Two-Week Test-Retest Reliability of the Scale of Accurate Personality Prediction

By

Lisa Dawn Stewart

Bachelor of Arts
Psychology, Theatre
Converse College
2013

Master of Science
Clinical Psychology
Florida Institute of Technology
2016

A doctoral research project
submitted to the School of Psychology at
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Psychology
in
Clinical Psychology

Melbourne, FL
December 2017
We the undersigned committee hereby approve the attached doctoral research project.

Two-Week Test-Retest Reliability of the Scale of Accurate Personality Prediction

By

Lisa Dawn Stewart

________________________  ____________________
Philip D. Farber, Ph.D.    Richard T. Elmore, Jr., Ph.D.
Associate Professor      Associate Professor
School of Psychology     School of Psychology
Committee Chair          Committee Member

________________________  ____________________
John Frongillo, Ph.D.     Mary Beth Kenkel, Ph.D.
Assistant Professor       Professor and Dean
School of Arts and Comm.  School of Psychology
  unication               Committee Member
Abstract

TITLE: Two-Week Test-Retest Reliability of the Scale of Accurate Personality Prediction

AUTHOR: Lisa Dawn Stewart, M.S.

MAJOR ADVISOR: Philip D. Farber, Ph.D.

The Scale of Accurate Personality Prediction (SAPP) is a measure derived from the Sixteen Personality Factor Questionnaire (16PF Fifth Edition) with the objective of assessing one’s level of self-knowledge. The SAPP measures self-knowledge by appraising the extent to which individuals are able to accurately predict their personality traits. Previous research on the SAPP has provided support for its construct validation. The present study served as a replication of prior reliability studies to further assess the SAPP’s temporal stability. Test-retest reliability data of the SAPP was obtained by comparing SAPP scores from 34 individuals across two testing sessions that were two weeks apart. Statistical analyses using Pearson’s correlation coefficient were conducted, which revealed a significant moderate correlation between the two derived SAPP scores ($r^2 = .584$, $p<.01$). Implications and limitations of this study, as well as suggestions for future research are discussed.
Table of Contents

Abstract ........................................................................................................................................ iii
List of Tables ................................................................................................................................. v
Acknowledgements ..................................................................................................................... vi
Literature Review ....................................................................................................................... 1
Statement of Purpose and Hypothesis ....................................................................................... 32
Method ......................................................................................................................................... 33
Results ......................................................................................................................................... 36
Discussion .................................................................................................................................... 39
References ..................................................................................................................................... 44
Appendices ................................................................................................................................. 51
List of Tables

Table 1: Demographic Percentages..........................................................48

Table 2: Test-Retest Reliability Data.......................................................49

Table 3: SAPP Ranges, Means, and Standard Deviations..........................50
Acknowledgements

I am fortunate to have such a wonderful support system, with many people to offer my thanks. First, I would like to express my sincere gratitude for Dr. Philip Farber for his continuous support on this project and through my other graduate school endeavors. Thank you for all of the wisdom and laughter you have shared with me over the past few years. You have taught me a great deal about psychology, but you have taught me even more about life. I am thankful to call you my mentor.

Additionally, I am thankful to Dr. Elmore and Dr. Frongillo for their time serving on my committee and their thoughtful feedback on this project. I would also like to thank my friends in the graduate program for making these four challenging years more enjoyable than I ever could have imagined. To Jared and Taylor, I am grateful to have collaborated with you both on SAPP research, and for your patience with the bombardment of text messages with logistical questions you received from me.

I would also like to thank my family for their unconditional love and guidance, and for teaching me how to properly use a comma. To Ashley, thank you for keeping me sane through all graduate school milestones, and for bringing my life so much joy. My appreciation also extends to my best friend of nearly fifteen years, Allison.

Thanks for always uplifting me with words of encouragement or a thoughtful “surcee.” Last but not least, I would like to thank “the Gracies” for their loyalty, and for empowering one another to use our talents as active leaders for change.
Psychological Perspectives of the Self

William James was among the first psychologists in the nineteenth century to distinguish the “I” from the “Me” in regard to the self. He posited that the I-Self, or self-as-subject, referred to the active agent that has a sense of self-awareness. The Me-Self refers to the subject that is observed by other people or by the I-Self. James’ conceptualization of the self proposed the notion that the self is multifaceted. Sigmund Freud expanded upon this construct to posit that people have a multitude of “selves” that are manifested differently, depending on their contexts. Most notable psychologists of the first half of the twentieth century, including Karen Horney, Carl Jung, Abraham Maslow, and Carl Rogers rejected this idea. Rather, they were in support of the idea of a singular and stable self (Broderick & Blewitt, 2014).

Leary and Tagney (2003) outlined five different ways in which the “self” has been conceptualized by behavioral and social scientists. First, they note the tendency for the “self” to be equated with the entire “person”; i.e., as synonymous terms. They assert that this is problematic from a psychological standpoint, as a person is not typically thought of as being a self; rather, a person is thought of as having a self. Secondly, Leary and Tagney (2003) report that the self has also been conceptualized as referencing a person’s personality or personality traits. However,
this conceptualization is also precarious, as it questions whether personality researchers are studying the self or a separate concept. Leary and Tagney (2003) reconcile this theory by asserting that while personality and the self are separate constructs, the self is pertinent to comprehending facets of one’s personality.

Leary and Tagney’s (2003) next two conceptualizations (i.e., their third and fourth categories) are seen as extensions of James’ definitions of the self as the experiencing subject (I-Self,) and the observed self (Me-Self). The self-as-subject (I) has been likened to the inner psychological being that centralizes a person’s experiences. Therefore, the self-as-subject is thought to be the observer that perceives, processes, and organizes perceived stimuli to evoke emotion and guide behavior. In contrast, the self-as-object (me) refers to one’s ability to be evaluated by others and by the self-as-subject. For instance, the self-concept, self-schema, and self-beliefs are often produced by the self-as-subject’s reflection and intent to describe the self-as-object. The fifth common conceptualization of the self includes the self as executive agent. That is, the self is thought to be at the core of decision-making and response implementation/inhibition processes. Thus, terms such as self-control and self-regulation reflect the role of the “self” in higher order executive functions (Leary & Tagney, 2003). These various conceptualizations offer a glimpse into the amount of interest and difficulty that has arisen in efforts to objectively conceptualize and clarify the “self.”

Several advancements within the fields of psychology and sociology in the second half of the twentieth century led to an increased emphasis on the self. This
emphasis resulted in research done on self-esteem in the 1950s and 1960s, the emergence of cognitive psychology with focus on internal thought processes, and the publication of psychological measures pertaining to self-attributes in the 1960s and 1970s (Leary & Tangney, 2003). The self currently remains a central topic within psychological domains and has implications for personality theory and assessment.

Cultural Variations of the Self

The notion of “the self” is a construct that is deeply intertwined within the realm of psychology. Common psychological terms such as self-esteem, self-worth, self-efficacy, self-concept, self-awareness, self-actualization, and self-knowledge exemplify the relevance the self has in the consideration of psychological functioning. Considering its significance, a comprehensive conceptualization of the self is crucial to understanding its facets and applications. Mainstream psychological methodology is primarily rooted in Western ideology (Yang, 2012). As our world continues to diversify, multicultural approaches to psychological conceptualization and treatment is paramount for accurate and global understanding of human cognition and behavior. Thus, it is necessary to note cultural variations in defining the self when seeking to evaluate how well people understand themselves.

The term for how people perceive themselves compared to others is known as their self-construal (Ross & Murdock, 2014). Self-construals differ largely
depending on one’s cultural environment. Those in Western/individualistic cultures are more likely to hold independent self-construals, while those in Eastern/collectivistic cultures are more likely to have self-construals that are interdependent. Those with independent self-construals have a singular, fixed, and stable sense of self that does not vary depending on social context. Primary elements of independent self-construals include personal attributes, skills, and predilections. In contrast, those with interdependent self-construals have a sense of self that is directly linked to social context. Thus, those with interdependent self-construals have a sense of self that is more malleable and variable than those with independent self-construals. Interpersonal relationships with specific others and with in-groups are central features of interdependent self-construals (Kiuchi, 2006).

Ross and Murdock (2014) assert that those with interdependent self-construals still maintain a separate sense of self but posit that it is one that includes interpersonal relationships as being a component of the self. Hence, their sense of self is largely dependent on situational variables, reflecting their values for harmony and societal expectations over individual desires (Ross & Murdock, 2014). These values are exemplified in many Eastern religious and philosophical ideologies. For instance, according to the tenets of Confucianism, individuality is regarded as a self-centered quality and deemed only necessary in regard to moral judgment and scholastic endeavors. Those with interdependent construals emphasize effort as opposed to individual abilities. Thus, failure is perceived as the result of inadequate effort. Success, however, is not solely attributable to effort but
to the assistance of others as well. This further reflects the importance of social relationships and in-group membership (Harter, 2012).

