2-2018

An Investigation into the Test-Retest Reliability of the Scale of Accurate Personality Prediction with a Six Week Interval

Jared Marshall Barrow

An Investigation into the Test-Retest Reliability of the Scale of Accurate Personality Prediction with a Six Week Interval

By

Jared Marshall Barrow

Bachelor of Science
Engineering
Franklin W. Olin College of Engineering
2010

Master of Science
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
Doctorate
In
Psychology

Melbourne, Florida
February 2018

We the undersigned committee
hereby approve the attached doctoral research project.

An Investigation into the Test-Retest Reliability of the Scale of Accurate Personality Prediction with a Six Week Interval

By

Jared Marshall Barrow

Philip D. Farber, Ph.D. Associate Professor School of Psychology Committee Chair	Richard T. Elmore, Jr. Ph.D Associate Professor School of Psychology Committee Member
John C. Frongillo, Ph.D. Assistant Professor School of Arts and Communication Committee Member	Mary B. Kendel, Ph.D. Dean and Professor School of Psychology

Abstract

TITLE: An Investigation into the Test-Retest Reliability of the Scale of Accurate Personality Prediction with a Six Week Interval

AUTHOR: Jared M. Barrow, M.S.

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

The Scale of Accurate Personality Prediction (SAPP) is a measure to determine the accuracy of an individual's ability to predict his/her own personality traits. Previous studies have demonstrated the validity of SAPP as a measure, and this study serves as one in a series of study on the reliability of the SAPP. To establish the test-retest reliability of the SAPP, the SAPP scores of 22 participants were calculated from two testing trials separated by a six week interval. A Pearson correlation indicated a significant correlation of moderate strength between the SAPP scores from the two testing trials ($r=.572$, $p<0.01$). The implication of these results within the context of the previous test-retest reliability studies and limitations to the current study are discussed.

Overview

The topic of personality has been a key focus among psychologists and philosophers for many years. While many psychology theorists have come to different conclusions about the definition, origins, and component elements of one's personality, most recognize and highlight the importance of personality factors in understanding and possibly determining human behavior. One sub-topic of personality that has gained much interest in the last few decades is the degree to which individuals are aware their own personalities, and more specifically, how well they can correctly identify their unique personality traits. This area of personality typically involves the ability to accurately predict the direction and degree of the more accepted personality traits. Such accurate personality prediction has been identified as one way to empirically determine one's level of self-knowledge. Within applied clinical psychology, treatments may attempt to enhance self-knowledge to influence a patient's behavior. Though the construct of self-knowledge is an important area of study within the personality field for this reason, few assessment tools have been developed to assess one's level of this type of self-knowledge.

In 2000, Miller developed the Scale of Accurate Personality Prediction (SAPP), which was hoped to provide a measure of a person's self-knowledge. The SAPP index is based upon the comparison of the individual's predicted and obtained scores on the Sixteen Personality Factor Questionnaire, Fifth Edition (16PF). Since the time of its initial construction, the SAPP has gone through a

good deal of construct validation efforts (Afanador, 2006; Anderson, 2002; Blankmeier, 2007; Glywasky, 2003; Grossenbacher, 2006; Hadricky, 2009; Hickey, 2004; Hood, 2001; Layton, 2004; Winter, 2002; Wolf, 2006) with results suggesting that the scale may well be a viable instrument for measuring a form of one's self-knowledge.

In addition to establishing the validity of the measure, studies have been conducted to evaluate the test-retest reliability of the measure. Silva (2011) attempted the first exploration of the test-retest reliability of the SAPP. Silva's study had participants complete the 16PF and predict their own scores, with a two-week interval between test administrations. Silva's study yielded a significant correlation, however, it was below what is typically expected for a test-retest correlation, and it was suggested that this was a result of the study's small sample size. Subsequent studies employed a similar methodology, examined the test-retest reliability at two-week (Hirsch, 2012), four-week (Sverdlova, 2012), and six-week (Elghossain, 2012) intervals. Each of these studies also yielded significant correlations. The relatively small sample sizes in all of the above studies suggest that replication of these reliability efforts is necessary. The current study contributed to the test-retest reliability data of the SAPP, utilizing the six-week interval period. With a sample of 22 participants, the Pearson correlation of r=.572 was found to be statistically significant ($p<0.01$) and of moderate strength. However, with this small sample size, the possibility of an enlarge effect size remains.

Table of Contents

List of Tables

Table

Acknowledgements

I would like to thank Dr. Philip Farber, who has help to shepherd me through this research process. Additionally, he has provided sage advice through my graduate school career, along with great laughs and fascinating conversations. I currently lack the vocabulary to express the depth of my appreciation, and how grateful I have been to have him as an advisor to help guide me through this process. I would also like to thank my parents, David and Grace. Through their actions and support, I have felt a great deal of love over the years. I would like to thank my grandmother, Moy, for her constant encouragement. Thank you to my co-researchers, Lisa and Taylor, they have been great collaborators and I have been very fortunate to have them. I would like to thank Dr. Elmore and Dr. Frongillo for taking the time to serve on the committee. Additionally, I would also like to thank Michael Stowers, Psy.D and the staff at PSI/PAN, who made this research possible by providing grant in-kind funding for it, in the form of online administrations and scoring of the 16PF along with necessary data files.

