY

NICOLE MAYER
B.A., The Ohio State University 2007

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Chicago, 2012

Chicago, Illinois

Defense Committee:

Daniel Cervone, Chair and Advisor
Mary Murphy
Tracy Caldwell, Dominican University
ACKNOWLEDGMENTS

I would like to thank my thesis committee—Daniel Cervone, Mary Murphy, and Tracy Caldwell—for their support and assistance. Their guidance from the proposal to the defense was crucial in helping me to successfully accomplish my research goals. I would especially like to acknowledge Daniel Cervone as my advisor, who oversaw the development of the proposal and whose contributions were important to the completion of this very long study.

I would also like to thank my undergraduate research assistants – Pearl Patel, Sammy Hamad, Kevin Bautista, Allison Upton, and Timir Bharucha. These individuals were extremely helpful with data collection, maintenance, and general input; I could not have done it without them.

NDM
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. The significance of reliability</td>
<td>1</td>
</tr>
<tr>
<td>B. Psychometrics</td>
<td>3</td>
</tr>
<tr>
<td>1. Internal consistency</td>
<td>3</td>
</tr>
<tr>
<td>2. Retest reliability</td>
<td>4</td>
</tr>
<tr>
<td>C. An explanation of reliability</td>
<td>6</td>
</tr>
<tr>
<td>D. Validity</td>
<td>8</td>
</tr>
<tr>
<td>E. Self-schemas</td>
<td>9</td>
</tr>
<tr>
<td>F. Knowledge and Appraisal Personality Architecture</td>
<td>13</td>
</tr>
<tr>
<td>G. Overview of study one</td>
<td>16</td>
</tr>
<tr>
<td>II. STUDY ONE: INTERNAL CONSISTENCY</td>
<td>17</td>
</tr>
<tr>
<td>A. Methods</td>
<td>17</td>
</tr>
<tr>
<td>1. Participants</td>
<td>17</td>
</tr>
<tr>
<td>2. Procedure</td>
<td>17</td>
</tr>
<tr>
<td>3. Measures</td>
<td>18</td>
</tr>
<tr>
<td>a. Schema information</td>
<td>18</td>
</tr>
<tr>
<td>b. Self-report trait measures</td>
<td>18</td>
</tr>
<tr>
<td>B. Results</td>
<td>19</td>
</tr>
<tr>
<td>1. Preliminary analyses</td>
<td>19</td>
</tr>
<tr>
<td>a. Number of participants in each group</td>
<td>20</td>
</tr>
<tr>
<td>b. Chi-square test</td>
<td>20</td>
</tr>
<tr>
<td>c. Trait means and standard deviations</td>
<td>21</td>
</tr>
<tr>
<td>2. Primary analyses</td>
<td>22</td>
</tr>
<tr>
<td>C. Discussion</td>
<td>24</td>
</tr>
<tr>
<td>III. STUDY TWO: RETEST RELIABILITY</td>
<td>26</td>
</tr>
<tr>
<td>A. Overview of study two</td>
<td>26</td>
</tr>
<tr>
<td>B. Methods</td>
<td>27</td>
</tr>
<tr>
<td>1. Participants</td>
<td>27</td>
</tr>
<tr>
<td>2. Procedure</td>
<td>28</td>
</tr>
<tr>
<td>3. Measures</td>
<td>29</td>
</tr>
<tr>
<td>a. Schema information</td>
<td>29</td>
</tr>
<tr>
<td>b. Practice questions</td>
<td>30</td>
</tr>
<tr>
<td>c. Independence questionnaire</td>
<td>30</td>
</tr>
<tr>
<td>d. Dependence questionnaire</td>
<td>31</td>
</tr>
<tr>
<td>e. Control items</td>
<td>31</td>
</tr>
<tr>
<td>f. Reaction times</td>
<td>32</td>
</tr>
<tr>
<td>g. Latency response task</td>
<td>32</td>
</tr>
<tr>
<td>h. Demographics and suspicion probes</td>
<td>33</td>
</tr>
<tr>
<td>C. Results</td>
<td>34</td>
</tr>
<tr>
<td>1. Preliminary analyses</td>
<td>34</td>
</tr>
<tr>
<td>a. Trait mean comparison</td>
<td>35</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (continued)

2. Primary analyses ........................................................................................................ 36
   a. Reaction times ......................................................................................................... 36
   b. Latency response task ............................................................................................... 39
3. Retest alpha comparison .............................................................................................. 43
D. Discussion ..................................................................................................................... 44

IV. GENERAL DISCUSSION ................................................................................................. 47
   A. Schemas ..................................................................................................................... 47
   B. Limitations ................................................................................................................ 51
   C. Future Directions ...................................................................................................... 53

APPENDICES ...................................................................................................................... 55
   Appendix A ................................................................................................................... 55
   Appendix B ................................................................................................................... 57
   Appendix C ................................................................................................................... 58
   Appendix D ................................................................................................................... 59
   Appendix E ................................................................................................................... 60

CITED LITERATURE ............................................................................................................ 61

IRB forms ........................................................................................................................... 73

VITA .................................................................................................................................... 77
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. TRAIT MEANS AND STANDARD DEVIATIONS</td>
<td>22</td>
</tr>
<tr>
<td>II. HEXACO FACET-LEVEL ALPHA COMPARISONS</td>
<td>23</td>
</tr>
<tr>
<td>III. INTER-ITEM CORRELATIONS AMONG INDEPENDENCE SCALE ITEMS</td>
<td>37</td>
</tr>
<tr>
<td>IV. INTER-ITEM CORRELATIONS AMONG DEPENDENCE SCALE ITEMS</td>
<td>37</td>
</tr>
<tr>
<td>V. DESCRIPTIVES FOR LATENCY RESPONSES TO INDEPENDENCE WORDS</td>
<td>40</td>
</tr>
<tr>
<td>VI. DESCRIPTIVES FOR LATENCY RESPONSES TO DEPENDENCE WORDS</td>
<td>41</td>
</tr>
<tr>
<td>VII. DESCRIPTIVES FOR LATENCY RESPONSES TO NEUTRAL WORDS</td>
<td>42</td>
</tr>
</tbody>
</table>
SUMMARY

Reliability typically is regarded as a property of scales. If one’s scores are (in)consistent from time 1 to time 2, the test is assumed to be (un)reliable. However, an alternative conceptualization of reliability is that it is a property of persons and their responses; test scores are 'reliable' when a person responds to test items in a manner that is consistent, which implies that a test could be reliable for one person but not for another.

Two studies were conducted to test the hypotheses that people who are schematic for a trait will provide more consistent responses both within test items (internal consistency) and across test sessions (retest reliability) than people who are aschematic for the trait. An internet study provided evidence partially supportive of the hypothesis; schematics displayed higher internal reliability than aschematics on two trait measures. In the second study, conducted in two waves in a psychology laboratory, although we obtained evidence for successfully measuring schemas, contrary to the retest hypothesis, there were no differences between schematic and aschematic alphas across the testing sessions.
I. INTRODUCTION

A. The Significance of Reliability

Is personality stable over time? Like most answers to many psychological questions (e.g., nature vs. nurture), there does not appear to be one that is fundamentally right and one that is fundamentally wrong, and there is evidence suggesting that personality is both temporally stable and dynamic. For example, McCrae and Costa (1982) suggested that by age 30 our personalities are so fixed that they can predict our personalities at age 80. In line with this, other studies have also found that personality traits (the Big Five) do tend to be stable over the lifetime (e.g., Costa & McCrae, 1997; Moss & Susman, 1980). However, studies that have examined personality over more than 20 years have found less stability (see Haan, 1981), and this finding has been echoed in several meta-analyses (e.g., Ardelt, 2000; Roberts, Walton, & Veichtbauer, 2006). What should we make of these seemingly inconsistent findings? Are people’s personalities stable through life or not, and as importantly, do people even exhibit the same traits from one day to the next? Self-report and brain activation studies (Svrakic, Przybeck, & Cloninger, 1992; Canli et al., 2004) indicate that moods may vary from day to day in response to a number of factors including stress (DeLongis, Folkman, & Lazarus, 1988), novelty- and sensation seeking behaviors (Larsen & Kasimatis, 1990), menstrual cycle in women (Rossi & Rossi, 1977), and even day of the week (Ryan, Bernstein, & Brown, 2010). Given the many factors that may affect how a person responds to a scale on any given day, coupled with the existing contention within personality stability research, one may wonder how to obtain a high reliability coefficient for a measurement tool. To address this question, this article will examine the psychological factors that contribute to different types of reliability, namely internal consistency and retest reliability.

A person could feel less outgoing on one day than on another, and this discrepancy may affect psychologists’ findings, especially when conducting a longitudinal study. If an individual’s personality and/or mood is not stable over time, and a psychologist obtains low retest reliability on a certain scale,
does that mean that the scale is not reliable, and if so, how exactly do we assess retest reliability for a measurement instrument? Do the differences in one’s answers truly reflect real differences in personality, or are they merely measurement error? These are all questions of a psychometric nature, and ones on which we may shed some light by examining the very nature of the concept of retest reliability. Perhaps the answers lie not only in what psychologists measure, but in how they measure it.

The assumptions that researchers make about reliability generally stem from Classical Test Theory (CTT; Gulliksen, 1950; Lord & Novick, 1968). CTT, an expansion and reformation of one of the tenets of the naïve model of veridicality (that one’s answer actually reflects what the instrument is attempting to measure), can be summed with the following equation:

\[ X = T_x + E_x \]

where \( X \) is a person’s observed score, \( T_x \) is the person’s true score, and \( E_x \) is the person’s error score. This theory asserts that the true score and error score are uncorrelated (Novick, 1966) so therefore, any error that a researcher observes (say, when calculating retest reliability) is purely that of measurement and not of personality (the ‘true’ score). But perhaps the true score and error score are correlated. If a person completes a trait scale, CTT assumes that the person’s true score is that which is affected by the trait being measured and nothing else. Mood would fall under error source here, and although there is evidence that traits and moods vary independently of one another, there is also evidence that certain traits are correlated with moods. For example, higher levels of trait harm avoidance were shown to correlate with such mood states as hostile, anxious, confused, and tired (Svrakic, Przybeck, & Cloninger, 1992). This sort of error variance may not affect internal consistency and would more than likely affect temporal stability of the test (McCrae et al., 2011), but it would not necessarily mean that the test is unreliable. Thus, it does not seem so easy to neatly separate true scores from error scores, and it is at this junction that CTT starts to fall apart as a basic measurement theory for personality.
According to CTT, any error that occurs when assessing personality must be a problem with the measurement tool and not with the person taking the test. Therefore, researchers who assess retest reliability may be inclined to throw out data, or claim that the measurement instrument is unreliable if they obtain low correlation coefficients due to measurement error that is too large. Of course, as mentioned previously, there are many facets of personality that are not stable over time and indeed, Kaplan and Saccuzzo (2001) assert that “This type of analysis is of value only when we measure ‘traits’ or characteristics that do not change over time” (p. 106).

Internal consistency and retest reliability are both crucial statistics, especially in scale construction and unfortunately, researchers may prematurely dismiss significant findings due to an erroneously conflated measurement error because of the emphasis placed on the purity and stability of personality. What is needed is a different conceptualization of reliability; one that rests on different assumptions that will allow researchers to uncover significant findings that may otherwise be buried in the rubble of the stable personality assumption. Internal consistency, test-retest, inter-rater, and parallel forms are all ways that researchers obtain the alpha coefficient, and in the following sections I will discuss current issues with internal consistency and retest reliability, as they are the most common forms of reliability that researchers test (Hogan, Benjamin, & Brezinsky, 2000), and subsequently propose a study that may demonstrate their new utility and importance.

B. **Psychometrics**

1. **Internal consistency**

   Internal consistency can be defined as the relationship between scores “on the basis of alternate configurations of the items across a single administration of the instrument” (Onwuegbuzie & Daniel, 2002, p. 90) or more simply, how well a set of items that purports to measure a construct are correlated. Because of the ease and efficiency with which researchers are able to calculate it, internal consistency is the most popular reliability estimate that is reported in journal articles (Hogan, Benjamin,
& Brezinski, 2000; Onwuegbuzie & Daniel, 2002). It is also considered by some to be the most important indicator of reliability (Charter, 2003; although for disagreement see Boyle, 1991). The proliferation with which internal consistency is calculated makes it a prime target to test; researchers could benefit from this new concept of reliability, and it is important to demonstrate its effects on the coefficient with which most people are familiar.

For researchers constructing tests, the general agreement is that alphas should not be below .7 (Nunnally & Bernstein, 1994) nor higher than .9 (Streiner, 2003), and these numbers are even higher for clinicians measuring individuals or high-stakes decision-makers, ranging from .70 to .95 (see Pontorotto & Ruckdeschel, 2007), with a majority of researchers indicating an acceptable level at .85 or higher. Of course, internal consistency can be affected by a number of factors including number of items, inter-item correlations (Carmines & Zeller, 1979), standard deviations, and sample size (Nunnally, 1978), and Pontorotto and Ruckdeschel (2007) set forth new guidelines for calculating internal consistency while taking the aforementioned factors into account.

Much like Pontorotto and Ruckdeschel, I am also proposing another factor to consider when calculating internal consistency. It is important to keep factors such as sample standard deviation, sample size, etc. into account, and perhaps in future studies I will control for them. However to keep this study simple in other respects, I will only adopt the .7-.9 acceptability of range for calculating internal consistency.

2. **Retest reliability**

The widely accepted operationalization of retest reliability is that it is the ability of a measurement instrument to produce consistent scores over time (Schuerger, Tait, & Tavernelli, 1982; McCrae et al., 2011). In fact, the definition of reliability (or temporal stability) is so fixed that one does not even generally define it in a methods or results section, and instead its meaning is inferred from the description of the method or the trait under study (e.g., Schuerger, Zarrella, & Hotz, 1989; Watson &
Walker, 1996). The concept of “reliable” does indeed imply consistency, even to the general population. If one’s friend has a tendency to be late to get-togethers occasionally, or sometimes cancel at the last moment, that person would be thought to be unreliable, at least in a social gathering context. Thus, I maintain that reliability requires a person’s scores to be consistent over time. This is not a new idea; however, where my concept and the general concept differ is in the emphasis placed on the mechanism underlying it, which I will discuss in the next couple of sections.

Studies that report retest reliability are more scant than those that report internal consistency (Watson, 2004; Roberts & DelVecchio, 2000), but that does not mean that retest reliability is any less important, especially since its source of variance is different and it does not have the same implications as internal consistency (Sawilowsky, 2000). One main problem with retest reliability is that it generally requires a longitudinal study design of some sort, which can be both expensive and time consuming. It is also not without its share of contention.