Contrary to their interdependent counterpart, those with individual self-construals are likely to credit their successes to personal abilities, while attributing their failures to external variables (Harter, 2012). Rather than viewing interpersonal relationships as being an element of the self, those with individual self-construals emphasize their differentiation from others. Individual preferences, beliefs, and characteristics are integral to independent self-construals (Hannover, Birkner, & Pohlmann, 2006). As to be expected, classic Western psychology mirrors the independent self-construal’s notion of a self that is stable and able to be examined over time (Broderick & Blewitt, 2014). However, independent and interdependent self-construals are not rigidly bound to culture. Therefore, high levels of each type of self-construals are found in people of both individualistic and collectivistic societies (Ross & Murdock, 2014). This indicates that both types of self-construals have significance in the way psychologists seek to conceptualize the self. The integration of cultural considerations into the operationalization of the self is essential yet complex. Kiuchi (2006) inquired about the manner in which one develops a self-concept that is both somewhat consistent and stable yet flexible and contextually adaptive. Thus, the challenge of cohesively reconciling these two types of self-construals remains. For the purposes of this research effort, the more independent/Western self-construal will serve as the focus of the self, along with one of its more important elements, self-knowledge.
Defining personality has proven to be as difficult as defining the self. Indeed, Engler (2003) contends that most theorists of personality have differing definitions of personality, based on each of their unique perceptions of the term. One such theorist, Gordon Allport, detailed more than fifty definitions (Engler, 2003). He also established “trait” as the fundamental concept pertaining to personality, which impacted the way it would be assessed. Furthermore, the notion of personality traits influenced the perception of a stable, unchanging personality regardless of context (Winter, John, Stewart, Klohnen, & Duncan, 1998). This illustrates the parallel among conceptualizations of the self and personality.

Indeed, some theorists have used the two terms interchangeably (Leary & Tangney, 2003). Leary and Tangney (2003) proposed that it is better to differentiate the terms. They asserted that personality can be delineated as the sum of facets that make one “psychologically unique” and note the relevance the concept of the self has in relation to comprehending elements of personality (Leary & Tangney, 2003).

Freud played a vital role in the inception of a formal personality theory. According to Freud, the structure of one’s personality is comprised of a blend of unconscious and conscious variables; namely, the id, ego, and superego. He conceded that each component of the personality serves a function to motivate one’s behavior to satisfy his or her numerous need states (Engler, 2003).

Influenced by Freud’s theory, Carl Jung also conjectured that the structure of the
personality includes both unconscious and conscious factors. Unlike Freud, who held a more deterministic viewpoint in regard to behavior, Jung posited that potential future endeavors also motivate behavior (Hall & Lindzey, 1957).

The emergence of the behaviorism movement at the turn of the twentieth century within psychology brought about a significant contrast in the way personality would be conceptualized as compared to the origins of personality theory. One of the pioneers of behaviorism, B.F. Skinner, purported that it was not practical to attempt to comprehend people’s behaviors by studying unobservable internal factors. Instead, he asserted that objective observations and analysis of organisms’ environments were essential to understanding and modifying behavior. Skinner found it unnecessary to attribute causality or internal motivational states to behavior. Rather, Skinner was interested in factors that could be precisely manipulated, such as antecedents and consequences of behavior. The concept of personality did not cease to exist under behaviorist theories. Rather, personality, as defined largely by internal characteristics, was regarded as superfluous within a scientific domain (Engler, 2003).

In an effort to expand the restricted concepts of human nature offered by classical psychoanalysis and radical behaviorism, more humanistic theories emerged within the psychological field. Abraham Maslow’s theory did not directly contradict contributions from psychoanalysis or behaviorism. Instead, his theory complemented these movements to include a concept of human nature that accentuates positive human attributes. In a similar manner, Carl Rogers’ theory of
personality has as its prominent notion that the self is, and emerges as, an integral component of human experience. Thanks to Rogers’ contributions, the self reemerged in the mid twentieth century as a practical consideration in the comprehension of personality. Rogers’ theory is influenced by phenomenology, an approach that focuses on how something is perceived, rather than what elements constitute the object or the experience itself. Rogers postulated that the individual holds the best perspective for understanding himself or herself. Rogers elaborated on this notion of the self to describe one’s self-concept as a differentiation from the sum of one’s experiences to encompass people’s perception of themselves. One’s perceived self can sometimes vary from one’s objective experience, resulting in what Rogers’ calls incongruence. Rogers asserts that people have a need for their perceived self to be congruent with their objective experiences (Engler, 2003). One’s ability to accurately differentiate his or her perceived self from their “real” self is likened to one’s breadth of self-knowledge.

The existential movement within psychology continued to expand upon the reemergence of the self in personality theory. Existentialist Rollo May challenged the notion that scientific examination and the concept of the self are mutually exclusive. Rather, he proposed that discounting the self as a feature of humanity results in the loss of an important variable of a person’s psychological functioning. Before the existential movement, Western psychology focused on the supposed fixed laws and guidelines that were believed to govern human behavior, including conditioned responses, motivational drives, and unconscious forces. Existentialists
have defined these fixed principles as being the essence of beings. However, they offered an alternative approach to understanding people that rejected reducing human beings to their components (i.e. essences). Instead, existentialism has asserted that the study of the entirety of one’s existence was more paramount to the more complete understanding of the self, as it prevented the objectification of the human experience. Existentialists, in general, believe that the veracity of the self is that it exists within each individual’s unique perception of him or herself (Engler, 2003).

*Self-Knowledge*

The examination of self-knowledge has primarily emphasized people’s accuracy at assessing and evaluating their internal conditions. This includes personality characteristics, emotional states, and viewpoints. Considering the dynamic nature of humans, another component of self-knowledge includes memory processes, such as the ability to recall and evaluate past emotional experiences, and then to predict future emotional responses. Thus, self-knowledge includes understanding of the past self and future self as well as the present self. Self-knowledge corresponds to consciousness. That is, the assertion that one possesses self-knowledge suggests that he or she is conscious of internal states and processes. Therefore, conditions not within the realm of conscious awareness thwart the capacity for self-knowledge. For instance, people can knowingly or unknowingly inhibit or forget undesirable emotions and/or thoughts (Wilson & Dunn, 2004).
Such limitations indicate the importance of consciousness in the acquisition of self-knowledge. Indeed, Vazire & Wilson (2012) purport that individuals possess “blind spots” in their knowledge of self that are the result of limitations of awareness.

The multidimensional nature of the self also complicates the attainment of self-knowledge. For instance, several components of the self, such as self-narratives and self-identification, may conflict with principles of knowledge. That is, individuals are generally regarded as the authority of truth of their respective lives. Vazire & Wilson (2012) assert that knowledge requires belief in the truth and justification of a premise. However, cognitive distortions and biases can alter people’s objectivity. Therefore, the fallibility of human perception can impede the acquisition of self-knowledge. Nevertheless, achievement of self-knowledge has utility across various domains of human functioning (Vazire & Wilson, 2012).

Greater levels of self-knowledge are predicted to correspond with behaviors that are more consistent with internal states and values. Comprehension of the self allows for individuals to more accurately evaluate their responses to experiences. Furthermore, self-knowledge likely benefits interpersonal relationships as it allows for people to understand others’ interactions with them (Vazire & Wilson, 2012). Therefore, awareness of how much self-knowledge an individual possesses may be useful in reducing conscious “blind spots” and thus increasing one’s self-knowledge as it allows for the opportunity of self-discovery.
Considering the similarities and overlap between the constructs of self and personality, it can be postulated that personality assessment can have inferences for the study of the self and self-knowledge. Objective assessments of personality provide a quantitative analysis of measurable personal traits. Results from objective measures of personality are useful for gaining a comprehensive understanding of someone’s psychological functioning. Bearing in mind the barriers to one’s appraisal of their self-knowledge, it can be presumed that people may be aware of certain traits they possess and incognizant of other characteristics. Therefore, objective personality assessments offer the opportunity to examine how well individuals know themselves. This implies that the usefulness of objective measures of personality expands beyond the study of the self to yield germane information regarding one’s self-knowledge.

What follows at this point will be first a review of the various attempts within the psychological literature to measure the construct of self-knowledge, and subsequent to that review, an in-depth look at a recently developed scale presumed to measure the accuracy of personality self-prediction utilizing the Sixteen Personality Factor Questionnaire (16PF).

Attempts to Measure Self-Knowledge

Several self-report measures have been developed and marketed as assessments of self-knowledge. For instance, in 1975 Fenigstein, Scheier, & Buss (cited in Blankemeier, 2007) developed the Self Consciousness Scale (SCS). This
23-item measure contains a “Private self-conscious” factor that is deemed similar to self-knowledge in that it refers to routine reflection and attendance to internal experiences. Those who score high on this factor have been correlated with more accurate self-knowledge. However, the difficulty with the SCS and other similar self-report measures is that they are not direct measures of self-knowledge, but instead assessments of similar constructs or correlates with self-knowledge (Blankemeier, 2007). A measure developed by Alschuler, Weinstein, Tamashiro, and Smith (1977) (as cited in Elghossein, 2012) involved applying an intricate coding system to responses given by subjects regarding affective states in relation to past experiences, which was based on the presumption that self-knowledge is an interaction of the summation of one’s life experiences with their ability to process them. Sedikides (1993) developed a measure that included asking subjects about personality characteristics they believed themselves to have.