Literature Review

Self-Knowledge

When a person presents for psychological treatment, a clinician assesses many aspects of his or her personality, including the areas of judgement and insight. An assessment of an individual's judgment examines his or her thought processes to see how those processes cause that person's actions. An examination of one's insight, however, considers how the individual reflects on those actions, their effects, and their meanings. The assessment of a person's insight analyzes introspective processes and one's knowledge of self. Psychologists have attempted to define what exactly is "self-knowledge," but it has proved to be no easy task. Hart & Matsuba attempted to create a definition by building on the foundation established by Scheffler in 1983 and Steup in 2008. Scheffler and Steup focused their work on defining the concept of knowledge and concluded that knowledge requires belief, truth of the belief, and a way to verify that truth. However, their work alone was inadequate to define self-knowledge, because the added layer of the self adds complexity to the analysis. Using Scheffler's and Steup's works as a base, Hart & Matsuba define self-knowledge as "the true beliefs one has about one's self" (2012, p. 7). They propose that three facts must be true for self-knowledge to be displayed. Their first assertion is that true beliefs are a subset of a person's beliefs about one's self and the truth aspect makes self-knowledge "difficult to assess" (Hart & Matsuba, 2012, p. 8). The second assertion is that the knowledge must be validated with a process "known to track truth" (Hart &

Matsuba, 2012, p. 8). Third, self-knowledge requires "*belief* in a proposition" (Hart & Matsuba, 2012, p. 8). Their research lays the framework for assessing self-knowledge.

Different theorists define the self in varied ways based on their perspective. Per Carl Rodgers, the self is composed of the psychological processes that dictate an individual's behavior (Engler, 2003). Within this conception, the self emerges from the interaction between the person and the environment. This is further explored with the self-concept, which is how the person perceives him or herself (Engler, 2003). In contrast, Bandura's concept of the 'self' focuses more on cognitive structures (Engler, 2003). Bandura (1978) uses the term self-system to describe "cognitive structures that provide reference mechanisms" which include a "set of subfunctions for perception, evaluation, and regulation of behavior" (as cited in Engler, 2003, p. 247). Within Bandura's social learning theory, the 'self' describes these cognitive structures and processes which allow people to interact with their environment and help shape their behavior (Engler, 2003). William James (1890/1998) stated that the self comes from consciousness and that "a man's self is the sum total of all that he can call his" (as cited in Hart & Matsuba, 2012). Hart & Matsuba draw three implications from James' statement. The first is that an individual can reflect on him or herself. The second is that the individual is the ultimate authority on what is included as part of the self. The third is that there is a self-awareness present along with an emotional investment in specific elements of the self. The authors further relate two components to the self: personal memories,

and a narrative of the self (Hart & Matsuba, 2012). These personal memories help form representations and generalizations about themselves in areas such as appearance, skills, relationships, and psychological characteristics. The narrative of self helps to integrate the self-characteristics from the personal memories and provide a frame which allows these memories and representations to be evaluated by the self (Hart & Matsuba, 2012). Vogt and Colvin (2005) note that the major research into self-knowledge has primarily focused on the process in which people "come to know themselves", opposed to the accuracy of people's self-perceptions.

While the concept and nature of self-knowledge has been discussed, the various aspects of self-knowledge can make it difficult to assess. A simple self-report of self-knowledge is unlikely to be useful because if a person has a low level of self-knowledge, he or she is likely to lack insight into this deficit and not report it (Vogt & Colvin, 2005). Kruger and Dunning (1999) proposed that this is a result of the skills need to be competent in an area are often the same skills needed to evaluate competence. Additionally, seeking external ratings of self-knowledge from friends and family is also problematic, because the rater would have to know both how the individual views him or herself, and what the individual is like (Vogt & Colvin, 2005). Early research conducted on accuracy of self-knowledge operationalized accuracy as the summed difference between a person's self-ratings on his/her own personal and an external criterion rating for these traits (Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Funder & Colvin, 1997; as cited in Vogt & Colvin, 2005). In this case, the criterion ratings came from a friend's ratings of

the individual's characteristics per guidelines. Cronbach (1995) noted that this method has the potential confound of conflating idiosyncrasies of the rater's response styles with the individual differences in the sensitivity. Another method for assessing self-knowledge is by calculating a correlation from an individual's ratings on personality traits and criterion ratings for the same traits (Vogt & Colvin, 2005). However, this process relies on the rank ordering and does not necessarily indicate if the individual's profile is like that generated by that of the criterion.

Researchers have theorized that accurate self-knowledge is associated with improved psychological well-being. An example of this is in decision making, where accurate self-knowledge may result in positive outcomes with enhanced psychological well-being (Vogt, 1998; as cited in Vogt & Colvin, 2005). Conversely, when decisions are made based on inaccurate self-knowledge, a person's psychological well-being may be negatively impacted (Funder, 1999; as cited in Vogt & Colvin, 2005). This is connected to the idea that many therapeutic approaches are based the assumption that accurate self-knowledge of problematic personality characteristics and maladaptive behavioral patterns leads to positive behavioral changes (Brown, 1991). When discussing self-knowledge, the question arises as to what exactly is being known about the self? One way to describe these aspects of the self is through the concept of personality. This results in self-knowledge being characterized as knowledge of one's personality.

Defining personality

While the subject of personality has been a topic of interest throughout history, there has not been a single unifying standard definition. Raymond Cattell (1950) defined personality as "that which permits a prediction of what a person will do in a given situation" (as cited in Ellis, Abrams, & Abrams, 2009, p. 231). Allport described and classified over 50 different definitions of personality in his first book, *Personality: A Psychological Interpretation*. Part of the difficulty in defining personality comes from the complexity of phenomena that psychologists and researchers attempt to explain by using it, along with the attempt to have a unifying link between people's actions and an overarching internal process to explain such actions. Some psychologists, notably B.F. Skinner, have attempted to explain human behavior in the absence of these constructs. Maddi (1980, p. 10) defines personality as a "stable set of characteristics and tendencies that determine those commonalities and differences in the psychological behavior (thoughts, feelings, and actions) of people that have continuity in time and that may not easily be understood as the sole result of the social and biological pressures of the moment ".