There is a trend for contemporary researchers to conduct retest reliability in a superficial and inadequate manner (Watson, 2004), especially in longitudinal designs. In a critique of the current usage of retest reliability, Watson (2004) noted that researchers tend to include samples that are too small, test over intervals that are too short, conclude that their reliability coefficients are satisfactory with disregard to the actual size, ignore the fact that some stability values are higher than others, and make conclusions based on one measurement. Additionally, not only is the use of retest reliability scant, but in some cases it is interchanged with other forms of reliability. For example, in a longitudinal study on the consistency of individual differences, Conley (1984) used measures of internal consistency to calculate the stability of true scores, and Guadiano (2006) substituted retest for inter-rater reliability, recommending that, “other indices of reliability can be used instead if test-retest estimates of reliability are unavailable” (p. 15). However, because different coefficients represent varying sources of error, they should not be mixed (Sawilowsky, 2000).
If our use of reliability is, to say the least, reliable, then first we all need to get on board with what it is and how we should measure it. The concept of reliability that I am proposing will hopefully do just that; one of my goals is to provide psychologists with a personality assessment method and measure that is easy to conduct and calculate, and an explanation that incorporates theory that explains reliability at an individual level.

C. An Explanation of Reliability

As it stands, reliability consists of the correlation between individuals’ scores obtained at either two separate points in time or across items, and is generally regarded as a property of the measurement instrument, not of individuals taking the test, even though multiple researchers have asserted that reliability is a characteristic of individuals’ scores (Caruso, 2000; Pedhazur & Schmelkin, 1991). This is a bit curious as the way that we conceive of reliability, a psychometric property, is that it pertains to a person’s scores (the “metric”), yet has nothing to do with the actual person (the “psyche”) taking the test. By mere semantic deconstruction, one can see that there appears to be a discrepancy between our concept of reliability and its basic etymology. To bring the concept more in line with its roots, I propose a person-level account; reliability is a property of the individual (or an individual difference), and therefore differences in individuals will result in differences in alpha coefficients across both items and separate times.

Why propose a person-level explanation of reliability? Mischel (1968) argued for the importance of the situation in personality by situation interactions. Many mistook his argument to mean that personality did not matter at all, but he nevertheless found, in a review of previous studies, that personality traits (e.g., Need for Achievement, see Kagan & Moss, 1959) were correlated at most at .30 across situations and time. This did not, and does not, bode well for the argument that personality is stable.
In light of these findings, one may question what would lead individuals to respond reliably at all. This is precisely what an idiosyncratic assessment of reliability can help to address. First, consider trait theory; according to trait theory, individuals possess mean levels of traits, aggregated across situations (Costa & McCrae, 1995). Thus, if we assign a person a mean number “3” of shyness and then try to use that number to predict behavior across situations in which that person actually would score a 1 or 5, it would prove to be a very poor predictor of behavior. Different situations have different meanings to each individual, and Mischel (1968) claimed that too much information was lost in averaging responses; to obtain much higher cross-situational consistency across individuals, one would have to conduct research using idiographic methodologies. This idea can be extended to reliability. To determine whether a scale is reliable, instead of averaging and correlating all participant responses, we should examine each individual, as some may provide more reliable responses than others. If we do find that some people answer more reliably than others as expected, then it may be due to some underlying factor(s).

Within the social cognition literature, schemas, defined as knowledge of a concept (Fiske & Taylor, 2008), have been identified as cognitive structures that drive affective responses (Fiske, 1981), affective extremity (Linville, 1982), and perceptions of stimulus attributes (Tesser, 1978). Schemas have also been shown to affect people’s judgments of person (old flame, college majors) categories (Fiske, 1982), consumer products (Sujan, 1985), and loaded political issues (Sears, Huddie, & Schaffer, 1986). If general schemas can drive peoples’ responses to related or relevant targets, then a person’s knowledge of himself (self-schema) could drive his response to some self-relevant target (a personality scale). Thus, in this paper, I will be testing the cognitive construct of self-schematicity as an underlying mechanism that leads a person’s responses to be reliable. The hypothesis is as follows: participants who are schematic for a trait will provide more reliable responses on a questionnaire that assesses that trait than
participants who are aschematic for the trait. To my knowledge, there have been no studies to date that have examined temporal stability as an individual difference.

The following sections will describe and discuss self-schemas (Markus, 1977) and the Knowledge and Appraisal Personality Architecture (KAPA; Cervone, 2004) model upon which the hypothesis is grounded. First, however, a seminal paper by Borsboom, Mellenbergh, and van Heerden (2004) that has addressed a similar problem in psychometric theory and has proposed an analogous solution will be discussed.

D. **Validity**

In their 2004 article, Borsboom and colleagues raised a similar concern about the current state of validity. Their claim was that the concept of validity has evolved from its uncomplicated original conceptualization, that a test should measure what it purports to measure (Kelley, 1927; Cattell, 1946), to much more complicated versions (construct and criterion validity) that involve placement within complex nomological networks (Crohnbach & Meehl, 1955). Their main argument is that too often, researchers conceptualize validity as being correlational; that is, a construct is valid if it is appropriately situated within a theorized nomological network. For example, if one wants to validate a scale that measures agreeableness, one would also have to show that agreeableness is positively correlated with extraversion and negatively correlated with narcissism as well as both constructs’ empirical measures (e.g., measures of bragging for narcissism). However, the authors argue that measurement instruments are valid if the target (i.e., the person taking the test) systematically exerts a causal impact on the instrument, and thus validity can be conceptualized in a much simpler way. For example, a blood pressure monitor is a valid measure because the person’s blood pressure causes the dial to move. For the monitor to be valid, one need not correlate blood pressure with heartbeats-per-minute, Body Mass Index, etc. Thus, an individual’s responses always cause the test scores (Borsboom et al., 2004). This
distinction allows one to claim that an individual’s answers caused their scores on the test, whereas a
correlational explanation of validity does not.

Although I did not manipulate the independent variables in my study, and thus cannot claim a
causal explanation (yet), Borsboom et al. (2004) provided an inspiring argument to an analogous
problem that started this project moving in that direction. For a test to be valid it must first demonstrate
reliability (although for a refutation see McCrae et al., 2011) and reliability, like validity, is
conceptualized as a correlational statistic when it should perhaps be conceptualized as causal. If we
make the claim that a test is only reliable if two scores correlate, we are not making any claims about
the direction of measurement (i.e., that a person’s traits caused his or her scores).

If an individual’s answers cause the scores on a measurement scale, and individuals are unique
in the construct that the scale attempts to measure, then it is possible for a test to be differentially
reliable for different people. This is a possibility that has not been formally addressed or tested with the
current concept of reliability, which rests on the assumption that individuals’ personalities are stable
(Knowles & Condon, 2000). What is needed is a personality theory that is sensitive to individual
differences, and that may explain and predict those who would display reliable responses and those
who would not. The KAPA model (Cervone, 2004), coupled with the concept of self-schemas (Markus,
1977) provides exactly that. In the following sections I will discuss the placement of self-schemas within
the KAPA model as the theoretical framework that provide a sound base for my proposition.

E. **Self-Schemas**

When individuals complete personality questionnaires, we assume that their answers are not
drawn from a void, and we like to think that they are thinking about and assessing the extent to which
each item describes them. Thus, participants may draw from a well of self-knowledge, and use that
knowledge to answer the items.
Self-schemas (Markus, 1977) are believed to be somewhat stable cognitive structures that are comprised of self-knowledge about the self in particular domains (i.e., abilities, competencies, traits). One then draws on this information to make decisions about behaviors that are pertinent to future related experiences and thus, schemas may account for behavioral consistency. Indeed, in her landmark study, Markus (1977) found that participants who were schematic for the independence/dependence trait were more likely to endorse trait-related words and supplied more trait-consistent descriptions than those who were aschematic for the trait. This is relatively unsurprising; however schematic participants also had faster reaction times to trait-relevant words than aschematic participants. The automaticity of response on this measure provided stronger evidence for the self-schema as an important cognitive structure.

Further research on self-schemas has revealed that they facilitate encoding, evaluation, and retrieval of domain-relevant information (Bargh, 1982; Markus, Crane, Bernstein, & Siladi, 1982). Additionally, people who possess schemas in a certain domain are sensitive to information and cues relevant to that domain, and this knowledge is readily accessible (they are quick to use it) in domain-pertinent situations (Anderson & Ross, 1984; Markus, Cross, & Wurf, 1990; Kuhl, 1985). In contrast, people who are aschematic, or who do not see a certain trait as important to their self-concept, in relation to schematics, have been found to respond slower to trait-relevant words (Markus, 1977; Bem, 1981), recall less trait-relevant adjectives (Bruch, Kaflowitz, & Berger, 1988), report less physiological arousal in trait-relevant situations (Cyranowski & Andersen, 1998), and report less emotion to schema-relevant stimuli (Jung & Lennon, 2003).

Self-schemas have been found to be a relatively consistent predictor of responses and behaviors across varying contexts. For example, self-schemas have been found to predict prosocial behavior (Froming et al., 1998), dieting behaviors (Kendzierski & Whitaker, 1997), and processing of political information (Duncan, 2005) and trait-relevant information (Fekken & Holden, 1992). Schemas also play a
large role in clinical therapy; changing depressive patients’ self-schemas is thought to lead to behavioral
change, and thus is targeted in therapeutic interventions (Beck, 1967; Dozois et al., 2009). Shadel,
Niaura, & Abrams (2000), using the measures that Markus (1977) devised and an idiographic approach,
examined smoker self-schemas (knowledge about the self as a smoker). They found that the dynamic
nature of schemas was likely to predict assessment of smoking behaviors across a variety of situations.
Nasby (1989) found that participants high in private self-consciousness and thus hypothesized to
possess a self-consciousness schema, scored more consistently across testing sessions than those with
low private self-consciousness. Although these results are compelling and preliminarily support my
claim, Nasby reported that reliability was driven by participants’ self-schemas, but he only administered
the Self-Consciousness Scale (SCS; Fenigstein et al., 1975) and so we cannot be sure that the reliability
observed was due to an actual cognitive structure and not a trait difference. The current study will
address this discrepancy by measuring self-schemas in the same manner as Markus (1977).

Of course, it is also important to distinguish what a schema is not in addition to what it is. First, a
trait may seem like a similar or even the same construct as a schema, but there are some conceptual,
theoretical, and semantic differences that separate them. Traits are generally regarded as “the extent to
which an individual is high, moderate, or low on a personality dimension” (Burke, Kraut, & Dworkin,
1984, p. 568) and can further be defined in terms of behavior as the tendency to act consistently with
that level on that personality dimension (Magnussen & Endler, 1977). In contrast, schemas are not
dependent on one-to-one mappings of personality dimensions to behavior; instead they are dependent
on the personal importance, relevance, and salience of the trait dimension to individuals (Burke, Kraut,
& Dworkin, 1984). One who scores highly on a measure of extraversion would be considered to be
extremely outgoing, but that does not mean that extraversion is important or central to that person’s
self-concept. Traits are also structurally different from schemas, although it is questionable as to
whether traits are structures at all. Rather, they are broad categories, mainly used for descriptive
purposes, that originated from factor analyses (Costa & McCrae, 1995; Digman, 1990) and are applied to entire populations. Schemas, on the other hand, are stable cognitive structures that are unique to each individual (Fiske, 1981).

Second, while I am concerned with how important and central certain traits are to the participants in these studies, I am ultimately measuring schemas and not trait centrality (Asch, 1946). Trait centrality, much like a semantic network, pertains to how connected a certain trait is to other traits (Asch, 1946; Orehek, Dechesne, Fishbach, Kruglanski, & Chun, 2010). The importance of the trait in this instance is not a subjective evaluation from a person (“being independent is an important part of who I am”), but a reported network of connectivity, for example, measured by the number of trait adjectives that a person checks in relation to a target trait (Wishner, 1960).

Last, self-schemas are theoretically and empirically different from metatraits (Baumeister & Tice, 1988), which refer to the amount of trait that a person possesses. One could see how being highly traited might be confused with having a schema for a particular trait but, as with traits, metatraits do not take into account subjective judgments of importance or centrality to the individual. Most studies that measure metatraits do so by comparing individuals’ standard deviations on a certain trait (e.g., Hershberger, Plomin, & Pedersen, 1995; Baumeister, 1991) or by using a single self-report measure of trait relevance (e.g., Britt, 1993, Paunonen, 1988; Ben & Allen, 1974). The latter method appears to be the closest to the method used to measure schemas as it involves a value judgment, however, relevance is not interchangeable with importance and centrality. For example, a man may report that yes, the trait of spontaneity was relevant when he decided to go bungee jumping on a whim, but he may respond differently when asked whether being spontaneous is important and central to who he is.

Because of their personal relevance, centrality, and importance, self-schemas should differ between individuals to produce varying levels of reliability on trait measures. Specifically, participants who possess a self-schema for a certain trait will be more likely to provide consistent responses across
different testing sessions and items than those who do not possess schemas for that trait. Therefore, schematic individuals’ responses will be more reliable than those who are aschematic.

While schemas have been shown to predict behavior, what is the causal role, if any, that they play in personality functioning? What is needed is a broader but more finely-tuned theory of personality structure, one that incorporates schemas into its framework to ultimately explain consistency in responses to personality questionnaires. The KAPA model (Cervone, 2004) incorporates schemas into its personality architecture concept, and thus provides a relevant and well-suited theoretical base for my proposition.

F. **Knowledge and Appraisal Personality Architecture**

Cervone (2004) introduced the KAPA model as a theoretical framework to organize personality and explain cross-situational variance and consistency within individuals. Using an idiographic approach, the model is person-centered and examines intra-individual personality structures as opposed to inter-individual structures (see McCrae & Costa, 1996; Goldberg, 1993). The KAPA model is comprised of two components, knowledge and appraisals, that interact when individuals attempt to construct meaning in different situations. Knowledge refers to a structural personality feature that can be either generalized or domain-specific, and consists of beliefs and mental representations of ourselves, others, and the environment (Lazarus, 1991; Cervone, 2004). Appraisals refer to the dynamic evaluative judgments that people make about personally relevant occurrences (Lazarus, 1991), and possess an affective component (as opposed to the “cold cognition” of knowledge), such that emotions arise from the meanings that individuals give to situations (Cervone, 2004). Thus, individuals appraise certain situations, or decide upon courses of action or responses, by drawing from their wells of knowledge.