Miller (2000) reviewed several studies that were conducted to examine the relationship between subjects’ prediction of scores on a personality measure with their obtained score. For instance, Furnham and Varian (1988) (as cited in Miller, 2000) administered the Eysenck Personality Inventory and found that subjects were able to accurately self-predict their scores on constructs which they were familiar with, such as neuroticism and extroversion. Miller (2000) also cited studies that utilized the 16PF to examine accuracy of self-reports in comparison to obtained test scores (e.g. Friedman, Sasek, & Wakefield, 1976; DeBlassie & Franco, 1983; Furnham, 1989). Miller (2000) noted that none of these previous efforts
emphasized individual personality factors that may influence accuracy of the ability to self-predict. To date, the SAPP is the first measure of individual’s self-knowledge of personality and psychological traits (Mazur, 2015).

*The 16PF*

Raymond B. Cattell developed the 16PF in 1945, using multiple factor analyses to objectively measure personality traits. Since it was first published in 1949, the 16PF has been revised four times. The most current version of the 16PF, the Fifth Edition, was developed in 1988. The 16PF contains 185 items that correspond to 16 primary factors and five global factors. Three validity scales are also included to determine the respondent’s attitude toward test-taking to rule-out the possibility of random, biased, or other types of responding that may invalidate the profile (Karson, Karson, & O’Dell, 1997).

The 16 primary factors are displayed on a respondent’s profile so that the scales that have the most impact on people’s behavior and personality are listed first; that is, the ones first identified through the series of factor analyses (Cattell, 1989) (see Appendix D for a blank 16PF profile sheet).

The first of these factors is Factor A: Warmth. Warmth is a measure of interpersonal amiability, agreeableness, and willingness to engage in social interactions. High scores on Factor A are indicative of people who have a large affinity for others and are supportive, charitable, and well-liked by others. Extremely high scores may indicate markedly high need for social contact to the
extent that the needs of others are prioritized above the needs of the individual. Those with low scores on Factor A are typically partial to solitude and are not likely to be as attentive to the needs of others. They are more likely than their high A counterparts to lack concern about the implications their behaviors have on others. Extremely low scores denote tendencies for social withdrawal and significant deficits in need for interpersonal contact (Karson, Karson, & O’Dell, 1997).

The second factor, Factor B: Reasoning, is an approximate measure of intelligence, in this context defined as, “the capacity to discern relationships in terms of how things stand, relative to one another” (Cattell, 1989, p. 30). Specific abilities measured by this factor include, “recognizing analogies and similarities, and being able to classify events and form typologies” (Cattell, 1989, p.30). Higher scores on Factor B are indicative of abstract thinking patterns while lower scores allude to the presence of more concrete thinking patterns. Factor B was not intended to be interpreted as a measure of intellectual capacity, per se, but rather as a means of discerning an individual’s approach to solving problems in their daily life (Karson, Karson, & O’Dell, 1997).

Factor C: Emotional Stability is the third of the 16 primary factors. Factor C corresponds to an individual’s ability to cope with external and internal stressors. Low scores on Factor C are indicative of individuals who are reactive and emotionally changeable in response to stressors. Therefore, those with low scores typically demonstrate less effective and adequate coping and are thus more prone to
adjustment difficulties. Conversely, people with high scores on scale C are mature, adaptive, emotionally stable, and overall more adjusted. However, extremely high scores may reflect denial of difficulties (Karson, Karson, & O’Dell, 1997).

The fourth of the 16 primary factors is Factor E: Dominance. Karson, Karson, & O’Dell (1997) assert that this factor might better be described as a measure of assertiveness rather than dominance. The reasoning for this is the scale measures styles of self-expression rather than styles of interpersonal control. However, Cattell (1989) states that average scores on this scale indicate that respondents have the capacity to appropriately assert themselves as a means of self-protection and setting healthy boundaries within interactions. She distinguishes dominance as a construct that infers aggression rather than self-defense. Nonetheless, those who score high exhibit a proclivity to engage in forthright and candid interpersonal interactions. They are more willing to express themselves freely and typically prefer an authority-based approach to their environment. Low scorers on this factor are more deferential, submissive, and typically avoid conflict and/or confrontation (Karson, Karson, & O’Dell, 1997).

Factor F: Liveliness constitutes the fifth primary factor of the 16PF and is a measure of exuberance (Cattell, 1989). Low scores indicate restrained, cautious, and serious approaches to engagement with the environment, while high scores indicate more lively, animated, and spontaneous attitudes. It is hypothesized that parental punishment styles influence respondents’ approach to engaging with the
environment. Those who are more precautious and restrained are presumed to have experienced excessive punishment as children (Karson, Karson, & O’Dell, 1997).

The sixth primary factor, Factor G: Rule-Consciousness refers to the extent that individuals are accustomed to conforming to societal and conventional standards. Those with low scores on this scale are more expedient and non-conforming. They are more likely to question the rationale behind rules and regulations and let their experiences guide their principles and/or behavior. Those with lower scores may be less concerned with impression management than their high-scoring counterparts. People with high scores on Factor G are more conscious of established guidelines of behavior and may consider themselves especially moralistic (Karson, Karson, & O’Dell, 1997).

Social Boldness, known as Factor H, is the seventh primary factor and is a measure of people’s disposition toward risk taking. The polar states of this scale are correlated with physiological responses. Those who have high scores demonstrate less reactivity to stress. This indicates that their parasympathetic nervous system dominates their stress response. High scoring individuals are therefore more likely to be venturesome, sensation-seeking, and risk-takers. Conversely, those with low scores typically have a dominant sympathetic nervous system, making them more likely to exhibit heightened reactivity to stressors. Therefore, low scorers are more likely to be inhibited when it comes to risk-taking and are often shy, timid, and sensitive to impending threats (Cattell, 1989).
The eighth primary factor is Factor I: Sensitivity. Cattell (1989) likens this scale to Jung’s notion of a polarity between “thinking” and “feeling.” Those with low scores align with the “thinking” category and have a propensity to be objective, utilitarian, and unsentimental (Cattell, 1989; Karson, Karson, & O’Dell, 1997). These individuals are more likely to discredit or remain oblivious to internal emotional stimuli when making evaluations than their high scoring counterparts. Those with high scores are more likely to interpret their environment based on internal emotional experiences, individual value systems, aesthetics, and subjective intuitions (Cattell, 1989).

Factor L: Vigilance is a measure of people’s trust or amount of suspicion in others. High scorers on this ninth primary factor are inclined to be suspicious, skeptical, guarded, and vigilant in their interpersonal interactions. Very high scores have been found to be one of the greatest markers of psychopathology that can be determined from the 16PF. Indeed, extremely high scores are more indicative of paranoia than of suspiciousness. Criteria for paranoia include, the assertion that, “the forces about which suspicions are harbored must be projections of some aspect of the self” (Karson, Karson, & O’Dell, 1997, p.50). Thus, paranoia is rooted in people’s tendency to attribute their own hostile interpretations to others. Mid-range scores are associated with people’s anger and suggest the degree to which they carry resentments as well as the rate at which they are likely to become angry. Low scores indicate one who is trusting and unsuspecting of others (Karson, Karson, & O’Dell, 1997).
Abstractedness is the construct measured in Factor M, the tenth primary factor. Karson, Karson, & O’Dell (1997) liken the scale to a measure of impracticality. Those with high scores have a propensity to be absent-minded, imaginative, and engrossed with ideas. High scorers are designated as more impractical. Their low scoring counterparts, are inclined to practicality and are solution-oriented and grounded. Low scorers demonstrate more focus on their immediate environment (Cattell, 1989; Karson, Karson, & O’Dell, 1997).

The eleventh primary factor, Factor N: Privateness refers to the degree of an individual’s propensity for self-disclosure. Those with low scores are more likely to be more forthright and likely to disclose personal experiences. Very low scores may be indicative of immaturity, self-centeredness, or hypomania. Very low scorers lack tact and discretion when divulging personal details. Those with high scores are less likely to disclose personal information and are more private and discreet (Karson, Karson, & O’Dell, 1997). High scores may manifest as a manipulative social mask in which people do not show their true feelings or motives for personal gain. Conversely, high scores may indicate tendencies for people to astutely observe surroundings and to judiciously select details they wish to disclose to others (Cattell, 1989).

The twelfth of the primary factors is Factor O: Apprehension. This factor is an assessment of a person’s feelings of self-worth (Cattell, 1989). People with low scores are complacent, self-assured, and untroubled. They are less likely to experience guilt than their high-scoring counterparts. Those with high scores have a
proclivity to experience self-doubt, apprehension, and worry (Karson, Karson, & O’Dell, 1997). High scores can be adaptive in that apprehensive tendencies may facilitate preparations for future endeavors (Russell & Karol, 1994). However, high scores can also lend to maladaptive characteristics such as low self-esteem (Karson, Karson, & O’Dell, 1997).