Despite the differing theoretical definitions of personality, Hjelle and Ziegler (1992) highlighted four similarities among them. The first is that the definitions of personality should emphasize the importance of individuality, or distinctiveness. The second is that these definitions of personality should describe personality within a hypothetical structure or organization. This layer of abstraction is typically based on inferences from behavioral observations. The third similarity

is that most definitions of personality should view personality within a development context or in terms of a life history (Hjelle & Ziegler, 1992). This helps to characterize personality as an evolving entity, subject to potential influences such as environmental characteristics and biological inclinations. The final similarity is that most theoretical definitions of personality should see personality as "representing those characteristics of the person that account for consistent patterns of behavior" (Hjelle & Ziegler, 1992, p. 5). This similarity views personality as being stable over time and across contexts, and emphasizes a sense of continuity. While all the definitions may vary from theorist to theorist, all attempt to explain, and to some degree predict, the behavior of the individual.

Much like the various personality theorists incorporating a developmental perspective, it helpful to discuss the history of personality and psychology to understand the development of the concept and study of personality.

History of Personality

The history of personality begins not in the field of psychology, but in the field of philosophy. Throughout history, many philosophers have developed a conception of personality. In Ancient Greece, there were several philosophers who speculated on the nature of personality and its components. One of the notable early Greek ideas of personality originated from Plato. Plato viewed the soul as the seat of personality and his conception of the soul consisted of the forces of reason, emotion, and appetite (Ellis, Abrams, & Abrams, 2009). Plato also argued that the powerful force of reason worked to keep the more primitive forces of appetite and

emotion in check. His student, Aristotle, named the seat of personality the psyche and thought that the psyche was the product of a biological process. Because of this, Ellis et al., (2009) referred to Aristotle as the first biological psychologist. Aristotle conceptualized the psyche as a set of faculties with a distinct hierarchy. At the bottom of this hierarchy was the human's drive to fulfill basic needs, a faculty referred to as nutritive. The next level of the hierarchy was the perceptual one, which was the "aspect of the mind that interprets sensory data" (Ellis et al., 2009, p. 3). The final and highest level was the intellectual one. The intellectual faculty was the only one, per Aristotle's model, unique to humans, with the other two faculty also being present in animals (Ellis et al., 2009). Hippocrates of Cos, also known as the father of western medicine, proposed that humans contained four distinct elemental fluids: blood, phlegm, yellow bile, and black bile, and that if these fluids were out of balance an illness would develop (Ellis et al., 2009). Claudius Galen built upon this idea and thought that if any of these fluids/humors should dominate the others, specific changes in the individual's personality would take place (Ellis et al., 2009). While these earlier explanations are very different from modern conceptions of personality, they do share some commonalities. Even in these earlier definitions, personality is composed of various elements and the dynamic interaction between them, and within a few these ancient conceptions of personality, there is the implicit idea of a structure, or hierarchy, that later theorists employed. In these explanations, there is also an implicit idea of balance being the key to good mental health.

The late 19th century represents the temporal birthplace of psychology. With the advent of the field of psychology, the study of personality became more codified. Wilhelm Wundt established the first psychology lab in Leipzig, Germany in 1878. Wundt is also the origin of the school of psychology known as structural psychology. Structuralism was concerned with the human psyche and characterizing it "in terms of what they considered its fundamental elements or structures" (Hjelle & Ziegler, 1992, p. 360). The structuralists considered sensations, images, and affects as the basic components of the mind (Ellis et al., 2009). The structuralists also were interested in studying the process by which these "elements of the mind combined to form more complex aspects of consciousness" (Ellis et al., 2009, p36). Their method for studying this was through introspection. This approach stands in contrast to the school of psychology that arose in the United States in the late 1890s called functionalism. Functionalism was more focused on the adaptive value of behavior and less concerned with understanding consciousness or its underlying structure (Ellis et al., 2009). Early proponents of this school of thought included William James, John Dewey, James Rowland Angell, and Edward L. Thorndike (Ellis et al., 2009). While introspection was not immediately disregarded as a research tool, the emphasis on behavior and its adaptive value resulted in introspection being de-emphasized and becoming insignificant as a research method for the functionalists. While these were the first psychological approaches to personality, they had a narrow focus compared to the comprehensive scope of the later personality theories.