The KAPA model specifically identifies self-schemas as accessible self-knowledge that may be activated in relevant situations and as important contributors to personality coherence (Cervone, 2004; Cervone et al., 2008). For example, if a woman who has a “socially shy” self-schema went to a party, she
may become withdrawn (or more social as she tries to overcome it) because her shyness, activated by the situation, may be likely to dictate her thoughts/actions. Conversely, if this same woman went to the pound to adopt a dog, she may be shy while interacting with an employee, but her schema of shyness would probably not inform her decision because the decision would be irrelevant. It is important to note that the situation-specific criteria for activating schemas means that idiographic measurements are important; another person who also has a social shyness schema may become shy in completely different contexts.

Schematic information, when interacting with situational beliefs, is likely to drive consistency in responses (Cervone, 1997; Shadel & Cervone, 2006). Although I did not manipulate or measure situations/context in the current study, according to the KAPA model and previous work on schemas, it is the case that self-schemas inform our appraisals, and thus may drive consistency in responses that require individuals to draw on schematic information. Thus, in the present study, I predict that people who possess schematic information about themselves on a number of traits will be more likely to provide consistent responses between a set of items (internal consistency) and at two different times (test-retest) than those who are aschematic for this trait.

While the current concept of reliability is based on the same principles of correlation as the nomothetic approach, my hypothesis that reliability is a property of individuals calls for the sensitivity to individual idiosyncracies that that the KAPA model (Cervone, 2004) endorses. That is, based on social-cognitive theories, if one accepts the idea of personality as the interaction of various psychological processes such as affect and cognition, with perceptions of capabilities in certain environments (Mischel & Shoda, 1995; Cervone & Shoda, 1999; Bandura, 1997), one cannot accept the current concept of reliability, which takes into account none of the aforementioned properties of personality functioning. Instead, the current view of reliability is simply that if an individual’s scores are consistent from one point in time to another, then the test is reliable, and if an individual’s scores are not consistent, then
the test itself is not reliable. This correlational approach violates a basic assumption of personality assessment such that it disregards how an individual’s personality might affect his or her scores. By assessing self-schemas for the trait being measured, I am ostensibly tapping a personal relevance construct that will vary between individuals and thus produce varying levels of reliability.

Reconceptualizing reliability in this manner could have important implications for researchers in different fields of psychology; defining and measuring reliability as an individual difference could result in social, personality, and even clinical researchers conducting studies a bit differently than they otherwise do. For example, there may be multiple scales that are reliable for people who are schematic for the attribute being measured, and not so reliable for people who are not schematic for that attribute. Depending on what is being measured, researchers may want to administer different scales to people with and without relevant schemas, find a different measure that is reliable for both groups, or simply measure the attribute a completely different way (i.e., implicitly). However, schema measurement may also affect reliability across different methodologies, such as the Implicit Association Test (IAT; Greenwald et al., 1998). Retest reliability for implicit measures has generally been found to be quite low, ranging from .16 to .69 (for a review, see Egloff, Schwerdtfeger, & Schmukle, 2005) and perhaps one reason for this could be the exclusion of a schema measurement. Participants who take the Implicit Association Test – Anxiety (IAT – Anxiety; Egloff & Schmukle, 2002) but do not necessarily possess a schema for anxiety (i.e., they do not see themselves as particularly anxious or calm) may vary more in their results across testing sessions. Measuring schemas when administering the IAT may help to unearth the root of this discrepancy. Of course it is important to note that schemacity may not be the only factor that influences retest reliability; other factors such as age (Schuerger et al., 1982), the use of single adjective versus statement items (Watson, 2004), and the observability and evaluative neutrality of a trait (John & Robins, 1993) may also affect retest reliability. It is interesting to think about
how past, and subsequently current research results would be interpreted if the authors had conceived of reliability as a measure of individual differences and not measurement instruments.

G. **Overview of Study One**

The purpose of study 1 was to examine internal consistency among people who vary in schematicity for a set of traits. Specifically, study 1 was designed to test the hypothesis that people who are schematic for a trait will display higher psychometric reliability than people who are aschematic for the trait. The hypothesis was tested using the six traits (honesty/humility, emotionality, extraversion, agreeableness, conscientiousness, openness to experience) of the HEXACO model (Ashton & Lee, 2009), a model of individual differences in global dispositional traits that was created through factor analysis of everyday personality descriptors.

This study was conducted via Mechanical Turk (mturk.com), an internet service through which work activities are outsourced to internet users in return for payment. Participants were given a list of the six aforementioned traits and instructed to rank-order them in terms of importance. Participants who ranked a trait as important were considered to be schematic for that trait and participants who ranked a trait as unimportant were considered to be aschematic. Participants could choose any traits as being most and least self-descriptive. We hoped to identify, within the pool of six traits, trait variables for which the numbers of people schematic and aschematic for the trait were in balance, that is, for which a substantial and roughly equivalent number of people were schematic and aschematic. We also searched for traits for which aschematic and aschematic participants did not differ substantially on mean levels of the trait. These attributes would be ones for which a test of the main hypothesis would be conducted with sufficient statistical power and without a confound of trait level with psychometric reliability. Thus, provided that approximately equal numbers of participants indicate being schematic and aschematic for each HEXACO trait, and that they do not differ on mean levels of the traits, schematics should provide higher alpha coefficients than aschematics for each of the HEXACO traits.
II. STUDY ONE: INTERNAL CONSISTENCY

A. **Methods**

1. **Participants**

   Three hundred seventy-five participants, all over 18, received $.20 for taking part in the study. After removing responses from participants who did not complete the survey, provided the same response for every item, and indicated English as a second language, the final sample consisted of three hundred twenty participants ($M_{age} = 33.85$ years, $SD_{age} = 13.42$). The racial breakdown of the sample was 1.27% American Indian/Alaska Native, 8.23% Asian/Asian-American/Pacific Islander, 6.97% African-American/Black, 80.38% Caucasian/White, 4.43% Hispanic/Latino, and 1.27% identified as Other. Gender information was not collected for this study.

2. **Procedure**

   The study was programmed using qualtrics, an online platform for distributing surveys. Individual screens of information presented a given measure and included a “continue” button that participants used to advance to the next screen of the survey.

   After deciding to take part in the study, participants viewed and agreed to a consent form. On the next screen they completed the measure of schematicity/aschematicity; as in prior work (Burke, Kraut, & Dworkin, 1984), we reasoned that self-reports of the importance of certain psychological attributes would indirectly measure whether or not participants possessed a schema for those attributes. Thus, participants rank ordered the importance of the six HEXACO (Ashton & Lee, 2009) traits (honesty/humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience). For purposes of the present analyses, the two traits ranked as most important to the self for a given participant (see details below) were categorized as schematic traits for that participant, and the two traits ranked as least important were categorized as aschematic traits.
Next, participants completed the self-report 60-item short version of the HEXACO-60 (Ashton & Lee, 2009). Finally participants completed demographic items: age, race/ethnicity, and English as a first language.

3. **Measures**

   a. **Schema information**

      The rank-ordering task used to measure schematicity began with the following instructions, “We are interested in which of these traits are central, important, and meaningful to who you are. In other words, which traits do you give a lot of thought to on a day-to-day basis?” Participants were then instructed to “…rank the following traits from 1 to 6, where 1 is the trait that you never think about or is irrelevant to who you are/an unimportant aspect of your personality and 6 is the trait that you think about the most/is the most important aspect of your personality.” Below these instructions was a 6x6 matrix in which the rows were labeled with one of the six HEXACO traits (Honesty/Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, Openness to Experience) and each column was labeled with a number (1 – 6). Participants clicked in one box below the numbers 1-6 for each trait to indicate the importance of that trait; those who indicated a “1” or “2” (trait is unimportant) for any given trait would be considered to be aschematic for that trait and participants who indicated a “5” or “6” (trait is important) for a certain trait would be schematic for that trait.

   b. **Self-report trait measures**

      On the next screen, participants completed the 60-item HEXACO, with items randomized between participants. Participants were instructed to indicate their agreement/disagreement with each statement.

      The HEXACO-60 (see Appendix A) is a 60-item scale that includes 6 scales, each of which measures one of the HEXACO traits (e.g., H: I wouldn’t use flattery to get a raise or promotion at work,}
even if I thought it would succeed; E: I would feel afraid if I had to travel in bad weather conditions; X: On most days, I feel cheerful and optimistic; A: I rarely hold a grudge, even against people who have badly wronged me; C: I plan ahead and organize things, to avoid scrambling at the last minute; O: I’m interested in learning about the history and politics of other countries). Each trait measure of the HEXACO also contains 4 subscales that ostensibly measure facets of that trait embedded within it (H: Sincerity, Fairness, Greed Avoidance, Modesty; E: Fearfulness, Anxiety, Dependence, Sentimentality; X: Social Self-Esteem, Social Boldness, Sociability, Liveliness; A: Forgivingness, Gentleness, Flexibility, Patience; C: Organization, Diligence, Perfectionism, Prudence; O: Aesthetic Appreciation, Inquisitiveness, Creativity, Unconventionality). The response items are coded from 1 – 5 (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree).

Items are coded (See Appendix A for reverse codes) such that higher scores indicate higher levels of the trait; for example, a person who scores a mean of five on the conscientiousness scale would be considered to be highly conscientious and a person who scores a one would not be considered to be conscientious.

Internal consistency (alpha) has been found to be in the acceptable range (.7 - .9) for each subscale of the HEXACO-60 (Ashton & Lee, 2009; H: α = .79; E: α = .78; X: α = .80; A: α = .77; C: α = .78; O: α = .77).

A. Results

1. Preliminary analysis

Prior to testing the main hypothesis, we conducted a preliminary analysis by examining the numbers of participants who were schematic for each trait as well as comparing schematic and aschematic trait means. Although the study was set up so that participants were allowed to choose any of the traits as most and least important, in theory allowing every person to choose, for example, agreeableness as the least important and openness as the most important trait, we did not anticipate
substantial differences between numbers of participants who were schematic and aschematic for each trait. An ideal test of the main hypothesis would be one for which the numbers of schematics and aschematics for each trait should be balanced. Additionally, if there is a difference in trait means between schematics and aschematics, then this measurement confound renders interpretation of the main analysis difficult.

a. **Number of participants in each group**

For a number of the HEXACO traits, the numbers of participants who were schematic versus aschematic was quite uneven. Specifically, there were more honesty/humility schematics than aschematics ($n_{schem} = 161, n_{aschem} = 36$), less emotionality schematics than aschematics ($n_{schem} = 68, n_{aschem} = 97$), less extraversion schematics than aschematics ($n_{schem} = 37, n_{aschem} = 182$), more conscientious schematics than aschematics ($n_{schem} = 128, n_{aschem} = 40$), and more openness schematics than aschematics ($n_{schem} = 105, n_{aschem} = 68$). The two groups appeared to be balanced on agreeableness ($n_{schem} = 70, n_{aschem} = 62$).

b. **Chi-square test**

Chi-square tests of chance were conducted for each trait to examine whether it was equally probable that participants indicated being schematic or aschematic for the trait. The analysis showed that for four of the six HEXACO traits, the numbers of schematics and aschematics was significantly out of balance, that is, one was more common than the other; there were more honesty/humility schematics, $\chi^2(1, N = 197) = 82.79, p < .001$, more extraversion aschematics, $\chi^2(1, N = 219) = 98.67, p < .001$, more conscientious schematics, $\chi^2(1, N = 168) = 47.87, p < .001$, and more openness schematics, $\chi^2(1, N = 173) = 8.69, p < .01$. There was no difference between the numbers of emotionality schematics and aschematics, $\chi^2(1, N = 165) = 3.61, p = ns$, and agreeableness schematics and aschematics, $\chi^2(1, N = 132) = .35, ns$.  

20
c. Trait means and standard deviations

A second consideration that is pertinent to the main hypothesis test is whether schematics and aschematics differ in their trait means. As discussed previously, schemas are structurally and qualitatively distinct from traits. There is no theoretical reason why schematics and aschematics would differ on their mean levels of each trait. Schemas are partially social constructions; a person’s surroundings determine, in part, what becomes encoded as a schema. For example, a person who is extremely athletic (the trait) may develop a schema for athleticism if this was a characteristic that was highly valued among his peers, family, teachers, etc. This same athletic person may not develop a schema for athleticism at all if he was raised on a farm, was home-schooled etc. In both instances, the level of trait remains the same, but the schema varies depending on the person’s environment. Likewise, there was no reason that the national sample of participants who completed the HEXACO should also vary systematically on their levels of traits. If schematics and aschematics significantly differ on trait levels, then the comparison of alpha coefficients will be confounded with the systematic variation of the trait measures.

To examine these differences, a one-way analysis of variance (ANOVA) was conducted on all trait measures (see Table 1 for means, standard deviations, and effect sizes). Levene’s test for homogeneity of variance indicated that there were no significant differences between schematic and aschematic variance on any of the trait measures. Participants who were aschematic for a given trait displayed significantly higher means (recall that this means they were higher on the trait) than participants who were schematic for all traits except Agreeableness, for which there was no significant difference. Specifically, aschematics indicated that they were more honest, more emotional, more extraverted, more conscientious, and more open to experience.
Table 1
Means and standard deviations of participants by trait in each condition

<table>
<thead>
<tr>
<th></th>
<th>Aschematics</th>
<th>Schematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>HEXACO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>2.92</td>
<td>.57</td>
</tr>
<tr>
<td>E</td>
<td>3.15</td>
<td>.60</td>
</tr>
<tr>
<td>X</td>
<td>3.09</td>
<td>.74</td>
</tr>
<tr>
<td>A</td>
<td>2.81</td>
<td>.54</td>
</tr>
<tr>
<td>C</td>
<td>2.64</td>
<td>.64</td>
</tr>
<tr>
<td>O</td>
<td>2.49</td>
<td>.61</td>
</tr>
</tbody>
</table>

Note. P-values correspond to the difference between the group means for each trait (row).

Agreeableness was the only trait for which equal numbers of participants were schematic and aschematic, and for which the groups’ trait levels did not differ. Therefore, in the subsequent analysis, data will be reported for all six HEXACO attributes, but agreeableness will be the focal variable, as it was the only one that was not psychometrically confounded.

2. **Primary Analyses**

Alphas were calculated separately for schematics and aschematics for each trait, and were compared using the Feldt test, a technique for determining the significance of the difference between two alpha coefficients on a single test for two independent groups (Feldt, 1969; Hakstian & Whalen, 1976). The results presented below are organized by trait. Alpha values for the trait scales as well as the subscales are reported in Table 2, but caution must be exercised in interpretation of the subscale values, as each subscale consists of only 2 to 3 items.