The remaining four factors are derived exclusively from questionnaire data rather than a combination of behavioral ratings, objective testing ratings, and questionnaire data. Hence, the letter “Q” serves to denote these scales. Factor Q1 refers to Openness to Change (Karson, Karson, & O’Dell, 1997). When interpreting this scale, “the higher the score, the more oriented the person is toward trying something new; the lower the score, the more directed the person is toward making do or savoring what is available” (Karson, Karson, & O’Dell, 1997, p. 56). Very low scores are indicators of problematic characteristics, such as inflexibility and adjustment difficulties. This scale is also related to the concept of fulfillment, where those with higher scores may desire change due to dissatisfaction with their current circumstances. Conversely, those who dislike change may be more content with their present conditions (Karson, Karson, & O’Dell, 1997).

Factor Q2: Self-Reliance refers to one’s propensity to engage in tasks with others versus doing them alone. The lower the score the more group-oriented and affiliated individuals are, and the higher the score the more a solitary and individualistic nature is suggested (Karson, Karson, & O’Dell, 1997). Karson, Karson, & O’Dell (1997) note the tendency for response patterns on this scale to be
indicative of people’s vocational preferences. Clinically, very high scores suggest that respondents may be interpersonally avoidant rather than demonstrating a penchant for solitude. Conversely, very low scores may signify dependent tendencies (Karson, Karson, & O’Dell, 1997).

Perfectionism, or compulsivity, is the construct measured in Factor Q3. Those on the higher end of the scale are more organized, self-disciplined, and perfectionistic. The lower the score the more respondents exhibit a tolerance for disorder and flexibility. Extremely high scores indicate intolerance for ambiguity and lack of adaptive flexibility. On the other hand, very low scoring individuals are more likely to be disorganized and lacking in efficiency (Karson, Karson, & O’Dell, 1997).

The last of the sixteen primary factors is Factor Q4: This factor measures “the unpleasant sensations that accompany autonomic arousal, colloquially referred to as ‘nervous tension’” (Cattell, 1989, p.294). Experiences of tension are conducive to the development of anxiety. Those with high scores are energetic and tense. Furthermore, they are often impatient and driven individuals. Those with lower scores are more relaxed, patient, and imperturbable. High scores on Q4 can also be used as an indicator of heightened subjective distress (Karson, Karson, & O’Dell, 1997).

The five global factors of the 16PF refer to the five largest second-order factors that result from factor analyses of the sixteen primary factors (Cattell, 1989). The first of the global factors is Extraversion (EX). The higher the score on
this scale signifies the more extraverted and actively social a person is. Therefore, the lower end of the range of scores is indicative of introverted and socially inhibited individuals. Primary factors that load onto this scale include warmth (A+), candIDness (N-), liveliness (F+), group orientation (Q2), and social boldness (H+). These factors are theorized to modify each other. For instance, A+ individuals are apt to receive social reinforcement and are thus less disposed to be shy and excessively discreet and are more likely to be enthusiastic. The correlation between these traits suggests an extraverted nature (Karson, Karson, & O’Dell, 1997).

The second global factor is that of Anxiety (AX). Interpretation of this factor has direct clinical value, as high scores are more indicative of possible psychopathology. A high score “either means the person is overwhelmed or that he or she is acknowledging a lot of problems” (Karson, Karson, & O’Dell, 1997, p. 73). The primary factors that load onto this scale are suggestive of typical anxiety symptomatology, such as dysphoric emotionality and negative thought patterns (Cattell, 1989). There are many similarities between AX and Ego Strength (C). However, the former contains a considerably greater number of items. Other factors that load onto this scale, aside from low ego strength (C-), include vigilance (L+), tension (Q4+), and apprehension (O+) (Karson, Karson, & O’Dell, 1997). When discerning whether or not a high AX score conveys a stable trait or reactive state, the presence of low self-sentiment (Q3-) in conjunction with low ego strength (C-).
is redolent of trait anxiety. Both of these primary traits are conjectured to be unchanging to differing mood states (Cattell, 1989).

The third global factor is Tough-Mindedness (TM). Originally, this scale was named “Cortertia,” a derivative of “cortical alertness” (Karson, Karson, & O’Dell, 1997). Those with high scores are postulated to have greater cortical acuity in response to their environment, yielding self-reliant, realistic, and tough-minded ways of processing information. In contrast, low scorers are attuned to subjective emotional responses and are more likely to contemplate whims, nuances, and ambiguity in response to environmental stimuli (Cattell, 1989). Primary factors that are associated with high scores are reservation (A-), practicality (M-), lack of sensitivity (I-), and inflexibility (Q1-). The inverse of these scales would yield a low score on this factor (Karson, Karson, & O’Dell, 1997).

Independence (IN) marks the fourth global factor and is a measure of an individual’s self-determination (Russell & Karol, 1994). Those on the high end of the scale are dominant (E+), socially bold (H+), vigilant (L+), and open to change (Q1+) (Karson, Karson, & O’Dell, 1997). Such people are inclined to base their behaviors on internal guidelines rather than merely acquiescing to social confines. If their actions are concordant with prevailing social norms it is likely that the conclusion was made on personal standards rather than conventional conformity. Low scores on each aforementioned scale would consequently result in a low score on this factor. Lower scores suggest individuals who are subdued and submissive to conventional expectations. Their submissive qualities suggest a need for external
guidelines and social support. These individuals require more structure in their life and are more resistant to change (Cattell, 1989).

The final global factor is that of Self-Control (SC). This factor examines the extent to which respondents inhibit their urges to regard rules and consequences of their actions. High scorers are thus more controlled individuals who exhibit rule-consciousness (G+), practicality (M-), compulsivity (Q3+), and low levels of liveliness (F-)(Karson, Karson, & O’Dell, 1997). People with high self-control typically exhibit adherence to conventional standards and values as a function to maintain social desirability (Cattell, 1989). Those who demonstrate less self-control are likely livelier, thus requiring more effort to control (Karson, Karson, & O’Dell, 1997). Low scorers are more unrestrained in regard to acting on urges and may demonstrate unconventional morals, idiosyncratic ideals, or a general indifference to personal values (Cattell, 1989).

The three validity scales of the 16PF include the Impression Management (IM) scale, the Acquiescence (ACQ) scale, and the Infrequency (INF) scale. The IM scale examines the extent to which a respondent may be answering items in an effort to look more or less socially desirable than could be reasonably accepted. The scale includes defensive items, laudatory items, and items assessing for conventionally accepted behaviors. Very high scores are indicative of a response pattern typical for those who are determined to present themselves in an overly positive light. On the other hand, very low scores indicate efforts to present oneself in a negative light and/or difficulties with self-esteem and interpersonal
interactions. The ACQ scale measures the respondent’s inclination to concur with item statements irrespective of their content. Profiles are not suitable for interpretation should the ACQ score be too high or low, as this indicates random responding or reduced comprehension of test items. The purpose of the INF also detects random responding and poor understanding of test items, as well as universal abnormality in responses. The INF scale is comprised of item responses that are “least frequently chosen by the norm sample” (Karson, Karson, & O’Dell, 1997, p. 64).

Interpretations of the 16PF have various applications. In addition to a psychological evaluation report, tailored interpretive reports exist for couple’s counseling, career development, competency, and so forth. The Scale of Accurate Personality Prediction (SAPP) expands upon the 16PF’s utility to provide a measure of an individual’s self-knowledge.

*Development of the SAPP*

Miller (2000) developed the SAPP using data collected through convenience sampling on the college population and members of the local community. Subjects were first administered a standard 16PF. Subsequently, the subjects were provided with a blank 16PF profile form (see Appendix D) and asked to indicate where they believed they would score on each factor. Then, each subject’s predicted score was compared to his/her actual obtained score. Using the comparison of scores, Miller derived the SAPP, which is a total score of the degree
of accuracy of self-prediction. The formula Miller developed for the SAPP is as follows:

\[
\text{SAPP} = [\text{OSA-PSA}] + [\text{OSB-PSB}] + [\text{OSC-PSC}] + [\text{OSE-PSE}] + [\text{OSF-PSF}] + [\text{OSG-PSG}] + [\text{OSH-PSH}] + [\text{OSI-PSI}] + [\text{OSL-PSL}] + [\text{OSM-PSM}] + [\text{OSN-PSN}] + [\text{OSO-PSO}] + [\text{OSQ1-PSQ1}] + [\text{OSQ2-PSQ2}] + [\text{OSQ3-PSQ3}] + [\text{OSQ4-PSQ4}] + [\text{OSEX-PSEX}] + [\text{OSAX-PSAX}] + [\text{OSTM-PSTM}] + [\text{OSIN-PSIN}] + [\text{OSSC-PSSC}]
\]

were OSA references the subject’s Obtained Score on Scale A, PSA reflects that subject’s Predicted Score on Scale A, and so on for the remaining 15 primary factors and the 5 global factors. Parentheses around the subtracted scores are meant to indicate the absolute value of the subtracted scores.