The first comprehensive model of personality was developed by Sigmund Freud. Freud developed what came to be called psychoanalytic theory, and it set the standard for personality psychology. "Freud proposed a comprehensive explanation of virtually all aspects of human behavior, both individual and collective. In addition, he attempted to explain how behaviors developed in the individual and how the individual develops as a member of the human species" (Ellenberger, 1970 as cited by Ellis et al., 2009, p. 81). Freud developed his theory based on his training in Paris and his clinical work with Josef Breuer. The earliest presentations of this theory were in books that would contain case studies. In 1895, Breuer and Freud published *Studies in Hysteria,* and the two men theorized that their patients' hysterical symptoms were the result of repressed memories of traumatic events (Hjelle & Ziegler, 1992). Freud would continue to publish and develop his theory, which would eventually grow to include "a theory of human development, a theory of mental functioning, and a set of propositions suggesting the ways in which disturbances in mental functioning lead to psychopathology" (Ellis et al., 2009, p. 109). The first part of Freud's theory included a topographical division of the personality. This organization included three levels of consciousness. The first is the conscious level, which consists of a person's current sensations and experiences (Ellis et al., 2009). Freud thought that this was only a small part of our mental life. The next level was the preconscious, which contains all the experiences that one may not be conscious of at the moment, but can be retrieved into awareness. The preconscious also serves as the bridge between the conscious and the unconscious

(Ellis et al., 2009). The unconscious is the largest and last level, and it is hypothesized to contains instinctual drives along with repressed emotions and memories (Ellis et al., 2009). In Freud's conception, these emotions and memories have been repressed because they are threatening to the conscious mind. In addition to these levels of awareness, Freud conceptualized three structures/processes that interacted result in behaviors, and the interaction of these structure would further help characterize personality.

Freud's structural model of the human mind divided the mind into id, ego, and superego. The id refers "exclusively to the primitive, instinctive, and inherited aspects of personality" (Hjelle & Ziegler, 1992, p. 88). Within this formulation, the id is located entirely within the unconscious and is "closely tied to [one's] instinctual biological urges (to eat, sleep, defecate, copulate)" (Hjelle & Ziegler, 1992, p. 88). The id is also primarily motivated by the pleasure principle, which reflects a constant quest for pleasure (Ellis et al., 2009, p. 88). Freud also viewed the id as a mediator between mental and somatic processes, and thought of the id as being "somewhere in direct contact with somatic processes, and takes over from them instinctual needs and gives them mental expression" (as cited in Hjelle & Ziegler, 1992, p. 89). The Ego is considered the negotiator between the demands of the outside world and the person's internal drives. Ellis et al., (2009, p. 90) mention that "Freud saw the ego as a metaphor for brain functions, specifically those located in the motor and sensory cortex". In contrast to the id, the ego functions based on the reality principle. The ego also is capable of secondary

process thinking, which includes rational, conscious, and logical thought. This allows the ego to control impulses and delay the gratification of needs to meet the demands of the external conditions (Ellis et al., 2009). This process results in the ability for the ego to problem solve. The ego is also distinguished by the being only part of the mind in this model that is in contact with the external physical and social reality (Ellis et al., 2009). The Superego is the last part of Freud's conception of the mind (Ellis et al., 2009). Freud stated that this was the last part of human psyche to be developed (Hjelle & Ziegler, 1992). The superego develops through the process of socialization which is the internalization of society's values, norms, and ethics. A strong influence on this process, and how the superego develops, is the child's interaction with the his/her caretakers. Freud further split the superego into two subsystems: the ego-ideal, and the conscience (Hjelle & Ziegler, 1992). The ego-ideal is the ideal image of the self, and represents the rewarding aspect which results in a sense of pride if achieved. The conscience is more punishment oriented and "it includes the capacity for punitive self-evaluation, moral prohibitions, and guilt feelings" (Hjelle & Ziegler, 1992, p. 91), which occur when the child does not meet internal expectations. Each of these components has different drives and motivations, and conflicts between the elements result in certain pathologies, referred to as anxieties. Moral anxiety occurs when the id or ego drive come into conflict with the superego (Hjelle & Ziegler, 1992). Neurotic anxiety occurs when unconscious impulses from the id intrude into the consciousness (Hjelle & Ziegler,

1992). These anxieties result in distress and influence the behaviors and thinking of an individual (Ellis et al., 2009).

Freud's theory of personality also contained a developmental component focused around psychosexual stages. His stages were defined by a specific focus for the person's libidinal or sexual energy. As one moves from state to stage, most of the person's sexual energy is redirected from a less mature focus to a more mature one. Conflicts, inadequate satisfaction, or excessive pleasure at any point during these stages can lead to a fixation where the target of the sexual energy remains attached to this lower level of maturity object (Ellis et al., 2009). This type of progression through stages also provided a model for future theorists to follow.

Another contrasting approach to psychology and personality was the learning-behavioral approach as championed by B.F. Skinner. Skinner believed in an approach that focused on how the environment directly affects the individual's behavior, without the need for any hypothesized internal abstractions. Thus, Skinner did not study a person's inner state, and rather treated humans as a "black box" whose inner works were not subject to empirical inquiry (Hjelle & Abrams, 1992). He also emphasized that this lack of focus on inner working and states of humans was due to the inability to reliably and objectively measure the hypothesized internal phenomena (Hjelle & Abrams, 1992). With this focus on behavior-environment interactions, Skinner's research was heavily characterized by an experimental approach.

Another movement within American psychology was the humanistic/third force movement. It was heavily influenced by existential philosophy, with one of its basic principles being the conceptualization of the individual as an integrated whole (Hjelle & Ziegler, 1992). This view does not lend itself to the analysis of individual components of personality. This movement also emphasized the role of the individual as a much more active agent, as compared to the views of the behaviorists and the psychodynamic perspective. The humanists also maintain the view that human nature was essentially good (Hjelle & Ziegler, 1992). With these assumptions, there are less emphasis on the components of personality. Yet, the humanists still focused on the motivational processes of people and their actions. One of early leaders of the humanistic movement in the 1960s was Abraham Maslow. Maslow conceptualized as a human as a "wanting organism", and he developed a hierarchy of innate needs, which direct much of human behavior. These need states include physiological needs, safety and security needs, belongingness and love needs, self-esteem needs, and self-actualization/personal fulfillment needs (Hjelle & Ziegler, 1992, p. 448). While these are placed in order of priority, Maslow acknowledged that there are exceptions to the arrangement of the hierarchy, and the needs may overlap and a person may be motivated by multiple needs at a time (Hjelle & Ziegler, 1992, p. 449). Maslow also described two categories of human motives; deficit motives and growth motives. Deficit motives are aimed at reducing tension from deficit states that are usually concerned with physiological and safety requirements Growth motives seeks to increase

human movement through new and challenging experiences, which fuels the urge to self-actualize (Hjelle & Ziegler, 1992).