For agreeableness, the alpha for aschematics (α = .666, n = 62) was significantly lower than the alpha for schematics (α = .777, n = 70), F(69, 61) = 1.50, p = .05. This result confirmed the hypothesis; for the groups that were balanced in number and did not differ in their trait means, as predicted, schematics were more consistent in their responses than aschematics.
Table 2
*Alpha comparison for facet-level trait measures of the HEXACO*

<table>
<thead>
<tr>
<th></th>
<th>Aschematic</th>
<th>Schematic</th>
<th>$F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Honesty/Humility</strong></td>
<td>.626</td>
<td>.725</td>
<td>1.36</td>
<td>.14</td>
</tr>
<tr>
<td>Sincerity</td>
<td>.571</td>
<td>.651</td>
<td>1.23</td>
<td>.24</td>
</tr>
<tr>
<td>Fairness</td>
<td>.831</td>
<td>.735</td>
<td>1.57</td>
<td>.03</td>
</tr>
<tr>
<td>Greed Avoidance</td>
<td>-.488</td>
<td>.425</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Modesty</td>
<td>-.140</td>
<td>.590</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Emotionality</strong></td>
<td>.732</td>
<td>.863</td>
<td>1.96</td>
<td>.00</td>
</tr>
<tr>
<td>Fearfulness</td>
<td>.560</td>
<td>.661</td>
<td>1.30</td>
<td>.12</td>
</tr>
<tr>
<td>Anxiety</td>
<td>.658</td>
<td>.767</td>
<td>1.47</td>
<td>.04</td>
</tr>
<tr>
<td>Dependence</td>
<td>.546</td>
<td>.695</td>
<td>1.49</td>
<td>.03</td>
</tr>
<tr>
<td>Sentimentality</td>
<td>.660</td>
<td>.631</td>
<td>1.09</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Extraversion</strong></td>
<td>.863</td>
<td>.856</td>
<td>1.18</td>
<td>.28</td>
</tr>
<tr>
<td>Social Self-Esteem</td>
<td>.753</td>
<td>.764</td>
<td>1.05</td>
<td>.40</td>
</tr>
<tr>
<td>Social Boldness</td>
<td>.730</td>
<td>.587</td>
<td>1.53</td>
<td>.07</td>
</tr>
<tr>
<td>Sociability</td>
<td>.662</td>
<td>.569</td>
<td>1.28</td>
<td>.19</td>
</tr>
<tr>
<td>Liveliness</td>
<td>.660</td>
<td>.617</td>
<td>1.13</td>
<td>.34</td>
</tr>
<tr>
<td><strong>Agreeableness</strong></td>
<td>.666</td>
<td>.777</td>
<td>1.50</td>
<td>.05</td>
</tr>
<tr>
<td>Forgiveness</td>
<td>.718</td>
<td>.619</td>
<td>1.35</td>
<td>.11</td>
</tr>
<tr>
<td>Gentleness</td>
<td>.714</td>
<td>.654</td>
<td>1.21</td>
<td>.22</td>
</tr>
<tr>
<td>Flexibility</td>
<td>.446</td>
<td>.498</td>
<td>1.10</td>
<td>.35</td>
</tr>
<tr>
<td>Patience</td>
<td>.590</td>
<td>.705</td>
<td>1.39</td>
<td>.09</td>
</tr>
<tr>
<td><strong>Conscientiousness</strong></td>
<td>.780</td>
<td>.813</td>
<td>1.18</td>
<td>.28</td>
</tr>
<tr>
<td>Organization</td>
<td>.542</td>
<td>.544</td>
<td>1.00</td>
<td>.52</td>
</tr>
<tr>
<td>Diligence</td>
<td>.211</td>
<td>.540</td>
<td>1.72</td>
<td>.03</td>
</tr>
<tr>
<td>Perfectionism</td>
<td>-.030</td>
<td>.622</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Prudence</td>
<td>.707</td>
<td>.544</td>
<td>1.56</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Openness to Experience</strong></td>
<td>.740</td>
<td>.756</td>
<td>1.06</td>
<td>.40</td>
</tr>
<tr>
<td>Aesthetic Appreciation</td>
<td>.540</td>
<td>.479</td>
<td>1.13</td>
<td>.28</td>
</tr>
<tr>
<td>Inquisitiveness</td>
<td>.511</td>
<td>.407</td>
<td>1.21</td>
<td>.19</td>
</tr>
<tr>
<td>Creativity</td>
<td>.752</td>
<td>.692</td>
<td>1.24</td>
<td>.16</td>
</tr>
<tr>
<td>Unconventionality</td>
<td>.486</td>
<td>.498</td>
<td>1.02</td>
<td>.47</td>
</tr>
</tbody>
</table>

*Note. $F$-value column contains † for uninterpretable negative alphas.*

* $p < .05$

** $p < .01$

For emotionality, the aschematic ($\alpha = .732$, $n = 97$) alpha was significantly lower than the schematic alpha ($\alpha = .863$, $n = 68$), $F(67, 96) = 1.96, p < .01$. This result also confirmed the hypothesis, and while the two groups were equally balanced, caution must be exercised in interpreting these results, as their means were significantly different.
For honesty/humility, the alpha for aschematics ($\alpha = .626$, $n = 36$) was lower than the alpha for schematics ($\alpha = .725$, $n = 161$) but this difference was not significant, $F(160, 35) = 1.36, ns$.

For extraversion, the aschematic alpha ($\alpha = .863$, $n = 182$) was slightly higher than the schematic alpha ($\alpha = .856$, $n = 37$); however, this difference was not significant, $F(181, 36) = 1.18, ns$.

For conscientiousness, the aschematic alpha ($\alpha = .780$, $n = 40$) was lower than the schematic alpha ($\alpha = .813$, $n = 128$); however this difference was not significant, $F(127, 39) = 1.18, ns$.

For openness to experience, the aschematic alpha ($\alpha = .740$, $n = 68$) was lower than the schematic alpha ($\alpha = .756$, $n = 105$); however, this difference was also not significant, $F(104, 67) = 1.06, ns$.

C. Discussion

Even though two major challenges (unequal numbers of schematics and aschematics and differences in trait levels between groups) rendered data analysis uninterpretable for most of the traits, due to the statistical power afforded by large numbers of aschematic and schematic people, and without the confound of differing means, we were able to test the main hypothesis on the trait of agreeableness. This was the only trait for which we can be confident that the difference in alphas was due to schematicity. The results confirmed our hypothesis; schematics responded more consistently across items than aschematics. Additionally, although schematics and aschematics differed in their emotionality means, and thus the results must be interpreted with caution, the numbers of people in each group were roughly equivalent, and so we tested the hypothesis and found support once again; schematics responded more consistently than aschematics.

As mentioned previously, aschematic and schematic alphas did not differ for honesty/humility, extraversion, conscientiousness, or openness to experience. This could have happened for a number of reasons. First, sample sizes may have been too inadequate or unbalanced to detect differences; for each of the traits, there was a large discrepancy between numbers of people in each group (see Table 1). For
example, 36 participants were aschematic and 161 were schematic for honesty/humility. Note that in the emotionality and agreeableness traits, the discrepancies were the smallest (29 and 8, respectively).

A closer look at some of the individual trait measures reveals that participants may have been responding to extraneous demands, or that the measure may not have been face valid. For example, the honesty/humility measure might have been problematic due to its morally-loaded nature; participants may have felt pressured to rate it as the most important when they may not have actually felt that way. The extraversion scale contained a subscale labeled Social Self-Esteem, which included the following three items: I feel reasonably satisfied with myself overall, I feel that I am an unpopular person (r), I sometimes feel that I am a worthless person (r). With the exception of the second item, this subscale seems to measure self-esteem with no social component and while extraversion and self-esteem have been found to be correlated (Abrams, 1988; Francis, 1997), it is conceivable that high self-esteem is not an inherent or necessary component of extraversion.

Interestingly, aschematic means were higher on each trait scale except those for agreeableness. This seems counterintuitive; the measure of schematicity required participants to indicate the trait’s importance, and one would think that a person who rates honesty as important would be more likely to rate honesty items a little higher than a person who indicates that it is not important. But perhaps this is where the separation between trait and schematicity is evidenced; just because a trait is important does not necessarily mean that the person possesses more of it, or is higher on it than someone who does not think it is important. The differences in means between aschematics and schematics (H = .46; E = .77; X = .47; C = .30; O = .22), while significant, were also very small which calls the meaningfulness of these findings into question. Effects sizes were all in the moderate range, except for emotionality, but this difference could have been driven by gender, thus resulting in a larger difference between the two groups. Importantly, the difference in means between aschematics and schematics on agreeableness did not differ, which allows one to be more confident that the difference between the schematic and
aschematic alphas was actually due to schematicity and not another systematic difference between the groups.

Study 1 provides some preliminary evidence that the amount of importance that a person places on a certain trait may affect the consistency of his or her responses on a measure of that trait. This study examined internal consistency, and the focus will now shift to the other type of reliability discussed previously: test-retest reliability.

III. STUDY TWO: RETEST RELIABILITY

A. Overview of Study Two

The purpose of study 2 was to examine retest reliability for people who were schematic or aschematic for the traits of independence and dependence. Specifically, study 2 was designed to test the hypothesis that people who are schematic for a trait would respond more reliably across different data collecting sessions than those who are aschematic. Initially, both dependence and independence traits were assessed; however, only 3 participants out of 1200 were either schematic or aschematic for dependence. Research suggests that, in terms of reference to self and others, dependence and independence might function as dependent constructs, measurable by using a bipolar spectrum (Witkin & Goodenough, 1977), and this is the way that Markus (1977) conceived of the construct. Thus, people who are schematic for independence may exclusively not be schematic for dependence. However, given that it may be a bipolar construct, ironically, one who is schematic for independence may automatically be schematic for “not-independence,” or, dependence as well. Because the response rate for dependent schematics was so low, the analyses of their data have been dropped; however, analyses examining independence schematics and aschematics responses to the dependence questionnaire and reaction times tasks will be included. Given the uncertainty of the relationship of dependence to independence, all analyses pertaining to the dependence questionnaire and dependence reaction times tasks are exploratory at this stage.
Participants completed this study in two waves. Independence schemas and trait independence were assessed during the first wave, and trait independence plus latency responses to trait-relevant adjectives were assessed during the second wave.

Latency response information was collected to ensure that schemas were actually measured; prior research (Markus, 1977) has shown that people who are schematic for a trait will respond more rapidly to trait-relevant adjectives. Thus, we expected that participants who were schematic for independence would respond more rapidly to independence-related self-descriptive adjectives than people who were aschematic for independence. Provided that we found evidence that we had measured schemas, we expected that schematic alphas would be higher than aschematic alphas across two different test sessions.

B. **Methods**

1. **Participants**

   Approximately 1200 undergraduate students enrolled in an introductory psychology course at the University of Illinois at Chicago participated in the first round of data collection for course credit. Out of those 1200 students, 557 participants qualified for the second wave of data collection; 477 met criteria for schematicity and 80 met criteria for aschematicity. The aforementioned students were invited by their personal identification numbers (PINs) to participate in the second wave of data collection through the university-maintained subject pool website (PECOLSUS), thus they could log in and view studies available to them, but they were not sent personal messages to return. Out of the 557 qualified students, 59 schematics and 19 aschematics returned to participate in the second round of data collection, but 3 participants were excluded because they remembered specific answers that they gave on the questionnaire during the first wave, 16 participants were excluded because they indicated that English was their second language, and 5 participants were excluded because they were more than 2 standard deviations away from the mean response time on the latency response task. 54 participants
remained (41 schematics, 13 aschematics), and thus all analyses include only those 54 participants. 24 men and 30 women participated (18 male and 23 female schematics, 6 male and 7 female aschematics), the age range was from 18 to 27 ($M_{age} = 19.43$ years, $SD_{age} = 2.02$), and racial identity was as follows: 40.7% White/Caucasian, 33.3% Asian/Pacific Islander, 16.7% Hispanic, 9.3% Black, and 3.7% Other.

2. **Procedure**

This study was conducted in two waves of data collection. The first wave of data was collected in a mass testing session and the following wave of data was collected in a psychology laboratory.

The first wave was conducted to identify participants who were schematic and aschematic for the traits of independence and dependence, and those participants who met the criteria for trait schematicity or aschematicity were eligible to participate in the subsequent wave. The first wave also served as pretest to confirm that there are people who are both schematic and aschematic for the given traits. First, participants completed a short questionnaire designed to assess schematicity for independence and dependence (Markus, 1977) and a questionnaire that assessed their dependence and independence traits (questionnaires described below), and those who met the criteria for schematicity and aschematicity were invited back for the second wave, identifiable only by university-assigned PIN numbers.

For the second wave, conducted approximately one to two months after the mass testing session, students arrived at the lab at their appointed times and completed the study in groups of 1 to 3. Upon arrival to the lab, students were greeted by one of three experimenters and were asked to choose one of three partitioned cubicles, each containing a computer on which the participant would complete the study, in which to sit. The experimenter then explained that the student would be prompted through the study on the computer and to let him/her know when the student was finished. After signing the paper consent form, participants completed the entire study, programmed using Media Lab,
on a computer. The experimenter remained in the room but separated from the participants by a mobile partition while the participants were prompted through the study.

First, they completed the same self-report questionnaire that they completed during mass testing, which consisted of independence, dependence, and control items (control items were added to the computer version). Reaction times, measured in milliseconds, were assessed for each of the dependence, independence, and control items and a separate latency response task was also completed. Last, participants completed the demographic measures. All measures are described below.