Interpretation of the SAPP score is quite simple. The lower the score, the more accuracy in self-prediction one possesses. Higher scores therefore indicate diminished ability to accurately self-predict. The highest score, and therefore poorest indicated degree of self-prediction is a 189. The lowest score possible is a 0 and indicates the highest and most exceptional degree of self-predictive accuracy (Miller, 2000).

In addition to assessing an individual’s accuracy in self-prediction, Miller was also interested in determining which of the 16PF scales were the best indicators of people’s ability to accurately predict personality traits. Therefore, Miller conducted a regression analysis and found that the global factor of Tough-
Mindedness, in the negative direction, was the highest indicator of participants’ ability to predict their score on the SAPP. Miller found that the low scoring and high scoring groups among participants had similarities. That is, those with lower scores demonstrated a profile pattern of A+, B+, F+, I+, L-, N-, N-, Q1+, Extraversion (+), and Tough-Mindedness (-). Therefore, the participants who were able to more accurately predict their personality traits were found to be warm, abstract, lively, sensitive, trusting, forthright, open to change, outgoing, and intuitive. In other words, better self-prediction was found in those with lower tough-minded scores. Those in the high scoring group that were less able to predict their personality traits were described as reserved, concrete, restrained, unsentimental, interpersonally wary, discreet, traditional, introverted, and lacking in empathy (Miller, 2000).

Subsequent to Miller’s (2000) initial development of and examination with the SAPP, additional studies have sought to corroborate the SAPP as a valid and reliable measure of self-knowledge.

Validity of the SAPP

Validity of a measure refers to its legitimacy, evidenced by research in theory, in its intended purpose (American Psychological Association [APA], 1985). Having sound validity is imperative for a measure to be regarded as useful. There are several types of validity, one being construct validity, which is of particular consideration for an assessment like the SAPP. Construct validity refers to the
extent an assessment of a purported conceptual variable, such as self-knowledge, is adequately measured. Convergent validity and discriminant validity are two other types of validity that are used to establish construct validity (Hood, 2001).

Convergent validity refers to a statistically significant relationship between two different measures that are purported to assess the same construct. Divergent validity refers to the lack of a statistically significant relationship between two different measures that are purported to assess differing constructs (APA, 1985). Hood (2001) examined the validity of the SAPP by seeking to replicate the original findings and by studying its convergent and divergent validity. Hood did this by comparing the SAPP with the Private Self-Consciousness factor of the Self-Consciousness Scale, hypothesizing that the two would be positively correlated as they measured similar constructs. Likewise, Hood compared the SAPP to a dissimilar measure of self-esteem, the Tennessee Self-Concept Scale, hypothesizing that the two assessments would not be significantly correlated. Based on the interpretation of 48 participants’ results on each measure, Hood was successful at providing evidence to support discriminant validity, as no significant correlation was found between the SAPP and the Tennessee Self-Concept Scale ($r = .188, p > .05$). However, Hood failed to yield convincing evidence supporting convergent validity with the Private Self-Consciousness factor of the Self-Consciousness Scale, although there was also only a moderate correlation with the SAPP in the predicted direction ($r = -.30, p > .05$).
In another effort to confirm convergent validity, Anderson (2002) assessed for correlation between the SAPP and the Self-Monitoring Scale, a measure of people’s behavioral adaptation involving self-awareness (Snyder, 1974). However, this study, conducted with 98 students, also failed to provide evidence for convergent validity as no significant correlation between the two measures was found. Anderson (2002) concluded that the SAPP is not related to an individual’s capacity for and willingness to engage in situational behavioral adjustments. Therefore, the constructs assessed by each measure are considered discrete.

Comparison among peer ratings of targeted subjects and the subjects’ self-ratings (via the SAPP score) was the basis for Layton’s (2005) study to provide an additional measure of construct validity. Subjects of the study consisted of graduate and undergraduate students at a local university. Analysis of 36 SAPP scores with 64 corresponding peer predictions did produce a correlation in the predicted direction (i.e., lower SAPP scores with the mean predicted scores of their peers), although it did not reach the level of statistical significance, (Layton, 2005). In an attempt to replicate Layton’s (2005) study, Wolf (2006) yielded significant findings indicating correlation between a subject’s SAPP scores and peer predictions. Wolf (2006) hypothesized and found evidence to suggest that those with lower SAPP scores (greater self-knowledge) would have a higher concordance with their peers’ ratings than those with higher SAPP scores (lower self-knowledge). Although Wolf (2006) sought to have a larger sample size than her predecessor, she only obtained useable data from 29 subjects.
Hickey (2005) took a similar approach as Layton (2005) in developing a concordance measure to confirm construct validity. Instead of peers, Hickey obtained predictions from two family members for each subject from whom a SAPP score was derived. Hickey’s 2005 study was also inhibited by small sample size, \( N = 34 \), and the results were not significant. In a similar effort, Afanador (2006) sought to examine the SAPP’s validity using SAPP scores of individuals in therapy in comparison with their therapist’s prediction of their amount of self-knowledge. However, results of this study were not significant and again, likely inhibited due to sample size. In 2007, Blankemeier replicated Hickey’s (2005) study with a somewhat larger sample size \( (N = 51) \) using a similarly derived concordance measure of an individual’s predicted personality score and corresponding family members’ predicted score. The results of this study revealed significant correlation between the concordance measure and participant’s SAPP scores (Blankemeier, 2007).

The overarching pattern of the validation studies conducted to date on the SAPP suggests some construct validation of the SAPP.

*Reliability of the SAPP*

The reliability of a measure refers to its consistency when repeated on individuals or groups of subjects (APA, 1985). Test-retest reliability is a type of reliability that includes administering a measure on the same group of people on two occasions. A measure has sufficient test-retest reliability if a significant
correlation is found between the sets of scores obtained on both testing occasions (Fekken, 2000). This indicates that the measure yields consistent results. While establishing the consistency of a measure is always paramount when considering the utility of a measure, it is especially critical when the measure in question intends to assess a stable construct, such as personality (Elghossain, 2012).

Silva (2011) conducted a study examining the test-retest reliability of the SAPP over a two-week period. She hypothesized that the subjects’ SAPP scores would maintain their stability over time, producing a comparable score on the second testing occasion. Silva’s study consisted of 62 participants obtained through nonrandomized sampling methods and mostly consisting of psychology graduate students. Results indicated a significant correlation between the subjects’ two derived SAPP scores, ($r^2 = .397, p < .05$). However, Silva noted that this is below what is typically deemed an acceptable correlation.

Hirsch (2012) replicated Silva’s (2011) study, with 58 participants completing both testing trials. Results of the study yielded a significant moderate correlation between the testing trials that were conducted over a two-week period. Sverdlova (2012) and Elghossain (2012) conducted similar studies of test-retest reliability of the SAPP over a four-week and six-week period, respectively. Sverdlova’s study was conducted on 58 participants and found the two SAPP scores to be significantly correlated but at a level that is below what is typically deemed acceptable. Elghossain’s results from the 47 participants who completed both trials revealed a significant, high correlation.
(2012) both found correlations within a psychometrically acceptable range, particularly when compared to the limiting reliability values of each of the 21 individual 16PF scales. Thus, the data to date collected would suggest that the SAPP is a generally reliable measure of self-knowledge.

**SAPP Sten Scores**

McElliggot (2015) replicated Miller’s (2000) original study to derive sten score equivalents from a database of 609 subjects who have had SAPP scores derived from their participation in earlier studies on the SAPP. McElliggot adjusted the SAPP scores using a simple linear transformation (subtraction from the highest possibly achieved SAPP score – 189) to reverse their direction. In other words, high SAPP scores, with this transformation, would then reflect higher levels of self-knowledge. McElliggot used two methods to develop sten scores to correspond with given SAPP scores, and an extraordinarily high correlation (r = .95) was found between these methodologies.
Statement of Purpose and Hypothesis

The intention of this study was to further investigate the test-retest reliability of the SAPP over a two-week period with a larger sample. Replication of prior studies with similar intent is paramount in confirming the consistency of SAPP scores over time. It was hypothesized that the subjects’ SAPP scores would maintain stability over time and would produce similar scores on a second trial that was administered two weeks following the initial trial. It was hypothesized that the test-retest reliability coefficient would be significant and would be equal to or exceed that which Silva obtained in 2011 and Hirsch obtained in 2012. Should the SAPP continue to yield evidence to support its reliability, it would then support one important psychometric condition necessary to potentially have it be one of the only singular measures of self-knowledge currently available.
Method

Subjects

Subjects of this study were obtained from the community through nonrandomized sampling, and were asked to participate on a voluntary basis. Emails with instructions on how to complete the online questionnaires were sent to 71 individuals. Those who responded to the first trial received a second trial two weeks later. 34 individuals completed both trials. Analysis and interpretation of data was only conducted on subjects who completed both trials.