Trait Theories

Another prominent approach to personality and behavior within the field of psychology is trait theory, which is guided by a dispositional perspective. Ellis et al., stated that "trait theory is the single approach to personality theory that is most directly based on and corroborated by research data" (2009, p. 219). This is evidenced by the fact that traits can be operationally defined, and thus allowing for them to be studied with the scientific method. While there are different definitions of traits, most of the definitions tend to share four elements: "Traits are stable within a given individual; traits vary among individuals; traits can be measured; traits are responsible for closely related behaviors" (Ellis et al., 2009, p.220). Gordon Allport was one of the earliest proponents of a trait perspective of personality. He characterized his approach as a blend of humanistic and personalistic approaches. "It is humanistic in its attempt to recognize all aspects of the human being, including the potential for growth, transcendence, and self-realization. It is personalistic in that it seeks to understand and predict the development of the real individual person" (as cited in Hjelle & Ziegler, 1992, p. 239). Trait theories can trace their origin to Allport's dissertation in 1922 which represented the first such study done in the United States on the traits of personality. Allport published several books on his work and co-developed two personality tests; The A-S Reaction Study and A Study of Values (Hjelle &

Ziegler, 1992). Before examining his conception of what a trait is, it is helpful to examine his definition of personality. He conceptualized personality as "the dynamic organization within the individual of those psychophysical systems that determine his characteristic behavior and thoughts" (Allport, 1961, p.28). This definition contains two important elements. First, it frames the person as a dynamic entity, and secondly, it emphasizes the connection between the psychological and the physical processes/structures that cause person's actions and thoughts. With respect to this second element, Allport's (1961) conceptualization of traits allow them to contain "neuro-psychic structure having the capacity to render many stimuli functionally equivalent, and to initiate and guide equivalent (meaningfully consistent) forms of adaptive and expressive behavior" (p. 347). Hjelle and Ziegler simplified Allport's definition of a trait, to see it as a "predisposition to act in the same way in a wide range of situations" (1992, p.242). This results in the trait acting akin to a filter, lending many stimuli the same functional meaning, resulting in generalization. Allport additionally made the argument that traits do not always lead to this generalization, and instead, may also be focal in nature, whereby a particular trait only dictates behavior within certain social situations. To this point, Allport divided traits into common and individual traits. Common traits are characteristics that are "shared by several people within a given culture" (Hjelle & Ziegler, 1992, p. 246). Nonetheless, these common traits can demonstrate diversity in presentation and "are never expressed by any two people in exactly the same way" (Hjelle & Ziegler, 1992). Allport also discussed individual traits which are

particular to the person and do not allow for comparisons among people (Hjelle & Ziegler, 1992). To reduce confusion, these individual traits were later referred to as personal dispositions while the common traits were referred to as traits. Allport later classified these dispositions into three types: cardinal, central, and secondary. A cardinal disposition is pervasive and has such a large influence on an individual that almost "everything a person does can be traced to its influence" (Hjelle & Ziegler, 1992, p. 247). However, due overwhelming nature of this type or disposition, Allport insisted that people rarely have these cardinal dispositions (Hjelle & Ziegler, 1992). In comparison, central dispositions are still generalized and have large effects, but are less all encompassing. They would comprise a person's most salient characteristics, such as what characteristics others would use to describe them. In Allport's research, the number of central dispositions ranged from 5-10 in number per person (Hjelle & Ziegler, 1992). Allport's final tier of dispositions, secondary dispositions, sees these dispositions as less generalizable and less consistent, such as preferences for food, clothing, etc., and specific attitudes. These can be more situationally determined, and Allport commented that a "person must be known quite intimately in order to discern his or her secondary characteristics" (Hjelle Ziegler, 1992, p. 248). Allport, in his work, focused considerably on these personal dispositions, and thus, he leaned toward idiographic research, which studies individual in depth. This may take the form of a person's diaries or letters, or individual interviews, and thus, these efforts do not seek to

compare people to one another, but instead look to describe the unique organization of traits within an individual.

Other trait theorists used a different approach to study personality. In particular, another group looked to empirically determine the number of human traits. There are several assumptions that have proven useful in this empirical search: 1) The dimensions of personality are universal; 2) People have enduring predispositions to respond in consistent ways and that there is a hierarchy to these personality dimensions; and 3) The degree and presence of these traits can be quantitatively measured (Hjelle & Ziegler, 1992). One of the more productive statistical methods used to identify these traits has been factor analysis. Essentially, factor analysis is a statistical process used to locate clusters of variables that correlate strongly with each other. Raymond Cattell was one of the earliest psychologists to employ the technique while studying personality. Cattell originally worked with Charles Spearman, the developer of factor analysis, who notably used the process in the study of intelligence (Ellis et al., 2009). Cattell viewed the primary advantage of factor-analysis as the combined consideration of clinically derived variables and scientific objectivity through an experimental method (Ellis et al., 2009). Cattell also described a distinction between surface traits and source traits (Hjelle & Ziegler, 1992). Surface traits are indicated by a set of behavioral characteristics and are related to one another. Though these surface traits may be causally observed, they lie on the "surface of personality" (Ellis et al., 2009, p.231), and, as such, do not explain the underlying and fundamental structure of