3. **Measures**

   a. **Schema information**

   We adapted our procedure for collecting schematic information directly from Markus (1977). First, participants were given three 11-point semantic differential scales (*Independent-Dependent, Individual-Conformist, Leader-Follower*) and instructed to indicate where they thought their personality best fell on each scale. The numbers at the low end of the scale corresponded to independence and those at the high end corresponded to dependence. It is important to note that these items were completed on paper during mass testing and on a computer during the lab portion. For the computer version, participants viewed one question on the screen at a time, and were taken to the next question only after they clicked the continue button at the bottom of the screen. Next, they were asked to indicate how important each of the three aforementioned traits were to their self description (e.g., *How important is this trait (dependence/independence) to your self-description? 1 = Not at all, 5 = Very much*). Last, participants completed an adjective checklist, randomized in the computer questionnaire, which included the words *responsible, intelligent, warm, independent, self-sufficient, dependent, agreeable, cultured, irresponsible, secure, reliant, calm, naïve, insecure, sad, introverted, athletic, and happy* with the instructions to check all of the words that, in general, describe their personalities (See Appendix B for complete instructions).
As a reminder, following Markus’ methods (1977), schematic participants are those who judge the independence/dependence trait as important to their self-concept and aschematic participants are those who do not see it as important to their self-concept. Participants who were schematic for the independence trait were those who indicated their personality as being on the extreme independence end (1-4) of two of the semantic differential scales, indicated that these traits were important (4-5), and checked “independent” on the adjective checklist. Participants who were schematic for dependence were those who indicated their personality as being on the extreme dependence end (8-11) of two of the semantic differential scales, indicated that these traits were important (4-5), and checked “dependent” on the adjective checklist. Participants who were aschematic were those who indicated their personality as being in the middle (5-7) of at least two of the semantic differential scales, indicated that these traits were relatively unimportant (1-2), and did not check either “independent” or “dependent” on the adjective checklist.

b. **Practice questions**

Because the dependent measures were timed, before participants completed them, we asked them to complete a few practice questions to familiarize them with the keys and mode of responding. Participants read the following instructions: “We would like to make sure that you are comfortable with the response buttons before you begin the next questionnaire. Therefore, these questions are filler questions and are not part of the experiment.” These items included: *I like vegetables, I am more of a dog person than a cat person, I like to listen to pop music, I have visited the state of Florida in my life, and I enjoy going to art galleries.* Item responses were on a 5-point scale (*1 = Not at all like me, 2 = A little like me, 3 = Moderately like me, 4 = Much like me, 5 = Very much like me)*.

c. **Independence questionnaire**

Both the independence and dependence questionnaires were taken from the International Personality Item Pool (IPIP) website, which provides to researchers, without charge, scales
comparable to numerous published psychometric instruments (Goldberg, 1999a, 1999b). Thus, the independence scale is a comparable IPIP version of the independence subscale (see Appendix C) from the Six Factor Personality Questionnaire (6FPQ; Jackson, Paunonen, & Tremblay, 2000). The IPIP items have shown to have a modest, if not fairly low internal consistency coefficient (α = .69) and the scale is highly correlated with the original 6FPQ scale (r = .86; http://ipip.ori.org/). These items included: I don’t care about dressing nicely, I feel it’s okay that some people don’t like me, I sail my own course, I want to be liked (rs), I love to be complimented (rs), and I need the approval of others (rs). Item responses were on a 5-point scale (1 = Not at all like me, 5 = Very much like me). High scores indicated more independence. Dependence and independence items were interspersed for mass testing and these items (and control items) were randomized on the computer version.

d.  **Dependence questionnaire**

The dependence questionnaire (see Appendix D) is a comparable IPIP version of the dependence subscale from the Temperament and Character Inventory (TCI; Cloninger et al., 1994). The internal consistency coefficient for the IPIP items is fairly low (α = .67), but the items are highly correlated (r = .90) with the original TCI items (http://ipip.ori.org/). These items included: I try to please everyone, I follow directions, I do what others want me to do, I give in to no one (rs), I want to be different from others (rs), and I don’t care what others think (rs). Item responses were on a 5-point scale (1 = Not at all like me, 5 = Very much like me). High scores indicated more dependence. Each scale was composed of six items and three were reverse-scored for a total of twelve trait items.

e.  **Control items**

Control items were included as a manipulation check in the second wave. To ensure that participants did not differ in other respects such as general response time and to provide evidence that the differences in reaction times were due to the schema-relevant nature of the questions, 12 control items were constructed; each control sentence matched a corresponding
dependence or independence item on syllabic length and frequency of word (Carroll, Davies, & Richman, 1971) in the English language (see Appendix E). These items were meant to reflect simple preferences and opinions, and the two groups were not expected to vary in response times to these questions. Items included: I try to work out a lot, I practice medicine, I do many things for my family, I don’t move often enough, I have many children, I like to shop for appliances, I don’t care about saving coupons, I feel it’s okay that people don’t recycle, I always behave well, I love to go water skiing, I want to be wealthy, and I need to drink water often. Item responses were on a 5-point scale (1 = Not at all like me, 5 = Very much like me). The independence, dependence and control items were randomized across participants, and they viewed one question at a time.

f. Reaction times

Reaction times for each of the self-report items were measured in milliseconds. Markus (1977) measured reaction times and found that those who were schematic for independence and dependence identified trait-relevant adjectives quicker than aschematics, therefore I also measured reaction times in two different ways. First, Medialab recorded the length of time that participants spent on each question. We expected that participants who were schematic for independence would be quicker than aschematics to answer the independence items, respectively. However, we did not anticipate a difference between schematics and aschematics on the control items.

g. Latency response task

After participants completed the dependence, independence, and control items, they completed the latency response task, programmed using DirectRT. Participants were told that they would be classifying words as either self-descriptive or not using just the ‘e’ and ‘i’ keys. They were instructed to go as quickly as they could while making as few mistakes as possible and they were given five practice words to classify (nice, cool, student, tired, employee) to become familiar with the task. During the task, the words ‘like me’ and ‘not like me’ appeared in the upper right and left hand corners
and the ‘e’ and ‘i’ keys corresponded to each category. To focus participants’ attention, a mask (XXXXXX) was flashed in the center of the screen for 120 milliseconds after which the target word appeared in the same place. Participants indicated whether the word described them or not by pushing the ‘e’ or ‘i’ key, and then the mask for the next word appeared. Participants responded to 60 randomized target adjectives.

Using Markus’ method (1977), the 60 words were chosen from Anderson’s (1968) list of 555 trait adjectives. 15 pretested words were related to independence and nonconformity (independent, individualistic, ambitious, adventurous, self-confident, dominating, self-sufficient, aloof, arrogant, egotistical, unconventional, outspoken, aggressive, assertive, uninhibited) and 15 words were related to dependence and conformity (dependable, cooperative, tactful, tolerant, unselfish, impressionable, conforming, dependent, timid, submissive, conventional, moderate, obliging, self-denying, cautious). 30 comparison words were also selected to match the schema-related adjectives. According to Anderson (1968), 10 were rated as positive (imaginative, open-minded, enthusiastic, ambitious, talented, creative, ethical, productive, inventive, interesting), 10 were negative (grouchy, shallow, petty, cold, childish, ungrateful, humorless, irresponsible, bragging, unfair), and 10 were neutral (inoffensive, perfectionistic, conservative, average, unpredictable, blunt, solemn, self-righteous, unlucky, restless).

We predicted that participants who were schematic for independence would respond more quickly to the independence words than aschematics, and that schematics and aschematics would not differ in their reaction times to the neutral words.

h. **Demographics and suspicion probes**

After the latency response task, to ensure participants were paying attention, participants indicated what key on the previous screen corresponded to ‘like me.’ Two participants were excluded for providing the wrong response. Because waves one and two of data collection were very similar, participants also completed some probes designed to assess whether and how much they
remembered from the first wave of data collection. Specifically, they were asked if they remembered any of the current studies’ tasks from mass testing, which questions (specifically) they remembered, and whether they remembered their answers on the previous questionnaire.

Finally, participants indicated their age, gender, race/ethnicity, and if English was their primary language. This question was used as a filter because the reaction times tasks require participants to respond as quickly as possible, if English is not the primary language, reaction times may be slower due to translation and not schematicity.

C. Results

1. Preliminary Analyses

Before reporting the preliminary and main analyses, it is important to note that, although 1200 people participated in wave one, only 54 participated in wave two. Furthermore, 41 were schematic but only 13 were aschematic for independence. The extremely low number of aschematic participants renders the test of the hypothesis virtually uninterpretable. Nevertheless, hypothesis testing was conducted, as it was possible that trends may be revealed.

Similar to study 1, prior to testing the main hypothesis, we conducted a preliminary analysis by comparing schematic and aschematic means on the independence and dependence questionnaires. If there is a difference in trait means between schematics and aschematics, then this measurement confound renders interpretation of the main analysis difficult.

We also examined the internal consistency of the independence and dependence items for schematics and aschematics in wave one. Following the evidence obtained in study 1, schematics should be more consistent within items than aschematics on the independence scale. However, it is important to note that it is possible to obtain a high retest coefficient even if internal consistency is low, as these two statistics represent different sources of variance; variance in scores between testing sessions is
considered to be a source of error and variance between items in one test (within person) is considered to be true variance (Guilford & Michael, 1950).

Additionally, we examined the correlations between the independence-dependence semantic differential scale (schematicity measure) and the independence and dependence questionnaires; if participants are conceptualizing the schematicity measure and the trait measure as similar or the same, the correlation between the schematicity measure and the independence questionnaire should be negative for schematics and non-significant for aschematics. Recall that lower numbers on the semantic differential scale and higher numbers on the independence questionnaire mean more independence.

While it is a bit less clear, previous research (Witkin & Goodenough, 1977) would suggest that the correlation between the schematicity measure and the dependence questionnaire should be positive for schematics and non-significant for aschematics; lower numbers on the dependence questionnaire mean more independence.

a. **Trait mean comparison**

To examine the differences between schematic and aschematic independence and dependence questionnaire means, a one-way ANOVA was conducted on both the first and second wave data. Schematics and aschematics did not differ in their respective levels of independence during wave one \( (M_{schem} = 3.01; M_{aschem} = 2.90), F (1, 52) = .31, ns \), and wave two \( (M_{schem} = 2.96; M_{aschem} = 3.00), F (1, 52) = .04, ns \). Additionally, they did not differ in their levels of dependence during wave one \( (M_{schem} = 2.93; M_{aschem} = 3.18), F (1, 52) = 1.38, ns \), and wave two \( (M_{schem} = 2.93; M_{aschem} = 3.08), F (1, 52) = .67, ns \).

b. **Wave one alphas and correlations**

To bolster statistical power, schematic and aschematic internal consistency alphas were calculated for the 563 participants (477 schematics and 80 aschematics) from wave one exclusively. Aschematic and schematic alphas were compared using the Feldt test for independent
samples. Results revealed that, on the independence questionnaire, aschematics responded more consistently than schematics (\(\alpha_{\text{aschem}} = .801, \alpha_{\text{schem}} = .580\)), and this difference was significant, \(F(75, 470) = 2.11, p < .001\). On the dependence questionnaire, aschematics responded more consistently than schematics (\(\alpha_{\text{aschem}} = .861, \alpha_{\text{schem}} = .545\)), and this difference was also significant, \(F(74, 471) = 3.27, p < .001\).

Because the preceding results were so robust, a Pearson product-moment correlation was conducted to examine the relationships between the items, specifically, the schematic and aschematic inter-item correlations on the dependence and independence scales separately. Results revealed that the aschematic inter-item correlations were higher for every item except one on both the independence and dependence scales (See Tables 3 and 4).

A Pearson product-moment correlation coefficient was computed to assess the relationship between the independence-dependence semantic differential scale and the independence and dependence questionnaires. As expected, for schematics, the differential scale and the independence questionnaire were negatively correlated, \(r(39) = -.36, p < .05\), and the differential scale and dependence questionnaire were positively correlated, \(r(39) = .36, p < .05\). As expected, for aschematics, the correlations between the differential scale and the independence questionnaire and dependence questionnaire were not significant.

1. **Primary Analyses**
   
   a. **Reaction times**

      First, a one-way ANOVA was conducted to compare schematic and aschematic reaction times to the practice items; as expected, their reaction times did not differ. Next, a one-way ANOVA was conducted to compare aschematic and schematic reaction times for the control items; the two groups were not expected to differ on any of the items. However, schematics and aschematics
Table 3
Aschematic and schematic inter-item correlations on the independence scale

<table>
<thead>
<tr>
<th></th>
<th>Aschematic</th>
<th></th>
<th></th>
<th></th>
<th>Schematic</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.550**</td>
<td></td>
<td></td>
<td></td>
<td>.199**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.497**</td>
<td>.591**</td>
<td></td>
<td></td>
<td>.129**</td>
<td>.428**</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.222</td>
<td>.164</td>
<td>.133</td>
<td></td>
<td>.251**</td>
<td>.070</td>
<td>.113*</td>
</tr>
<tr>
<td>5</td>
<td>.263*</td>
<td>.206</td>
<td>.162</td>
<td>.749**</td>
<td>.130**</td>
<td>.203**</td>
<td>.030</td>
</tr>
<tr>
<td>6</td>
<td>.319**</td>
<td>.387**</td>
<td>.297**</td>
<td>.645**</td>
<td>.788**</td>
<td>.009</td>
<td>.201**</td>
</tr>
</tbody>
</table>

* p < .05  
** p < .01

Table 4
Aschematic and schematic inter-item correlations on the dependence scale

<table>
<thead>
<tr>
<th></th>
<th>Aschematic</th>
<th></th>
<th></th>
<th></th>
<th>Schematic</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.683**</td>
<td></td>
<td></td>
<td></td>
<td>.280**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.647**</td>
<td>.660**</td>
<td></td>
<td></td>
<td>.505**</td>
<td>.323**</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.429**</td>
<td>.515**</td>
<td>.493**</td>
<td></td>
<td>.092*</td>
<td>.093*</td>
<td>.095*</td>
</tr>
<tr>
<td>5</td>
<td>.521**</td>
<td>.570**</td>
<td>.449**</td>
<td>.650**</td>
<td>.101*</td>
<td>.084</td>
<td>.113*</td>
</tr>
<tr>
<td>6</td>
<td>.307**</td>
<td>.343**</td>
<td>.366**</td>
<td>.480**</td>
<td>.533**</td>
<td>-.065</td>
<td>-.037</td>
</tr>
</tbody>
</table>

* p < .05  
** p < .01
differed in their reaction times to the control items; schematics were quicker than aschematics ($M_{schem} = 3496.13$ ms, $SD = 914.00$; $M_{aschem} = 4130.03$ ms, $SD = 1028.02$) to respond to the control items, $F(1, 52) = 4.47, p < .05$. A closer inspection of the items revealed that the groups did not differ on ten of the twelve items; however, for the item *I try to work out a lot*, schematics ($M = 2688.20$ ms, $SD = 1185.40$) were quicker to respond than aschematics ($M = 3568.62$ ms, $SD = 1512.81$), $F(1, 52) = 4.76, p < .05$. For the item *I do many things for my family*, schematics ($M = 2833.80$ ms, $SD = 1023.95$) were also quicker to respond than aschematics ($M = 5051.77$ ms, $SD = 6019.43$), $F(1, 52) = 5.30, p < .05$.

A one-way ANOVA was also conducted to compare reaction times for the independence items; schematics were expected to answer more quickly than aschematics. Schematics and aschematics differed in their reaction times to the independence items; schematics were quicker than aschematics ($M_{schem} = 3356.16$ ms, $SD = 861.82$; $M_{aschem} = 4154.21$ ms, $SD = 1409.43$) to respond, $F(1, 52) = 6.11, p < .05$. For all of the items, schematics were quicker to answer than aschematics; however, a closer inspection of the items revealed that reaction time differences for only two of the items reached significance. These items were *I feel it’s okay that some people don’t like me* ($M_{schem} = 3174.66$ ms, $SD = 1419.57$; $M_{aschem} = 4451.85$ ms, $SD = 2441.02$), $F(1, 52) = 5.50, p < .05$, and *I love to be complimented* ($M_{schem} = 2751.57$ ms, $SD = 1095.59$; $M_{aschem} = 3908.46$ ms, $SD = 2446.33$), $F(1, 52) = 5.74, p < .05$.