Instruments

The instruments used for both trials in the present study included a comprehensive letter of instruction (sent via email), and an online version of both the 16PF Fifth Edition and Scale of Accurate Personality Prediction.

Procedure

To determine test re-test reliability of the SAPP, each subject was requested to participate in two distinct trials of testing. There was a two-week period between trials. Each participant received an email including a comprehensive letter of instructions on how to complete the testing materials (see Appendix A). Each participant that completed the first trial received a second email with instructions two days before they were to complete their second trial. Examples of the emails
sent to participants for the first trial and second trial can be found in Appendices B and C, respectively (adapted from Hirsch, 2012).

The first email participants received contained instructions on how to complete the first trial of the study. The instructions contained a unique username and password combination to use to access the 16PF. They were also given a unique 4-digit ID code to enter on both questionnaires. The participants were provided with a link to the IPAT website to complete the 16PF online. Once the participant completed the 16PF they were directed to the Qualtrics website to fill out a questionnaire adapted from a blank 16PF profile form (Appendix D). They were advised to rate themselves on how they believe they corresponded to the 16 primary personality factors and 5 global factors.

Following two weeks after the initial testing, the subjects who completed the first trial were given a second email with instructions parallel to the first testing. They were given another unique 4-digit ID code and username/password combination. They completed an online version of the 16PF and the Qualtrics questionnaire, where they again predicted their scores on the 16 primary and 5 global factor scales. Once data collection was complete, copies of participants Basic Score Report was provided by IPAT. The subject’s SAPP score was derived using Miller’s (2000) formula, whereby the subject’s predicted score was first subtracted from their obtained score. Subsequently, the SAPP score was obtained through the summation of the absolute values of the discrepancy of scores, and then
the resultant score was subtracted from 189 (McElligott). Data from the two testing trials were compared and analyzed only for subjects who participated in both trials.
Results

There were 71 subjects who volunteered to participate and were provided with instructions on how to complete the study. Of those who initially volunteered, 52 completed the first trial of testing. Overall, 34 participants completed both trials of testing in their entirety in the necessary time frame, and only their data was used in the finding of this study. Participants were first directed to complete an online version of the 16PF, which was made available through IPAT. The online 16PF was viewed and initiated 98 times. This includes all potential participants and both testing trials. The test was started 88 times, and completed 87 times. Thus, the completion rate was 98.9% and only one participant discontinued the testing after starting the survey. Demographic variables were obtained and analyzed for each participant, including age, highest level of education obtained, race, ethnicity, gender, geographic location, and marital status.

Of the 34 participants who completed both trials of the study the mean age was 39.1 years, with ages ranging from 22 to 68. Most participants (35.3%) reported having a Master’s degree. Of all participants who completed both trials of the study, 32.45% indicated they had a Bachelor’s degree, 17.6% reported having completed some college, 11.8% indicated they had obtained a high school diploma, and 2.9% reported having a professional degree. When asked to choose one or more race participants identify as, 85.3% of participants (N=29) identified as Caucasian. 11.8% (N=4) of participants identified as Black or African American,
2.9% (N=1) identified as Brown, and 2.9% (N=1) identified as Middle Eastern. Of the total participants, 11.8% (N=4) identified as Spanish, Hispanic, or Latino. 28 participants (82.4%) identified as female, and 6 (17.6%) identified as male.

In regards to geographic location, 17 participants reported they were residents of South Carolina, 8 indicated they were Florida residents, 3 endorsed being residents of New York, and 2 indicated they were Louisiana residents. Additionally, there was 1 participant who indicated they resided in California, 1 who reported being a Georgia resident, 1 who endorsed living being a resident of Tennessee, and 1 who reported being a North Carolina resident. The majority of participants, 52.9% (N=18) indicated their marital status to be married, while 38.2% (N=13) reported their marital status to be single, and 8.8% (N=3) identified their marital status as divorced. A comparison of the demographic data of the present study with the normative sample of the 16PF and full database of research completed on the Scale of Accurate Personality Prediction can be found in Table 1. The sample of the current study is predominately comprised of individuals who identified as female, while both other samples have more of an equal gender distribution. In regards to racial and ethnic diversity, the current sample is comparable to the 16PF normative sample. A greater percentage of participants in the present study have at least a college degree than both the normative sample and SAPP database sample. Similar to the SAP database sample, the majority of participants in the present study live in the southeast.
To examine the test-retest reliability of the SAPP, the SAPP scores for each participant were derived during the initial testing trial and again two weeks later. SAPP scores of the sample during the initial testing trial revealed a mean of 150.76 and a standard deviation of 9.15, with scores ranging from 133.00 to 174.00. For the SAPP scores derived during the second trial six weeks later, a mean of 151.71 was found, with a standard deviation of 7.94 and scores ranging from 138.00 to 173.00. A Pearson correlation revealed a significant moderate correlation between the SAPP scores obtained during both trials ($r^2 = .584$, $p< .01$). The two-week test-retest reliability values obtained for each of the 21 factors are comparable to the 16PFs normative sample (Table 2). Additionally, the means from the present study are comparable to the mean of SAPP scores from the full database of research on the measure (Table 3). However, the present study yielded standard deviations that are somewhat lower than the standard deviation of the entire database, which can be attributed to the smaller range of scores.
Discussion

This study is a replication of two-week test-retest reliability of the SAPP, previously examined by Silva (2011) and Hirsch (2012). A Pearson correlation analysis was conducted, and the results revealed a significant moderate correlation ($r^2 = .584$) was revealed. As hypothesized, the correlation found in the present study exceeds what was obtained by Silva in 2011 ($r^2 = .397$) and Hirsch in 2012 ($r^2 = .566$). The moderate magnitude of the correlation found in the present study can partially be attributed to the correlation of the 21 factors that emerged over the same two-week period. As the SAPP is comprised of all 21 factors obtained on the 16PF, the strength of the correlation coefficient of the SAPP is limited by the test-retest reliability coefficient for each factor. Therefore, the test-retest reliability coefficient of the SAPP would not be expected to exceed that of the lowest test-retest reliability coefficient of the 21 factors of the 16PF, which in this study was found to be Factor N: Privateness ($r^2 = .721$). This is comparable to the correlation coefficient found in the normative sample (see Table 2). Thus, the significant moderate correlation found in the present study is considered to offer more support of the SAPP as a reliable measure of self-knowledge. However, limitations that may have influenced the present results should be taken into consideration.

A major limitation of the study is the small sample size. The intent of this study was to obtain a sample size greater than Silva (2011) and Hirsch (2012), who obtained 62 and 58 participants, respectively. However, the present sample size is
much smaller than that of both researchers (N=34). Several methods were implemented to increase sample size. For instance, like Hirsch (2012), testing materials were distributed online to make participation efficient and easily accessible. It was anticipated that participants would be more likely to volunteer to complete the survey online than by hand, which would require that they mail back the materials to the researcher upon completion. Social media was also utilized to help advertise the study, and participants were encouraged to share information on how to participate in the study with their friends and family. Multiple reminders were sent via email to participants to complete both the first and second trials of the study. Nevertheless, a much smaller sample size than what was hoped for was obtained. While the results indicated a significant moderate correlation, the small sample size may have enhanced the effect size, which could have increased the probability of a false-positive result.

Another limitation to the present study is the method in which it was conducted. While the online format of the 16PF and SAPP was convenient, several difficulties with this approach were noted. First, one participant contacted the researcher due to being unable to login with the username and password from IPAT they were assigned. This participant was given a new username and password and reported after data collection was complete that they were still unable to access the surveys. It is possible that similar technical difficulties occurred that were not reported to the researcher, which could have contributed to the low sample size or attrition rate.
The way in which testing instructions were presented to participants was another limitation to the study. In order to easily track each participant’s data from the initial and second testing trial, they were each assigned two 4-digit codes, one to use the first testing trial and one to use the second. The IPAT website required that each participant use only one username/password for testing trial. Therefore, in addition to the two 4-digit codes, participants were provided with two separate username and password combinations. While participants was provided with explicit instructions (See Appendices A, B, and C), approximately a third of participants who completed at least the first trial (N=18) did not follow instructions correctly or contacted the examiner due to confusion regarding the testing process. Additionally, following completion of the 16PF, participants were prompted to click on a hyperlink that would redirect them to the SAPP survey, and several participants (N=3) contacted the examiner to report that they were uncertain as to whether or not they were to complete the second questionnaire. Thus, instructions for completing the testing online may be perceived as more complicated than when Silva (2011) conducted the research in a pencil and paper format.

Lack of control of testing environment is another limitation of the present study. As the questionnaires were administered online, it cannot be guaranteed that each participant completed both trials without interruption and with adequate attention given to test items. This limitation exists even when conducted in a paper and pencil format, as Silva (2011) found. One benefit of using the online administration method as opposed to the paper and pencil method is that
timestamps were available notating when a participant completed the two questionnaires for both trials of the study. This helped ensure that the data to be analyzed was completed within the appropriate time frame.