personality. Cattell believed that surface traits were the manifestation of underlying source traits, and that these sources were the building blocks of personality, and what ultimately determined the consistency in people's behavior (Hjelle & Ziegler, 1992). Cattell used three types of data sources for his factor analysis: life record data (L-data), self-rating questionnaire data (Q-data), and objective test data (T-data or OT-data) (Hjelle & Ziegler, 1992). L-data consisted of measurements of a person's behavior in everyday situations, which could include peer interactions, scholastic performance data, and trait ratings provided people who knew the individual well (Hjelle & Ziegler, 1992). T-data came from objectively scored tasks within special situations, and these data were resistant to faking because of the participants being unaware of the evaluated dimension. The final source of data, Q-data, came from a person's self-ratings and reflect the person's introspections (Hjelle & Ziegler, 1992). Cattell developed utilized all three of these data sources to develop his most well-known instrument, the Sixteen Personality Factor Questionnaire (16 PF) (Cattell et al., 1970). From these data sources Cattell concluded that there were 16 source traits, or factors. The first 12 factors emerged across the L and Q data data sources, while the last four factors were found only in the Q-data (Hjelle & Ziegler, 1992). The 16 source traits are presented in order of the amount of variance each factor accounted for within the produced factor analyses. These traits are presented in Appendix A, with descriptions of the ends on the continuum.

The Sixteen Personality Factor Questionnaire (16PF)

One of the fruits of Cattell's research into personality was the Sixteen Personality Factor Questionnaire (16PF), originally published in 1949. Raymond Cattell's approach to developing the test was to base the assessment on fundamental building blocks of personality which he identified through factor analysis. This process began with the reduction of 17,953 trait words used by Allport and Odbert in 1936. Cattell's goal was to reduce these thousands of adjectives to optimal set of categories that preserved as much information as possible from the original list (Karson, Karson, & O'Dell, 1997). With the factor-analytic studies of behavior ratings and questionnaire data, Cattell was able to reduce the descriptors to 16 underlying dimensions, represented as unitary traits. There have been additional studies that have replicated Cattell's findings of a basic structure of 16 traits (Cattell & Krug, 1986). The 16PF has been used as an instrument for variety of uses, including the assessment of educational achievement, creativity, leadership, interpersonal skills, marital adjustment, and psychological adjustment. It has also been used in a variety of settings including research, educational, clinical and counseling, and industrial and organizational settings. As a personality measure it was not designed to "solve a clinical problem" (Karson, Karson O'Dell, 1997, p.3). This focus on the basic measure of personality separates the 16PF from assessments such as the Millon Clinical Multiaxial Inventory (MCMI) or Minnesota Multiphasic Personality Inventory (MMPI), which focus on classifying psychopathology. Per Heather Cattell (1989), the increasing

use of the 16PF as a measure in clinical space comes from the demands of managed care, the increasing use of therapeutic services by the "normal" population, the increased emphasis on ordinary personality traits in diagnosing clinical problems, and its use as a potential communication tool between the clinician and the client. Additionally, the 16PF has been translated into over 40 languages and research has confirmed the trait structure across cultures including France, Italy, New Zealand, Chile, Germany and Japan (Conn & Rieke, 1998). Because of these factors, the 16PF has become one of the most frequently administered and recommended personality questionnaires (Piotrowski & Keller, 1989) and one of the most often referenced in research articles (Graham & Lilly, 1984).

The 16PF has undergone 4 revisions since its first publication with the latest edition published in 1993. The fourth revision was conducted in 1988 and included re-standardizing on current population sample and updating item content. Efforts were made to improve internal consistency reliabilities of certain factors [M, N, Q1] and address the relative high scale intercorrelations between some primary factors [C, O, and Q4] that contribute to the same global factor (Conn & Rieke, 1998). The update of the item content came from selecting the "best items" from 5 forms of 16PF {Forms A, B, C, D and Clinical Analysis Questionnaire [CAQ]} and consolidate them into one new form (Conn & Rieke, 1998). This consolidation process came through four factor-analytic studies and a series of eight criteria that the items had to meet. These criteria were designed to reduce overlap between scales, improve comprehension, and remove bias. The original second order factors

emerged from the factor analysis using the national sample. While five, six, and seven factor solutions were examined, it was determined that the six-factor solution yielded the best simple structure (Conn & Rieke, 1998). This six-factor solution accounted for 70% of the total variance. In the fifth edition of the 16PF, these second-order factors were called global factors. This analysis yielded the five global factors of extraversion, anxiety, tough-mindedness, independence, and self-control (Conn & Rieke, 1998). The additional factor from the analysis was defined by a high loading factor on Reasoning and lower loading on Vigilance and was not used as a global factor. The test authors also established a minimum loading of .30 for a primary factor to be included in a global factor's calculation.

The Extraversion global factor compares extraversion to introversion. Heather Cattell (1989) related this construct to Jung's construct of the same name and made the comparison to Freud's "subject (self)/object (external) polarities" (p. 308). Heather Cattell further elaborated that Freud's polarity contrasted pleasure from the environment and pleasure from one's own ideas and imagination (Cattell, 1989). This mirrors the internal versus external focus of the global factor. The primary factors that contribute to this global factor are: Warmth [A+], Liveliness [F+], Social Boldness [H+], Privateness [N-] and Self-Reliance [Q2-] (Conn & Rieke, 1998).