Additionally, two other items were marginally significant: *I need the approval of others* ($M_{schem} = 2510.59$ ms, $SD = 1452.06$; $M_{aschem} = 3389.38$ ms, $SD = 2245.92$), $F(1, 52) = 2.74, p = .10$, and *I want to be liked* ($M_{schem} = 3070.78$ ms, $SD = 1595.11$; $M_{aschem} = 4003.46$ ms, $SD = 2429.98$), $F(1, 52) = 2.59, p = .11$.

Finally, a one-way ANOVA was conducted to compare reaction times for the dependence items; it was unclear whether schematics would be quicker than aschematics to respond and thus this analysis was exploratory; if participants viewed dependence and independence as 2 sides of the same construct, it is possible that schematics would respond more quickly to the dependence item. However, the results
revealed no significant differences between schematics and aschematics at either the level of the scale or the level of the individual items.

b. **Latency response task**

First, a one-way ANOVA was conducted for schematic and aschematic latency responses on the independent words; schematics ($M = 1054.20$ ms, $SD = 224.69$) responded more quickly than aschematics ($M = 1462.25$, $SD = 476.17$), $F (1, 52) = 18.03$, $p < .01$. A closer inspection of the items revealed that schematics responded more quickly than aschematics to eleven out of the fifteen words. These words included: *independent, individualistic, ambitious, adventurous, dominating, self-sufficient, arrogant, egotistical, unconventional, assertive, and uninhibited* (see Table 5).

A one-way ANOVA was also conducted for schematic and aschematic latency response times on the dependent words. Schematics ($M = 1179.68$ ms, $SD = 329.80$) responded more quickly than aschematics ($M = 1564.27$ ms, $SD = 595.51$), $F (1, 52) = 8.82$, $p < .01$. A closer inspection of the items revealed that schematics responded more quickly than aschematics on eight out of the fifteen dependent words. These words included: *dependable, tactful, tolerant, impressionable, conforming, dependent, submissive, and self-denying* (see Table 6).

Next, a one-way ANOVA was conducted for latency responses for the 30 neutral words; schematics and aschematics were not expected to differ on these items. Unexpectedly, schematics ($M = 1083.61$ ms, $SD = 219.03$) responded more quickly than aschematics ($M = 1378.75$, $SD = 359.98$), $F (1, 52) = 12.87$, $p < .01$. A closer inspection of the items revealed that schematics responded more quickly than aschematics to 10 out of the 30 words. These words included: *shallow, enthusiastic, petty, ambitious, unpredictable, ungrateful, humorless, irresponsive, inventive, and restless* (see Table 7).
Table 5
Means and standard deviations for latency responses to independent words

<table>
<thead>
<tr>
<th></th>
<th>Aschematics</th>
<th>Schematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Independent</td>
<td>1560.31 (1012.59)</td>
<td>874.10 (403.82)**</td>
</tr>
<tr>
<td>Individualistic</td>
<td>1856.08 (1093.86)</td>
<td>967.88 (290.26)***</td>
</tr>
<tr>
<td>Ambitious</td>
<td>1369.85 (821.31)</td>
<td>920.51 (304.53)**</td>
</tr>
<tr>
<td>Adventurous</td>
<td>1207.77 (486.38)</td>
<td>888.68 (324.08)**</td>
</tr>
<tr>
<td>Self-confident</td>
<td>1107.77 (431.70)</td>
<td>947.61 (467.48)</td>
</tr>
<tr>
<td>Dominating</td>
<td>1409.77 (703.55)</td>
<td>1068.20 (417.39)*</td>
</tr>
<tr>
<td>Self-sufficient</td>
<td>1518.85 (676.97)</td>
<td>1008.80 (424.16)**</td>
</tr>
<tr>
<td>Aloof</td>
<td>1230.62 (677.09)</td>
<td>1093.83 (498.25)</td>
</tr>
<tr>
<td>Arrogant</td>
<td>1331.00 (497.77)</td>
<td>975.44 (343.00)**</td>
</tr>
<tr>
<td>Egotistical</td>
<td>1703.92 (811.75)</td>
<td>1218.76 (544.67)*</td>
</tr>
<tr>
<td>Unconventional</td>
<td>1612.38 (658.98)</td>
<td>1253.22 (546.44)*</td>
</tr>
<tr>
<td>Outspoken</td>
<td>1528.15 (804.31)</td>
<td>1236.24 (769.33)</td>
</tr>
<tr>
<td>Aggressive</td>
<td>973.92 (286.80)</td>
<td>974.46 (389.38)</td>
</tr>
<tr>
<td>Assertive</td>
<td>1495.08 (605.58)</td>
<td>1070.63 (449.96)**</td>
</tr>
<tr>
<td>Uninhibited</td>
<td>2028.31 (1345.23)</td>
<td>1314.66 (679.05)*</td>
</tr>
</tbody>
</table>

* $p < .05$
** $p < .01$
*** $p < .001$
Table 6
Means and standard deviations for latency responses to dependent words

<table>
<thead>
<tr>
<th></th>
<th>Aschematics</th>
<th>Schematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
</tr>
<tr>
<td>Dependable</td>
<td>1374.23 (691.75)</td>
<td>1049.80 (409.91)*</td>
</tr>
<tr>
<td>Cooperative</td>
<td>1078.62 (567.79)</td>
<td>889.80 (222.21)</td>
</tr>
<tr>
<td>Tactful</td>
<td>1515.54 (975.92)</td>
<td>1021.54 (302.10)**</td>
</tr>
<tr>
<td>Tolerant</td>
<td>1745.69 (1838.51)</td>
<td>1065.44 (314.47)*</td>
</tr>
<tr>
<td>Unselfish</td>
<td>1603.38 (725.65)</td>
<td>1613.15 (1219.63)</td>
</tr>
<tr>
<td>Impressionable</td>
<td>2227.77 (1191.57)</td>
<td>1510.05 (816.75)*</td>
</tr>
<tr>
<td>Conforming</td>
<td>1774.85 (1274.71)</td>
<td>1148.54 (550.72)*</td>
</tr>
<tr>
<td>Dependent</td>
<td>1731.77 (1683.10)</td>
<td>1079.22 (681.63)*</td>
</tr>
<tr>
<td>Timid</td>
<td>1274.85 (461.34)</td>
<td>1054.37 (342.98)</td>
</tr>
<tr>
<td>Submissive</td>
<td>1888.92 (1237.94)</td>
<td>1172.10 (570.53)**</td>
</tr>
<tr>
<td>Conventional</td>
<td>1347.92 (463.79)</td>
<td>1169.68 (609.38)</td>
</tr>
<tr>
<td>Moderate</td>
<td>1344.62 (724.49)</td>
<td>1136.56 (621.90)</td>
</tr>
<tr>
<td>Obliging</td>
<td>1313.15 (513.92)</td>
<td>1368.41 (760.60)</td>
</tr>
<tr>
<td>Self-denying</td>
<td>2049.23 (1322.35)</td>
<td>1292.68 (605.08)**</td>
</tr>
<tr>
<td>Cautious</td>
<td>1193.46 (436.25)</td>
<td>1123.85 (457.09)*</td>
</tr>
</tbody>
</table>

* $p < .05$
** $p < .01$
<table>
<thead>
<tr>
<th>Trait</th>
<th>Aschematics M (SD)</th>
<th>Schematics M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaginative</td>
<td>1085.23 (423.09)</td>
<td>937.85 (300.03)</td>
</tr>
<tr>
<td>Inoffensive</td>
<td>1865.92 (1101.62)</td>
<td>1847.17 (1417.40)</td>
</tr>
<tr>
<td>Grouchy</td>
<td>980.62 (259.92)</td>
<td>940.10 (209.44)</td>
</tr>
<tr>
<td>Open-minded</td>
<td>967.77 (228.00)</td>
<td>856.12 (172.61)</td>
</tr>
<tr>
<td>Perfectionistic</td>
<td>1373.69 (447.42)</td>
<td>1206.51 (712.46)</td>
</tr>
<tr>
<td>Shallow</td>
<td>1328.85 (634.54)</td>
<td>948.05 (314.37)**</td>
</tr>
<tr>
<td>Enthusiastic</td>
<td>1164.85 (414.53)</td>
<td>874.15 (218.54)**</td>
</tr>
<tr>
<td>Conservative</td>
<td>1304.00 (550.38)</td>
<td>1089.73 (480.05)</td>
</tr>
<tr>
<td>Petty</td>
<td>1603.08 (752.50)</td>
<td>1194.22 (496.81)*</td>
</tr>
<tr>
<td>Ambitious</td>
<td>1488.92 (1455.19)</td>
<td>872.85 (243.60)*</td>
</tr>
<tr>
<td>Average</td>
<td>1185.77 (385.92)</td>
<td>1208.83 (558.39)</td>
</tr>
<tr>
<td>Cold</td>
<td>1379.77 (599.52)</td>
<td>1097.41 (412.12)</td>
</tr>
<tr>
<td>Talented</td>
<td>979.54 (243.98)</td>
<td>891.54 (349.40)</td>
</tr>
<tr>
<td>Unpredictable</td>
<td>2114.15 (1466.59)</td>
<td>1207.49 (407.62)**</td>
</tr>
<tr>
<td>Childish</td>
<td>1239.00 (332.63)</td>
<td>1040.83 (496.12)</td>
</tr>
<tr>
<td>Ethical</td>
<td>1418.00 (797.57)</td>
<td>1109.05 (462.32)</td>
</tr>
<tr>
<td>Blunt</td>
<td>1038.54 (218.60)</td>
<td>1079.76 (419.89)</td>
</tr>
<tr>
<td>Ungrateful</td>
<td>1666.85 (1221.04)</td>
<td>1078.66 (578.63)*</td>
</tr>
<tr>
<td>Productive</td>
<td>1028.00 (404.53)</td>
<td>851.61 (296.04)</td>
</tr>
<tr>
<td>Solemn</td>
<td>1791.62 (1163.84)</td>
<td>1283.88 (817.64)</td>
</tr>
<tr>
<td>Humorless</td>
<td>1978.23 (1365.04)</td>
<td>1227.78 (462.89)**</td>
</tr>
<tr>
<td>Creative</td>
<td>1158.38 (879.88)</td>
<td>864.17 (321.23)</td>
</tr>
<tr>
<td>Self-righteous</td>
<td>1444.69 (572.62)</td>
<td>1200.39 (535.79)</td>
</tr>
<tr>
<td>Irresponsible</td>
<td>1219.23 (419.51)</td>
<td>990.78 (307.95)*</td>
</tr>
<tr>
<td>Inventive</td>
<td>1804.62 (932.08)</td>
<td>1205.44 (563.19)**</td>
</tr>
<tr>
<td>Unlucky</td>
<td>1459.23 (658.84)</td>
<td>1216.24 (604.11)</td>
</tr>
<tr>
<td>Bragging</td>
<td>1248.46 (465.44)</td>
<td>1088.15 (418.49)</td>
</tr>
<tr>
<td>Interesting</td>
<td>926.85 (336.92)</td>
<td>872.88 (329.25)</td>
</tr>
<tr>
<td>Restless</td>
<td>2012.15 (1715.28)</td>
<td>1271.63 (672.88)*</td>
</tr>
<tr>
<td>Unfair</td>
<td>1106.54 (355.01)</td>
<td>954.93 (326.55)</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01
c. **Retest alpha comparison**

To assess retest reliability, the alpha coefficients for the independence and dependence questionnaires for each group were calculated; on independence, the schematic alpha was slightly higher than the schematic alpha between time one and time two ($\alpha_{schem} = .768$; $\alpha_{aschem} = .852$). However, Feldt’s test for independent samples indicated that this difference was not significant, $F_{(12, 40)} = 1.57$, $ns$. On the dependence questionnaire, again, the aschematic alpha was slightly higher than the schematic alpha ($\alpha_{schem} = .753$; $\alpha_{aschem} = .809$) but this difference was not significant.

Additionally, the test-retest coefficients for schematics and aschematics were compared using slightly different criteria for 43chematicity. Latency responses for trait-relevant and neutral words were examined as manipulation checks for measuring schemas; if schematics do respond to trait-relevant words more quickly than aschematics, then we should be able to examine test-retest reliability as a function of latency responses. A median split was used to divide the participants into two groups; because latency responses are being used as a proxy for schemas, the comparison of test-retest alphas for participants is the fast and slow groups should mirror the previous results. This is what was found; for latency responses to the independent words, there was no significant difference between fast responders ($\alpha = .717$) and slow responders ($\alpha = .649$) on the independence questionnaire, $F_{(27, 27)} = 1.24$, $ns$. There was also no difference between fast responders ($\alpha = .745$) and slow responders ($\alpha = .781$) on the dependence questionnaire, $F_{(27, 27)} = 1.16$, $ns$. Additionally, latency responses to the dependent words were also examined; again, a median split divided the participants into fast and slow responders. There was no significant difference between fast responders ($\alpha = .781$) and slow responders ($\alpha = .741$) on either the dependence scale $F_{(27, 27)} = 1.18$, $ns$, or the independence scale ($\alpha_{fast} = .709$, $\alpha_{slow} = .649$), $F_{(27, 27)} = 1.21$, $ns$. Finally, latency responses to the neutral words were examined; there is no reason to expect a difference between fast and slow responders on either scale. A median split divided the participants into fast and slow responders. There was no significant difference between fast
responders (α = .721) and slow responders (α = .763) on either the dependence scale $F(27, 27) = 1.17, ns$, or the independence scale ($\alpha_{fast} = .628, \alpha_{slow} = .725$), $F(27, 27) = 1.35, ns$.

D. Discussion

The unbalanced numbers of participants in each group, and the extremely low number of participants in the aschematic group did not provide a powerful test of the hypothesis. However, a lack of difference between schematic and aschematic trait means, as well as correlations in the expected direction between the schematicity measure and independence and dependence questionnaire allowed a test of the hypothesis. Schematics were quicker to identify a high percentage of trait-related words and a much lower percentage of neutral words, but contrary to the hypothesis, schematics and aschematics, measured using Markus’ (1977) criteria and latency responses, did not differ in the consistency of their responses across testing sessions.