A final limitation to be considered includes the homogeneity of the sample. The majority of participants were female, Caucasian, college graduates, and from the southeast. Thus, the sample is not representative of the general population. However, the racial and ethnic demographic composition of the present sample is more diverse than that of Silva (2011) and Hirsch (2012). For instance, 91.9% Silva’s (2011) sample and 100% of Hirsch’s (2012) sample were Caucasian, compared to 85.3% of the present sample. Although adequate racial and ethnic diversity is still lacking from the present sample, it is still more comparable to the normative sample (see Table 1).

Although the present study is not without limitations, the results are nonetheless valuable. The significant moderate correlation ($r^2 = .584$) that emerged further supports the test-retest reliability of the SAPP. Results from previous studies on the psychometric properties of the SAPP in conjunction with those found in the present study indicate it is an acceptable measure of self-knowledge. The SAPP has the potential for vast beneficial utility. For instance, the SAPP could be used as the initial platform for insight development by assisting with identification of discrepancies in a client’s personality prediction. Additionally, the SAPP can help mental health professionals in their conceptualization and intervention
planning, as they will be able to better tailor their treatment to account for their client’s level of self-knowledge.

Future research may want to emphasize refinement of the standardization of administration procedures based on its intended use. That is, if the SAPP is primarily intended for clinical use, future studies may want to replicate the environment and method by which it would be administered. To do so, it may first be worth establishing alternate form reliability between the paper and pencil and online version of the SAPP. Additionally, it is suggested that future research on the SAPP take place in a controlled testing environment, where the researcher is available to provide in-person instruction. This would improve the control of the testing environment and allow for clarification of instructions as necessary. This may help ensure that participants adequately comprehend the testing instructions and are giving forth their best effort. Obtaining larger and more diverse sample sizes should continue to be emphasized in future research on the SAPP. In conclusion, the support for the test-retest reliability of the SAPP found in the present study indicates the SAPP is worth further examination in order to establish it as a psychometrically sound measure of self-knowledge.
References


doi:http://dx.doi.org.portal.lib.fit.edu/10.1037/10522-014


Table 1. Demographic Percentages

<table>
<thead>
<tr>
<th>Demographic Category</th>
<th>Current Sample Percent (N=34)</th>
<th>SAPP Database Percent&lt;sup&gt;a&lt;/sup&gt; (N=645)</th>
<th>Normative Sample Percent&lt;sup&gt;b&lt;/sup&gt; (N=2500)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17.6%</td>
<td>42.0%</td>
<td>48.7%</td>
</tr>
<tr>
<td>Female</td>
<td>82.4%</td>
<td>58.0%</td>
<td>51.3%</td>
</tr>
<tr>
<td><strong>Race&lt;sup&gt;c&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American/Black</td>
<td>11.8%</td>
<td>2.3%</td>
<td>12.1%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.0%</td>
<td>9.3%</td>
<td>2.9%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>85.3%</td>
<td>71.0%</td>
<td>80.2%</td>
</tr>
<tr>
<td>Native American</td>
<td>0.0%</td>
<td>0.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Other</td>
<td>5.9%</td>
<td>5.3%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Hispanic Origin</td>
<td>11.8%</td>
<td>11.9%</td>
<td>9.0%</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 to 17</td>
<td>0.0%</td>
<td>1.2%</td>
<td>4.6%</td>
</tr>
<tr>
<td>18 to 24</td>
<td>5.9%</td>
<td>51.3%</td>
<td>13.8%</td>
</tr>
<tr>
<td>25 to 44</td>
<td>58.8%</td>
<td>34.4%</td>
<td>41.7%</td>
</tr>
<tr>
<td>45 to 54</td>
<td>0.0%</td>
<td>5.3%</td>
<td>12.9%</td>
</tr>
<tr>
<td>55 to 64</td>
<td>32.4%</td>
<td>5.9%</td>
<td>10.8%</td>
</tr>
<tr>
<td>65 and older</td>
<td>2.9%</td>
<td>1.9%</td>
<td>16.2%</td>
</tr>
<tr>
<td><strong>Education Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12 years</td>
<td>11.8%</td>
<td>5.0%</td>
<td>61.5%</td>
</tr>
<tr>
<td>12.5-16 years</td>
<td>17.6%</td>
<td>55.8%</td>
<td>22.7%</td>
</tr>
<tr>
<td>16+ years</td>
<td>70.6%</td>
<td>39.2%</td>
<td>15.8%</td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>38.2%</td>
<td>72.9%</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>52.9%</td>
<td>20.6%</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>8.8%</td>
<td>4.8%</td>
<td></td>
</tr>
<tr>
<td>Separated</td>
<td>0.0%</td>
<td>1.1%</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>0.0%</td>
<td>0.6%</td>
<td></td>
</tr>
<tr>
<td><strong>Geographic Location</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>88.2%</td>
<td>78.9%</td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td>2.9%</td>
<td>3.8%</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>8.8%</td>
<td>13.1%</td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>0%</td>
<td>4.0%</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>0%</td>
<td>0.2%</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>From the SAPP Database, which is an accumulation of data from multiple studies on the SAPP.


<sup>c</sup>Totals may exceed 100% since participants of the present study were allowed to choose more than one race. Additionally, in the present study and in the normative sample those who identified as Hispanic also endorsed at least one race category.
Table 2. Test-Retest Reliability Data

Test-Retest Interval

<table>
<thead>
<tr>
<th>Primary Factor</th>
<th>Two-Week (N=34)</th>
<th>Two-Week(^a) (N=204)</th>
<th>Two Month(^a) (N=159)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Warmth</td>
<td>.80</td>
<td>.83</td>
<td>.77</td>
</tr>
<tr>
<td>B Reasoning</td>
<td>.84</td>
<td>.69</td>
<td>.65</td>
</tr>
<tr>
<td>C Emotional Stability</td>
<td>.92</td>
<td>.75</td>
<td>.67</td>
</tr>
<tr>
<td>E Dominance</td>
<td>.83</td>
<td>.77</td>
<td>.69</td>
</tr>
<tr>
<td>F Liveliness</td>
<td>.92</td>
<td>.82</td>
<td>.69</td>
</tr>
<tr>
<td>G Rule-Consciousness</td>
<td>.86</td>
<td>.80</td>
<td>.76</td>
</tr>
<tr>
<td>H Social Boldness</td>
<td>.89</td>
<td>.87</td>
<td>.79</td>
</tr>
<tr>
<td>I Sensitivity</td>
<td>.83</td>
<td>.82</td>
<td>.76</td>
</tr>
<tr>
<td>L Vigilance</td>
<td>.89</td>
<td>.76</td>
<td>.56</td>
</tr>
<tr>
<td>M Abstractedness</td>
<td>.84</td>
<td>.84</td>
<td>.67</td>
</tr>
<tr>
<td>N Privateness</td>
<td>.72</td>
<td>.77</td>
<td>.70</td>
</tr>
<tr>
<td>O Apprehension</td>
<td>.92</td>
<td>.79</td>
<td>.64</td>
</tr>
<tr>
<td>Q1 Openness to Change</td>
<td>.91</td>
<td>.83</td>
<td>.70</td>
</tr>
<tr>
<td>Q2 Self-Reliance</td>
<td>.94</td>
<td>.86</td>
<td>.69</td>
</tr>
<tr>
<td>Q3 Perfectionism</td>
<td>.81</td>
<td>.80</td>
<td>.77</td>
</tr>
<tr>
<td>Q4 Tension</td>
<td>.88</td>
<td>.78</td>
<td>.68</td>
</tr>
<tr>
<td>SAPP</td>
<td>.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Global Factor

<table>
<thead>
<tr>
<th></th>
<th>Extraversion</th>
<th>Anxiety</th>
<th>Tough- Mindedness</th>
<th>Independence</th>
<th>Self-Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.93</td>
<td>.97</td>
<td>.93</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>.91</td>
<td>.84</td>
<td>.87</td>
<td>.84</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>.80</td>
<td>.70</td>
<td>.82</td>
<td>.81</td>
<td>.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Sample Trial One (N=34)</td>
<td>41.00</td>
<td>133.00</td>
<td>174.00</td>
<td>150.76</td>
<td>9.15</td>
</tr>
<tr>
<td>Current Sample Trial Two (N=34)</td>
<td>35.00</td>
<td>138.00</td>
<td>173.00</td>
<td>151.71</td>
<td>7.94</td>
</tr>
<tr>
<td>Current Sample Both Trials (N=34)</td>
<td>41.00</td>
<td>133.00</td>
<td>174.00</td>
<td>151.24</td>
<td>8.52</td>
</tr>
<tr>
<td>SAPP Database* (N=643)</td>
<td>69.30</td>
<td>101.30</td>
<td>170.60</td>
<td>147.04</td>
<td>13.05</td>
</tr>
</tbody>
</table>