The next global factor, Anxiety, attempts to account for the difference in intensity of discomfort from external threats and internal stimuli that people experience (Cattell, 1989). In the fourth edition of the 16PF, six primary scales

contributed to this second order scale. In the fifth edition, four primary factor scales contribute to the global factor: Emotional Stability [C-], Vigilance [L+], Apprehension [O+] and Tension [Q4+]. The scales of Social Boldness [H] and Perfection [Q3] were omitted from the new equation for anxiety because their loadings were below the required .30 cutoff (Conn & Rieke 1998).

The third global factor of Tough-Mindedness reflects a renaming process. In previous editions, it was labelled Tough Poise. This name change occurred to reduce confusion "in interpreting the concept of tough poise" (Conn & Rieke, 1998). This global factor removed primary scales from the previous version that contributed at a very low level or were gender specific (Conn & Rieke, 1998). The contributing primary factors to this global factor in the 16PF fifth edition are: Warmth [A-], Sensitivity [I-], Abstractedness [M-], and Openness to Change [Q1-] (Conn & Rieke, 1998).

Independence is the next global factor. Contributing primary factors included: Dominance (E+), Social Boldness (H+), Vigilance (l+), and Openness to Change [Q1+]. These factors were unchanged from the fourth edition, however, in the past several other factors also appeared [Factors G, M, N, Q2] but were omitted from the fifth edition because they fell below the .30 cutoff (Conn & Rieke 1998).

The last of the included global factors is Self-Control, which was labeled as 'Control' in previous editions. The "self" was added to emphasize the scale's focus on one's own thoughts, feelings, and behaviors (Conn & Rieke 1998). Included contributing primary factors were: Liveliness [F-], Rule-Consciousness [G+],

Abstractedness [M-], and Perfectionism [Q3+]. In previous editions Abstractedness [M] did not contribute to this global factor; however, with the revisions to the scale in the fifth edition, the scale has been more clearly defined and now contributes to it (Conn & Rieke 1998).

The 16PF fifth edition as an assessment measure consists of 185 items, with each primary factor containing between 10 to 15 items and 12 items comprising an impression management (IM) scale (Conn & Rieke 1998). In addition to the impression management scale, the 16PF contains two more validity scales, an Acquiescence (ACQ) scale and an Infrequency (INF) scale. The individual's raw scores on the factors are then converted into a standard ten (sten) scores and plotted on the profile sheet. The test publishers estimate that the administration time for the paper version ranges from 35-50 minutes. Additionally, for the fifth edition, the Reasoning (B) scale items are placed at the end of the test and separated from the personality items with separate directions because the nature of these questions is different, with distinct right or wrong answers opposed to the personality items. To further broaden the potential use of the 16PF, the test authors improved the readability of the test, placing it at a fifth-grade level for the fifth edition opposed to the seventh-grade level for the fourth edition.

With this strong theoretical backing, broad applicability, and a strong research presence, the 16PF was chosen as a basis on which to develop the Scale of Accurate Personality Prediction.

The Scale of Accurate Personality Prediction (SAPP)

As previously mentioned, measuring self-knowledge has been a complicated task, and few measures exist to address this need. The Scale of Accurate Personality Prediction (SAPP) was developed by Miller in 2000 to measure the accuracy with which an individual can self-predict personality traits. To accomplish this, individuals first complete the 16PF, then each individual rates him or herself (from 1 to 10) on the 16PF fifth Edition Individual Record Form, by using the High and Low score Meanings for each of the 21 factors (see Appendix B).

The SAPP score is then derived by subtracting the predicted scores (PS) from the obtained scores (OS) for each of the primary and secondary/global factors, calculating the absolute value of these differences, the summing the results across all 21 scales (Miller, 2000). The SAPP formula is included below:

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

In the formula, OS stands for obtained score and the letter or letters following it indicating the corresponding factor (A for Factor A, which is Warmth). Likewise, PS represents predicted score with the letter or letters that follow

indicating the corresponding factor. The SAPP includes a range of possible scores. A score of zero is the lowest possible score and indicates optimal accuracy; or in other words, no discrepancy between the any of the obtained or predicted score pairs emerges. The highest possible score is 189, which represents the poorest predictive ability. In Miller's original sample, score ranged from 18 to 79 with a mean of 42.07 and a standard deviation of 11.74. McElligott (2014) conducted a replication of Miller's study, using a larger archival dataset. With 607 participants, McElligott's data yielded a mean SAPP score of 43.14 and a standard deviation of 13.83, with the scores ranging from 18 to 91. McElligott then conducted a t-test to compare these results to Miller's original study and found them to be not statistically different. McElligott also reversed the SAPP scoring, by subtracting each obtained SAPP score from 189. This reversal and linear transformation, allowed for higher SAPP scores to now be associated with more accurate self-prediction, and lower scores with less levels of accurate self-prediction. In Miller's original study, Miller also examined which personality factors were the best indicators that accounted for more accurate predictions of self, through regression analysis. She found that high and low scorers had significantly different obtained scores on the following scales: Warmth (A); Reasoning (B); Liveliness (F); Sensitivity (I); Vigilance (L); Privateness (N); and Openness to Change (Q). As described by Silva (2011), Miller found that low SAPP scorers tended to be warm, have abstract reasoning, lively, sensitive, trusting, forthright, open to change, outgoing, and intuitive in contrast to high SAPP scorers who tend toward a

reserved disposition, concrete in reasoning, introversion, and a lack in empathy. Additionally, Miller's results indicated that the Tough-Mindedness (TM-) global factor was the highest predictor of the sample's ability to predict their SAPP score followed by Reasoning (B+), Independence (IN-), Tension (Q4+), and Anxiety (AX-). In 2015, Mazur performed a replication study using a large archival sample size. Mazur conducted linear regressions on the data and found that the most predictive primary scale was Suspiciousness (L-). Mazur results also indicated that accurate knowledge of Emotional Stability (C-), Sensitivity (I+), and Tension (Q4+) were also significant predictors of the SAPP score.