Although this study was meant to examine retest reliability, the internal consistency alpha was calculated for schematics and aschematics, in part to replicate the findings from study 1, but also to check the psychometric properties of the scales used in the study. The results were surprising: aschematics were more internally consistent than schematics on both scales. This could have happened for a number of reasons; perhaps the items were too disparate in meaning, or because of the low numbers of aschematics a Type II error was committed. These are possibilities, but perhaps the most plausible explanation pertains to the semantics of the word “independent.” The Oxford English Dictionary (Independent, 1971) lists 9 definitions of the word independent and 6 definitions of the word dependent (Dependent, 1971), and within each definition lists several uses. For schematics, the word independent probably possesses a special meaning in some way, but perhaps it pertains to any of the interpretations; they may be thinking of independence in various domains such as financial, parental, relationship, political, or gender-related. However, there may also be an English language prototype of the word; the first definition for independent listed in the Oxford English dictionary is, “Not depending
on the authority of another” (Independent, 1971), the first definition in the Merriam-Webster online dictionary is, “Not subject to control by others” (Independent, 2012), and the first definition in the American Heritage online dictionary is, “Not governed by a foreign power; self-governing” (Independent, 2011). When responding to the items, aschematics may all rely on this one prototype, meaning a general autonomy, and this would account for greater consistency among aschematics than schematics.

Additionally, because the difference between schematic and aschematic alphas was significant, to examine the relationship more closely, inter-item correlations were calculated for each group on the dependent and independent scales. Again, the results were surprising; aschematic inter-item correlations were larger than the schematic correlations on every item. This seems to follow the idea that schematics and aschematics have different concepts of the independence and dependence constructs. For example, Cervone (2004) found that two separate schematic participants who both indicated that independence was their most important trait indicated different situation profiles (with no overlap) in which independence was relevant. It remains to be seen whether schematic independent people think of the meaning of the word in the same way, but the trait itself is relevant in completely different contexts, which suggests that it is domain-specific at the least. One could also imagine a situation in which people who are schematic for one certain sport, may actually provide low reliability on a questionnaire that measures a general sportiness trait. These findings also suggest that people think of dependence and independence as a bipolar construct, as this analysis was conducted on participants who were schematic for independence, and unless independence and dependence are polar opposites, we would not expect a difference between schematic and aschematic correlations on this scale. A complete discussion of this difference is beyond the scope of this paper, but warrants further future investigation.
Although there were two exceptions, schematics and aschematics did not differ in their reaction times to most of the control items, as hypothesized. I made this comparison to show that the differences in their reaction times to answering the independence questions were specific to those questions, and therefore tapping the schema-relevant construct, as there was no reason why they should differ on items that were unrelated to the independence construct. There was no difference between schematics and aschematics on the dependence scale or individual items. For the independence items, schematics were quicker to respond to two of the six items, although they trended toward quicker reaction times for all of the items. There may be a few reasons why they were not quicker to respond to all of the items. First, and this is a problem for the entire study, there were only 13 participants in the aschematic condition; this really does not provide enough statistical power to detect significant differences. Second, the nature of the reaction time measure was a bit uncontrolled, that is, participants were not instructed to answer as quickly as possible, and there were five response options that were labeled with numbers (1-5) and words (Not at all like me – Very much like me). Thus participants not only had to read complete sentences (the items), but they also had to read the option labels and then decide which one out of five was best. It is conceivable that the participants varied in their reading speed and processing, so I also included the latency task measure to more tightly control for those individual differences.

The results of the latency response task were mixed; schematics responded more quickly to 73% of the independence words and 53% of the dependence words, but they also responded more quickly to 33% of the neutral words. On one hand, there is no theoretical reason as to why they would differ on any of the neutral words, but on the other, schematics responded quicker than aschematics to a much greater proportion of the schema-relevant words. A closer look at the data reveals that schematics were quicker to respond to virtually every word. The problem, of course, lies in the unequal distribution of participants in conditions as well as the extremely low number of participants in the aschematic
condition. Additionally, participants could have been primed with the construct by the high ratio of trait-relevant words, and as a result could have responded quicker in subsequent trials; in a latency response task, filler items should comprise 2/3 of the total words, however the filler words only comprised 50% of the total number. However, the comparison of fast and slow responder alphas mirrored the findings that compared alphas using Markus’ (1977) criteria for measuring schematicity, which suggests that the latency response task was tapping into schemas.

Despite these problems, the schematics did respond faster to a much higher proportion of the trait-relevant words than the neutral words, which suggests that trait-relevant knowledge is more accessible for those who possess the trait schema, and thus looks to be a promising start.

Unfortunately, for test-retest reliability, the alphas for schematics and aschematics did not significantly differ, and the trend was in the opposite direction as the hypothesis. This could have happened for several reasons. As stated before, the biggest problem in this study was sample size; a larger sample was needed to detect meaningful differences. The face validity of some of the items may have also been problematic; it is questionable whether the item, “I don’t care about dressing nicely” is actually tapping the independence construct, although I removed this item, reanalyzed the data, and the results were virtually the same. Additionally, the IPIP website reports the alpha for the independence scale as .69; by most standards, the reliability for these items is too low and would be considered unacceptable to use in a personality study. The items may simply be bad. However, a test that has a pre-established reliability that is not too high was perfect for use in this study; it suggests that participants had room to vary, and that the results would not be constrained by the ceiling effects of a high reliability test.

IV. GENERAL DISCUSSION

A. Schematicity
In two studies I tested the hypothesis that participants who possess a schema for a certain personality trait would provide more reliable responses than participants who are aschematic for that trait. Specifically, in study 1 I tested the hypothesis that participants who were schematic for a certain trait would provide higher inter-item alphas (internal consistency) than participants who were not schematic for that trait, and in study 2 I tested the hypothesis that participants who were schematic for independence would provide higher alphas across testing sessions (test-retest) than participants who were aschematic for independence. While the results were certainly not as clean-cut as my hypotheses suggested, I did obtain preliminary evidence for schema-driven reliability in study 1.

Specifically, in study 1, participants who indicated that agreeableness and emotionality were important and central to them had higher alphas on agreeableness and emotionality scales, respectively, than people who did not see these traits as central and important to who they were. Thus, these scales, as hypothesized, were differentially reliable for two different groups of people. Therefore, it can be concluded that reliability cannot be construed as a property of measurement tools because in that construal, there is no way to explain why reliability varied for two different groups of people; perhaps reliability as an individual difference offers a clearer explanation.

In study 2, I did not obtain evidence that schematics provided more consistent responses over time than aschematics, but schematics were quicker to respond to trait-relevant words, which at least offers evidence that schemas are partially responsible for driving people’s responses.

The results of these studies, namely study 1, suggest that a systematic difference between people will lead some to be more consistent with their responses to test items than others. Schemas are relatively stable structures that we all possess, that inform us in our opinions, decisions, and behaviors in almost any domain, and that can account for consistency across and within people and situations. Many times schemas are also culture-bound. For example, people who have a “birthday party schema,” will follow a certain protocol; they will bring gifts, know the words to “Happy Birthday,” and even be
familiar with typical birthday party games. A person who had just arrived in the United States from a foreign country may not have this schema, would not know what to do, and if we were unaware, we might find that person’s behavior odd or even rude; he or she would not sing to the birthday person, show up without a gift, and maybe not participate in the games. Similarly, a person who does not possess a schema for something on which they are being tested may not know what to do or how to answer; in this instance one would perhaps rely on heuristics, mood, or prototype examples that easily come to mind. Although, as suggested previously, relying on a prototype may lead to more consistent responses, but may not be a true score or indicative of what the test is trying to measure.

The importance of taking schemas into account seems especially pertinent when measuring traits that are latent constructs; it is difficult to use traits alone as the explanation for reliability. As Borsboom et al. (2003) have suggested, latent variables such as the five-factor traits fail at the individual level of explanation because it is impossible for them to meet the covariation criterion for causality. That is, to say that one thing caused another, one must be able to demonstrate that one independent event preceded the other, and that the second event would not have happened without the first. This is not possible with latent traits because the trait is not independent of the measurement instrument, and thus one could not show that one’s conscientiousness scores happened without conscientiousness because the scores are the only way to observe conscientiousness. Then, attempting to explain behavior or study findings becomes an exercise in tautology. If we were simply to test the reliability of the HEXACO scale (without measuring reliability) and then ask why reliability for honesty was high, we would be forced to say that it was because participants were high in honesty. While this will always be the case with self-report measures of traits, schemas may lend explanatory power in this ontological bind. If we are able to show that schemas (which are in and of themselves latent variables) co-vary with traits, especially intra-individually, to produce differential reliability coefficients, then we may be more confident that our psychometrics are capturing real psychological properties.
Of course, it only makes sense to measure schemas for trait reliability if we are able to clearly demonstrate, or have a thorough understanding that they are two clearly separate constructs; this is not so easy. There are several ways in which traits are conceived and measured: a physical entity, a behavior or set of behaviors, and latently (as in, the five factor traits). For example, a physical entity may be something like height; it is able to be visually observed and clearly measurable by using a device such as a yardstick. In this case, there is an obvious difference between the trait and a height-schema; one who thinks a lot about one’s height may be tall or short as would a person who never thinks about it, but there is probably no relationship between them. Athleticism is a trait that could be measured by a set of behaviors such as frequency, type, and number of sports in which one engages. Again, there seems to be a clear difference between the trait, defined as a set of behaviors, and the schema, defined as a knowledge structure. Here, however, there may be a correlation between the two; one who is highly athletic (or plays sports frequently) may also be schematic for athleticism, or give it a lot of thought on a day-to-day basis. Athletic orientation, a latent construct that is composed of a collection of thoughts, motives, and goals that are related to athleticism, would be the third type of measurable trait. Athletic orientation could be measured by importance of exercising, the value that one places on playing sports, and the degree to which a person enjoys playing sports, to name a few. Schemas and traits of this type may be inextricably linked, as the trait contains within it cognitive factors for which some people may be schematic.

Personality theorists have historically endorsed all three concepts; Eysenck conceived of traits as physical properties in the brain that exert a causal force on behavior (1991), the act frequency approach views dispositions as summaries of behaviors with no explanatory power, and as measurable only by a set of behaviors (Buss and Craik, 1983), and the five-factor theory (Costa & McCrae, 1995) treats traits as an umbrella for a set of thoughts, feelings, and behaviors, as evidenced by the items on the NEO-PI-R. This is also the concept of trait that the questionnaires in the current study measured. For example, on
the 6FPQ, participants answered such items as *I sail my own course* (behavior), *I love to be complimented* (feeling), and *I want to be liked* (goal/motivation), and although a bit more nuanced than the 6FPQ, on the HEXACO, participants answered such items as *On most days, I feel cheerful and optimistic* (feeling), *If I want something from someone, I will laugh at that person’s worst jokes* (behavior), and *I worry a lot less than other people do* (thought). One of the main problems in these studies, then, was the measurement of traits; the traits in our study were of the variety that is not able to be cleanly separated from schemas. For future studies, to avoid a conceptual confound, it may behoove us to measure traits as behaviors if possible. If this is not possible, then a different measure of schemas may serve the same purpose.

What would an ideal measure of schemas look like? In the current project, as in Markus’ (1977) experiment, both studies included a measure of personal importance, and even though the hypothesis was supported in study 1, retrospectively I have learned that this may not be the best way to measure a schema. Schemas act as templates that inform our decisions and behaviors, but they are not necessarily value-laden, such as a measure of personal importance suggests. By measuring schemas in this way, we have blurred the already-fuzzy line between schemas and traits. A better measure may be one that assesses the objective nature of knowledge without confounding it with a subjective measure of importance. Such a measure may be one that simply assesses the extent to which a person thinks about the schema in question.

B. Limitations

There were several limitation to these studies. In study 1, the morally-loaded nature of the honesty/humility trait and a social desirability bias may have led more participants to indicate that it was important to them when it may not have been; future studies should include a measure of social desirability. Additionally, the unequal numbers of participants who were schematic or aschematic for the traits was problematic; the reliability coefficient of a 25-person sample and the reliability coefficient
of a 120-person sample are not really comparable, but due to monetary and time constraints, it was necessary to discontinue data collection when the number of participants reached 375. Because these data have provided preliminary evidence that there may be schema-driven differences between individuals’ trait responses, I would like to replicate these findings in the future, with the exception that I would like to collect enough data so that each condition contains 50 people.

The major limitation in study 2, as noted, was the low number of subjects in the aschematic condition. This sample consisted entirely of college students, and for many late-stage adolescents, a time when they struggle to gain independence from their parents, the transition to college may be the pinnacle of an autonomous goal. Independence was very important to many people in the student population, and conversely, the lack of participants who did not place some importance on independence was virtually non-existent. If I were to run this study again with the student population, I would measure more than just one trait. Another limitation may be the age of the participants. Research suggests that attitudes and even core dispositions crystallize as people age, sometimes not until well past adolescence (Markus, 1979; Sears & Funk, 2007). If this is the case, then perhaps some students’ schemas may not be fully developed. However, attitudes and dispositions are structurally and functionally different from schemas; schemas represent stable cognitive structures that are believed to form early in life and remain relatively stable (Markus, 1977). While there is evidence that attitudes crystallize (Sears & Funk, 2007), they are also amenable to persuasion (Petty, Wheeler, & Tormala, 2003), and schemas are not.

Additionally, in study 2, the schema measurement was not easily separable from a trait. By including a measure of personal importance, which is an evaluative judgment and similar to self-report measurements of traits, it was not clear whether the schema that we were measuring was merely another trait measure. Future measurements of schemas should be more objective; for example, the question could simply ask how often the participant thinks of the attribute.
C. Future Directions

This study is the first step in a line of continuing research. For future research, I am interested in further examining the causal influence of schemas by manipulating (as opposed to simply measuring) them (i.e., making them salient or not), and then examining their impact on response consistency. Additionally, I would like to examine schema effects on popular or well-established personality scales with acceptable alphas. I would like to show that although reliability has been established at an acceptable level, this level drops (to below .70) when administered to people who do not possess a schema for the trait that the scale purports to measure. I would also like to explore the relationship of schemas to the reliability of other measures such as political and moral attitudes.

Gender was not recorded for study 1 but it seems as though it might be an important factor to consider for some traits, for example, emotionality. I would not claim that either men or women are more emotional than the other, but it is almost certainly more socially acceptable for a woman to display her emotions than a man, and the enculturation of such a norm could lead women to indicate that emotionality is more central to who they are, or even that they think about it from day to day. In the future, I would like to collect gender information to control for differences in schematics and aschematics.

Hopefully, this study is the first step in uncovering new methods and theory for measuring personality. I hope these finding will make salient the deficiencies not only in methods we use to construct new measurement instruments, but also the deficiencies in some of our approaches to designing studies meant to capture individual variance, and also to analyzing results using questionnaires that compare averages across individuals.