*From the SAPP Database, which is an accumulation of data from multiple studies on the SAPP.
Appendix A

Instructions for Participants

Dear Participant,

Thank you for agreeing to participate in this research study. The purpose of this study is to further explore the reliability of a new scale of self-knowledge for the Sixteen Personality Factor Questionnaire (16PF), which is the Scale of Accurate Personality Prediction (SAPP). More specifically, this study aims to establish test-retest reliability. This will require your participation on two distinct testing occasions. The second testing session will take place two weeks following the first testing session. Please read all of the following steps before beginning the study. After reading them carefully, follow them in order:

1. To complete the 16PF assessment, go to: https://www.netassess.ipat.com/ and enter the unique user name and password provided in the body of the email. Read the Terms of Service Provision and select Yes, I will. Select Continue. Please note: Exiting your web browser without agreeing to the Terms of Service, or responding No, will result in your passcode being locked and will require the code to be reset by the project team.
2. If this is your first trial, please enter the Trial One ID Code where indicated. There is no need to enter your name. You will use the Trial Two ID Code two weeks later. If this is your second trial, please enter the Trial Two ID Code.
3. The 16PF should take approximately 30min to complete. Once you have completed the 16PF, click on the link at the end. This link will redirect you to take the second questionnaire, the SAPP.
4. Please reuse the same ID you used on the first questionnaire on the second questionnaire.
5. Answer each question on the questionnaire.
6. After a two-week delay you will complete the questionnaires again with a second username, password, and ID code. As a reminder, you will receive an email two days prior to when you are to complete the second trial. These instructions will be sent to you again. Please remember to use the second username, password, and ID code that will be provided to you. It is requested that you complete the second trial within a 24hr period of the date that is exactly two weeks from when you completed the first trial.

Please be assured that the information you provide us is confidential. Your completion of the materials will serve as your consent to participate in this study. If you are interested in summary feedback concerning this study, please contact me. My email address is provided at the end of this page. Please note, to protect anonymity, individual feedback cannot be provided; only group summary results
will be available. These results will be available upon completion of the research project. Again, your assistance is appreciated. Please contact me if you have any further questions regarding the research.

Sincerely,
Lisa Stewart, M.S.
lstewart2014@my.fit.edu
Appendix B

First Trial Email to Participants

Dear Participant,

Thank you for agreeing to participate in my doctoral research study. Attached is a Word Document with detailed instructions on how to participate. There are two questionnaires to complete for the study, and the instructions will explain how to do so.

This is your first trial of the study. Two weeks from the day you decide to complete the first trial, you will complete the second trial. I will send you a reminder email two days prior to when you are to complete the second trial. Please make sure you complete the first trial on a date you know you can complete the second trial exactly two weeks later. Reference the ID codes below when completing the questionnaires today and two weeks later.

Trial One ID Code: 0111 (Use the first time you complete the questionnaires)
Trial Two ID Code: 0112 (Use the second time you complete the questionnaires)

You will be prompted to enter a username and password before completing the first part of the first trial. Please use the following username and password when completing the first trial:

Username: 22fc8faf0
Password: rimazuty

You are welcome to send me an email with any questions or concerns you may have regarding this project. I greatly appreciate your time!

Regards,
Lisa Stewart, M.S.
lstewart2014@my.fit.edu
Appendix C

Second Trial Email to Participants

Dear Participant,

Thank you for participating in the first trial of my study. Two days from now you will complete the second trial of my study Attached to this email is a Word Document with instructions on how to complete your participation in this study. Please read the instructions carefully. As with the first trial of the study, there are two questionnaires to complete. The instructions will explain how to do so.

Reference the ID code below when completing the questionnaire two days from now.

Trial Two ID Code: 0112

You will be prompted to enter a username and password before completing the first part of the second trial. **Note: Do not use the same username and password or ID code that you used during the first trial.** Please use the following username and password when completing the second trial:

Username: 22fc8faf1
Password: zecimagi

You are welcome to send me an email with any questions or concerns you may have regarding this project. I greatly appreciate your time!

Regards,
Lisa Stewart, M.S.
lstewart2014@my.fit.edu
Appendix D

Blank 16PF Individual Record Form

### PRIMARY FACTORS

<table>
<thead>
<tr>
<th>Factor</th>
<th>Stem</th>
<th>Left Meaning</th>
<th>Standard Test Score (STEN)</th>
<th>Right Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Warmth</td>
<td>Reserved, Impersonal, Distant</td>
<td>• • • • • •</td>
<td></td>
<td>Warm, Outgoing, Affable to Others</td>
</tr>
<tr>
<td>B: Reasoning</td>
<td>Conform</td>
<td>• • • • • •</td>
<td></td>
<td>Abstract</td>
</tr>
<tr>
<td>C: Emotional Stability</td>
<td>Resolute, Emotionally Changeable</td>
<td>• • • • • •</td>
<td></td>
<td>Emotionally Stable, Adaptive, Mature</td>
</tr>
<tr>
<td>D: Dominance</td>
<td>Dominant, Cooperative, Authoritative</td>
<td>• • • • • •</td>
<td></td>
<td>Declarative, Authorial, Assertive</td>
</tr>
<tr>
<td>F: Liveliness</td>
<td>Serious, Resistant, Careful</td>
<td>• • • • • •</td>
<td></td>
<td>Lively, Animated, Spontaneous</td>
</tr>
<tr>
<td>G: Rule-Consciousness</td>
<td>Expedient, Nonconforming</td>
<td>• • • • • •</td>
<td></td>
<td>Rule-Conscious, Diligent</td>
</tr>
<tr>
<td>H: Social Boldness</td>
<td>Shy, Threat Sensitive, Timid</td>
<td>• • • • • •</td>
<td></td>
<td>Socially Bold, Venturesome, Little Shy</td>
</tr>
<tr>
<td>I: Sanicality</td>
<td>Utilitarian, Obstinate, Unemotional</td>
<td>• • • • • •</td>
<td></td>
<td>Sensitive, Aesthetic, Sentiments</td>
</tr>
<tr>
<td>J: Vigilance</td>
<td>Invent, Unconcerned, Accepting</td>
<td>• • • • • •</td>
<td></td>
<td>Vigilant, Skeptical, Selective, Watchful</td>
</tr>
<tr>
<td>K: Abstraction</td>
<td>Generated, Imaginative, Idea-Oriented</td>
<td>• • • • • •</td>
<td></td>
<td>Abstracted, Imaginative, Idea-Oriented</td>
</tr>
<tr>
<td>N: Focussedness</td>
<td>Fortnight, Genuine, Airless</td>
<td>• • • • • •</td>
<td></td>
<td>Focused, Determined, Non-Delaying</td>
</tr>
<tr>
<td>O: Apogynism</td>
<td>Self-Assembled, Unwielded, Compliant</td>
<td>• • • • • •</td>
<td></td>
<td>Apogynistic, Self-Delaying, Wantied</td>
</tr>
<tr>
<td>Q1: Openness to Change</td>
<td>Traditional, Attached to Familiar</td>
<td>• • • • • •</td>
<td></td>
<td>Open to Change, Reorienting</td>
</tr>
<tr>
<td>Q2: Self-Reliance</td>
<td>Group-Oriented, Affiliative</td>
<td>• • • • • •</td>
<td></td>
<td>Self-Oriented, Solicit, Individualistic</td>
</tr>
<tr>
<td>Q3: Perfectionism</td>
<td>Tolerates Disorder, Unemotional, Flexible</td>
<td>• • • • • •</td>
<td></td>
<td>Perfectionistic, Organized, Self-Disciplined</td>
</tr>
<tr>
<td>Q4: Tension</td>
<td>Relaxed, Poised, Patient</td>
<td>• • • • • •</td>
<td></td>
<td>Tense, High Energy, Impatient, Onset</td>
</tr>
</tbody>
</table>

### GLOBAL FACTORS

<table>
<thead>
<tr>
<th>Factor</th>
<th>Left Meaning</th>
<th>Standard Test Score (STEN)</th>
<th>Right Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>EK: Extraversion</td>
<td>Introverted, Socially Inhibited</td>
<td>• • • • • •</td>
<td>Extraverted, Socially Participating</td>
</tr>
<tr>
<td>AX: Anxiety</td>
<td>Low Anxiety, Unperturbed</td>
<td>• • • • • •</td>
<td>High Anxiety, Perturbable</td>
</tr>
<tr>
<td>TM: Tough-Mindedness</td>
<td>Resolute, Open-Minded, Infatuated</td>
<td>• • • • • •</td>
<td>Tough-Minded, Resolute, Unempathic</td>
</tr>
<tr>
<td>Rn: Independence</td>
<td>Accommodating, Agreeable, Softness</td>
<td>• • • • • •</td>
<td>Independent, Persuasive, Willful</td>
</tr>
<tr>
<td>SC: Self-Control</td>
<td>Unrestricted, Reaches Urgent</td>
<td>• • • • • •</td>
<td>Self-Controlled, Inhibits Urges</td>
</tr>
</tbody>
</table>