These findings (and future related studies) could have a significant theoretical impact in the psychometric literature as well as among personality, social, and clinical psychology researchers. For instance, one hardship that many researchers face is the disappointment and frustration of a study that
did not prove fruitful and it is conceivable that test reliability is one demon at work. The current conceptualization of reliability could lend explanatory power to previously non-significant findings (and relief to panicked researchers), especially for those interested in personality scale construction.
Appendix A

(O-R) 1. I would be quite bored by a visit to an art gallery.
(C) 2. I plan ahead and organize things, to avoid scrambling at the last minute.
(A) 3. I rarely hold a grudge, even against people who have badly wronged me.
(X) 4. I feel reasonably satisfied with myself overall.
(E) 5. I would feel afraid if I had to travel in bad weather conditions.
(H) 6. I wouldn’t use flattery to get a raise or promotion at work, even if I thought it would succeed.
(O) 7. I’m interested in learning about the history and politics of other countries.
(C) 8. I often push myself very hard when trying to achieve a goal.
(A-R) 9. People sometimes tell me that I am too critical of others.
(X-R) 10. I rarely express my opinions in group meetings.
(E) 11. Sometimes I can’t help worrying about little things.
(H-R) 12. If I knew that I could never get caught, I would be willing to steal a million dollars.
(O) 13. I would enjoy creating a work of art, such as a novel, a song, or a painting.
(C-R) 14. When working on something, I don’t pay much attention to small details.
(A-R) 15. People sometimes tell me that I’m too stubborn.
(X) 16. I prefer jobs that involve active social interaction to those that involve working alone.
(E) 17. When I suffer from a painful experience, I need someone to make me feel comfortable.
(H) 18. Having a lot of money is not especially important to me.
(O-R) 19. I think that paying attention to radical ideas is a waste of time.
(C-R) 20. I make decisions based on the feeling of the moment rather than on careful thought.
(A-R) 21. People think of me as someone who has a quick temper.
(X) 22. On most days, I feel cheerful and optimistic.
(E) 23. I feel like crying when I see other people crying.
(H-R) 24. I think that I am entitled to more respect than the average person is.
(O) 25. If I had the opportunity, I would like to attend a classical music concert.
(C-R) 26. When working, I sometimes have difficulties due to being disorganized.
(A) 27. My attitude toward people who have treated me badly is “forgive and forget”.
(X-R) 28. I feel that I am an unpopular person.
(E) 29. When it comes to physical danger, I am very fearful.
(H-R) 30. If I want something from someone, I will laugh at that person’s worst jokes.
(O-R) 31. I’ve never really enjoyed looking through an encyclopedia.
(C-R) 32. I do only the minimum amount of work needed to get by.
(A) 33. I tend to be lenient in judging other people.
(X) 34. In social situations, I’m usually the one who makes the first move.
(E-R) 35. I worry a lot less than most people do.
(H) 36. I would never accept a bribe, even if it were very large.
(O) 37. People have often told me that I have a good imagination.
(C) 38. I always try to be accurate in my work, even at the expense of time.
(A) 39. I am usually quite flexible in my opinions when people disagree with me.
(X) 40. The first thing that I always do in a new place is to make friends.
(E-R) 41. I can handle difficult situations without needing emotional support from anyone else.
(H-R) 42. I would get a lot of pleasure from owning expensive luxury goods.
(O) 43. I like people who have unconventional views.
(C-R) 44. I make a lot of mistakes because I don’t think before I act.
(A) 45. Most people tend to get angry more quickly than I do.
46. Most people are more upbeat and dynamic than I generally am.
47. I feel strong emotions when someone close to me is going away for a long time.
48. I want people to know that I am an important person of high status.
49. I don’t think of myself as the artistic or creative type.
50. People often call me a perfectionist.
51. Even when people make a lot of mistakes, I rarely say anything negative.
52. I sometimes feel that I am a worthless person.
53. Even in an emergency I wouldn’t feel like panicking.
54. I wouldn’t pretend to like someone just to get that person to do favors for me.
55. I find it boring to discuss philosophy.
56. I prefer to do whatever comes to mind, rather than stick to a plan.
57. When people tell me that I’m wrong, my first reaction is to argue with them.
58. When I’m in a group of people, I’m often the one who speaks on behalf of the group.
59. I remain unemotional even in situations where most people get very sentimental.
60. I’d be tempted to use counterfeit money, if I were sure I could get away with it.
Appendix B

Schematicity Questionnaire

For the following questions, we are interested in your personality. There are no right or wrong answers; please just circle the answer that you think best represents who you are.

Please indicate where you think your personality best falls on the scales below:

1. Independent
   - 1  2  3  4  5  6  7  8  9  10  11
   - Dependent
2. Individual
   - 1  2  3  4  5  6  7  8  9  10  11
   - Conformist
3. Leader
   - 1  2  3  4  5  6  7  8  9  10  11
   - Follower

How important is this trait (dependence/independence) to your self-description?

1  2  3  4  5
Not at all Slightly Moderately Much Very much

How important is this trait (individual/conformist) to your self-description?

1  2  3  4  5
Not at all Slightly Moderately Much Very much

How important is this trait (leader/follower) to your self-description?

1  2  3  4  5
Not at all Slightly Moderately Much Very much
Appendix C
6FPQ Independence Subscale

Instructions: Please answer the following questions about your personality. There are no right or wrong answers; just click the option that best describe you.

1. I don’t care what others think
2. I don’t care about dressing nicely
3. I feel it’s okay that some people don’t like me
4. I love to be complimented (rs)
5. I need the approval of others (rs)
6. I want to be liked (rs)
Appendix D
TCI Dependence Subscale

1. I try to please everyone
2. I follow directions
3. I do what others want me to do
4. I don’t care what others think (rs)
5. I give in to no one (rs)
6. I want to be different from others (rs)
Appendix E
Control Items

1. I try to work out a lot
2. I practice medicine
3. I do many things for my family
4. I don't move often enough
5. I have many children
6. I like to shop for appliances
7. I don't care about saving coupons
8. I feel it's okay that people don't recycle
9. I always behave well
10. I love to go water-skiing
11. I want to be wealthy
12. I need to be social all of the time
CITED LITERATURE


Feldt, L. S. (1969). A test of the hypothesis that Cronbach’s alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*, 363-373.


August 30, 2010

Nicole Mayer, BA
Psychology
1007 W Harrison St
M/C 285
Chicago, IL 60612
Phone: (614) 596-5004 / Fax: (312) 413-4122

RE: Protocol # 2010-0687
“Reliability as Schema-Driven Concept”

Dear Ms. Mayer:

Your Initial Review (Response to Modifications) was reviewed and approved by the Expedited review process on August 30, 2010. You may now begin your research.

Please note the following information about your approved research protocol:

Protocol Approval Period: August 30, 2010 - August 29, 2011
Approved Subject Enrollment #: 1000
Additional Determinations for Research Involving Minors: The Board determined that this research satisfies 45CFR46.404, research not involving greater than minimal risk.
Performance Sites: UIC
Sponsor: None
PAF#: Not Applicable
Research Protocol(s):
  a) Reliability Study Protocol; Version 1; 08/09/2010

Recruitment Material(s):
  a) No recruitment materials will be used

Informed Consent(s):
  a) Waiver of Informed Consent granted under 45 CFR 46.116(d) for this research

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category:
(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

<table>
<thead>
<tr>
<th>Receipt Date</th>
<th>Submission Type</th>
<th>Review Process</th>
<th>Review Date</th>
<th>Review Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/27/2010</td>
<td>Response To Modifications</td>
<td>Expedited</td>
<td>08/30/2010</td>
<td>Approved</td>
</tr>
</tbody>
</table>

Please remember to:

→ Use your research protocol number (2010-0687) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the enclosure, "UIC Investigator Responsibilities, Protection of Human Research Subjects"

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 996-9299. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Marissa Benni-Weis, M.S.
IRB Coordinator, IRB # 2
Office for the Protection of Research Subjects

Enclosure(s):

1. UIC Investigator Responsibilities, Protection of Human Research Subjects

cc: Gary E. Raney, Psychology, M/C 285
Daniel P. Cervone, Psychology, M/C 28
November 19, 2010

Nicole Mayer, BA
Psychology
1007 W Harrison St., M/C 285
Chicago, IL 60607
Phone: (614) 596-5004 / Fax: (312) 413-4122

RE: Protocol # 2010-0898
“Reliability is Only as Good as Your Weakest Schema”

Dear Ms. Mayer:

Your Initial Review (Response To Modifications) was reviewed and approved by Members of IRB #2 by the Expedited review process on November 4, 2010. You may now begin your research.

Please note the following information about your approved research protocol:

**Protocol Approval Period:** November 4, 2010 - November 3, 2011

**Approved Subject Enrollment #:** 300

**Additional Determinations for Research Involving Minors:** The Board determined that this research satisfies 45CFR46.404, research not involving greater than minimal risk. Therefore, in accordance with 45CFR46.408, the IRB determined that only one parent's/legal guardian's permission/signature is needed. Wards of the State may not be enrolled unless the IRB grants specific approval and assures inclusion of additional protections in the research required under 45CFR46.409. If you wish to enroll Wards of the State contact OPRS and refer to the tip sheet.

**Performance Sites:** UIC

**Sponsor:** None

**Research Protocol(s):**

b) Reliability is Only as Good as Your Weakest Schema, Version 1, 09/28/2010

**Recruitment Material(s):**

a) UIC Psychology Department Subject Pool recruitment procedures will be followed.

**Informed Consent(s):**

b) Reliability and Schema Study, Consent form, Version 2, October 26, 2010
c) Reliability and Schema Study, Debrief, version 2, October 26, 2010

**Parental Permission(s):**
A waiver of parental permission has been granted under 45 CFR 46.116(d) and 45 CFR 46.408(c); however, as per UIC Psychology Subject Pool policy, at least one parent must sign the Blanket Parental Permission document prior to the minor subject’s participation in the UIC Psychology Subject Pool.

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category:

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

<table>
<thead>
<tr>
<th>Receipt Date</th>
<th>Submission Type</th>
<th>Review Process</th>
<th>Review Date</th>
<th>Review Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/18/2010</td>
<td>Initial Review</td>
<td>Expedited</td>
<td>10/20/2010</td>
<td>Modifications Required</td>
</tr>
<tr>
<td>11/01/2010</td>
<td>Response To Modifications</td>
<td>Expedited</td>
<td>11/04/2010</td>
<td>Approved</td>
</tr>
</tbody>
</table>

Please remember to:

→ Use your research protocol number (2010-0898) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the enclosure, "UIC Investigator Responsibilities, Protection of Human Research Subjects"

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 355-2939. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Jewell Hamilton, MSW
IRB Coordinator, IRB # 2
Office for the Protection of Research Subjects

Enclosure(s):

2. UIC Investigator Responsibilities, Protection of Human Research Subjects
3. Informed Consent Document(s):
   a) Reliability and Schema Study, Consent form, Version 2, October 26, 2010
   b) Reliability and Schema Study, Debrief, version 2, October 26, 2010

cc: Gary E. Raney, Psychology, M/C 285
    Daniel P. Cervone, Faculty Sponsor, Psychology, M/C 285
Nicole D. Mayer

University of Illinois at Chicago
Department of Psychology (M/C 285)
1007 W. Harrison Street
Chicago, IL 60607-7137
nmayer5@uic.edu
(614) 596-5004

Education

University of Illinois at Chicago 2009 – Present
Major: Social and Personality Psychology
Minor: Clinical Psychology
Advisor: Dan Cervone, Ph.D.

The Ohio State University 2004 – 2007
BA in Psychology with distinction

Honors and Awards

Graduated Summa Cum Laude 2007
3.97 GPA 2007
Dean’s list 2004 – 2007
National Society of Collegiate Scholars 2004 – 2005
Campus Scholar 2005 – 2006
Scarlet & Grey Scholar 2005 – 2006

Research Interests

My research interests include: (a) psychometric theories of personality assessment, including the concept of reliability as an individual difference, (b) neural versus social underpinnings in cross-cultural experienced emotions, (c) the interaction of cognitive and affective regulation and its implications for personality functioning

Research Experience

Research Assistant 2007 – 2009
Stanford University Graduate School of Business, Stanford, CA
Faculty Supervisor: Zak Tormala, Ph.D.
Brian Lowery, Ph.D.

**Research Assistant**
The Ohio State University, Columbus, OH
Faculty Supervisor: Richard Petty, Ph.D.

2006 – 2007

**Teaching Experience**

**Teaching Assistant**
University of Illinois at Chicago
Industrial/Organizational Psychology  2012
Abnormal Psychology  2012
Theories of Personality in Psychology  2012; 2011; 2009
Laboratory in Social Psychology  2011
Introduction to Psychology  2010
Writing in Psychology  2010

**Professional Memberships**

Society for Personality and Social Psychology
Association for Psychological Science
American Psychological Association

**Publications**


**Conference Paper Talks and Invited Symposia**


Presentations

Mayer, N. D. (2012, February). *The concept of reliability: Putting the “psyche” back in psychometric.* Talk given at the University of Illinois Chicago orientation for prospective graduate students, Chicago, IL.

Mayer, N. D. (2011, November). *Power and overconfident decision making: The role of sense of power.* Talk given at the University of Illinois Chicago Social and Personality brown bag meeting, Chicago, IL.

Mayer, N. D. (2011, June). *Carl Rogers’ Theory: Applications, related theoretical conceptions, and contemporary research.* Invited lecture given to University of Illinois Chicago Theories of Personality undergraduate class, Chicago, IL.

Mayer, N. D. (2011, February). *The concept of (retest) reliability: Putting the “psycho” back in psychometric.* Talk given at the University of Illinois Chicago orientation for prospective graduate students, Chicago, IL.


Poster Presentations


Media Mentions

*Thinking vs. Feeling*
Econometrix, New York Times, May 2010

*Psychology PhD Student Asks: Are the Words ‘I Think’ or ‘I Feel’ More Persuasive*
e-Atlas, Chicago, IL, December 2009

*Is ‘Thinking’ or ‘Feeling’ More Persuasive?*
Stanford GSB News, Stanford, CA, October 2009
Research in Progress

Reconceptualizing reliability as from a property of measurement instruments to a property of individuals (Collaborator: Dan Cervone)

Examining cross-cultural and differential experiences of emotions in the United States, India, and Africa (Collaborators: Dan Cervone, Ritu Tripathi, Jacob Ainoo)

Creating personality simulations using novel network analysis methods (Collaborators: Dan Cervone, Gregory Bartoszek, Lara Mercurio)

Professional Service: Poster Review

Poster Reviewer for:
Society for Personality and Social Psychology
13th annual convention, 2012, San Diego, CA

Poster Reviewer for:
Association for Psychological Science
24th annual convention, 2012, Chicago, IL
Category: General

Poster Reviewer for:
Association for Psychological Science
23rd annual convention, 2011, Washington D.C.
Category: General