Validation of the SAPP

An important element of test development is the validation of the test. The concept of test validity seeks establish that the test accurately measures what it intends to measure. This validity can be established through several means. One potential method of establishing the validity of a measure is through establishing construct validity. Within psychology, construct validity is often demonstrated through measuring convergent and discriminant validity. Convergent validity measures the degree in which two constructs or measures are related and discriminant validity measures the degree that unrelated constructs or measures are in fact unrelated. In 2001, Hood sought to establish construct validity through measuring both convergent validity and discriminant validity. To establish convergent validity, Hood attempted to correlate the results of the SAPP with a score on the Private Self-Consciousness subscale on the Self Consciousness Scale.

For discriminant validity, Hood attempted to compare the individuals SAPP score with their scores on the Tennessee Self-Concept Scale 2, which presumes to measure the construct of self-esteem. With a final sample size of 48, Hood found that there was no significant correlation with either the Private Self-Consciousness subscale ($r=-.030$, $p> .05$) or the Tennessee Self-Concept Scale-2 ($r = .188$, $p> .05$). While a non-significant correlation is expected in the case of the Tennessee Self-Concept Scale, for the convergent validity Hood's results did not support the hypothesis. In the discussion, Hood suggested that Private Self-Conscious scale measures self-awareness, which is separate from the construct of self-knowledge. In 2003, Glywasky also attempted to study construct validity with Hood's procedures and expanding the sample size to 211 participants. However, increasing the sample did not result in a correlation between Private Self-Consciousness and the SAPP. This lends support to the potential that the SAPP and the Private Self-Consciousness subscale are measuring different constructs.

In 2005, Hickey attempted to establish convergent validity for the SAPP through comparing the subjects' SAPP scores to family members' predictions of the subject's personality traits. A measure of concordance was used to indicate the amount of agreement between the raters. The hypothesis of the study, was that the concordance measure would be correlated with the SAPP score to indicate similarities in prediction. Additionally, a high SAPP score group was compared to a low SAPP score, and it was predicted that there would be a significant difference between the two SAPP groups on the concordance measure. The results from

Hickey's study indicated that the SAPP scores to the concordance measure had a positive correlation, although the finding was short of being statistically significant ($r=.302$, $p<.09$). Likewise, the comparison between a high SAPP score group and a low SAPP score group did not yield any significant findings. In 2007, Blankemeier attempted a replication of Hickey's 2005 study. Blankemeier employed a similar procedure while increasing the pool of target subjects to 51. Blankemeier found a significant yet low correlation between the SAPP scores and the concordance measure ($r=.283$, $p<.05$).

Layton conducted a similar convergent validity study in 2005 to Hickey. Using the same concordance measure, Layton utilized the subjects' peers instead of family members to gain predictive ratings. Layton found no significant correlation between the subjects SAPP scores and the concordance measure score ($r=.095$, $p>.05$). In 2006, Wolf attempted to replicate Layton's study once again utilizing peers for use in the concordance measure. Wolf found a significant correlation between the SAPP scores and the concordance measure scores ($r = .419$, $p<.05$). Due to Wolf's sample size, she was unable to perform inferential testing to compare the correlations of the high SAPP score group with a low SAPP score group on the concordance measure, but she did report that the mean differences were in the predicted direction.

Additional construct validation studies were also performed. In 2002, Anderson attempted to correlate the results of the SAPP with the Self-Monitoring Scale. Anderson hypothesized that lower scorers on the SAPP would correlate

highly with high self-monitors. It was conceptualized that behaviors associated with high self-monitors, such as more self-awareness and attunement to situational cues, would help to build greater self-knowledge. With a sample size of 77 participants, the participants were categorized into high self-monitor group and a low self-monitor group. Comparing the groups SAPP scores with an independent t-test revealed no significant difference ($t= 1.41$. $P >.05$). Anderson then proceeded to compare the SAPP scores with the raw score of the Self-Monitoring Scale and found no significant correlation ($r=.001$, $p>.05$). Anderson's conclusion that the SAPP score did not relate to an individual's sensitivity to others or their ability and willingness to adjust behaviors in different social contexts. In 2002, Winter compared two groups that were identified a priori to be different on their ability to "know themselves" to provide construct validity for the SAPP. Winter compared graduate psychology students to graduate engineering students, and hypothesized that the psychology students would be better at predicting their personality traits and thus have lower SAPP score than the engineering students. Winter used t-tests to compare the two groups and no significant differences were found ($t (29) = .68$, $p>.05$). in 2006, Grossenbacher attempted to expand on Winter's study with the addition of participants who had obtained degrees and were practicing in the respective fields. Grossenbacher findings revealed a significant difference between the two groups ($t = -4.247$, $p \leq .01$), indicating that the graduate psychology students and practitioners group demonstrated more accurate self-prediction of personality traits that the group consisting of graduate engineering students and

engineers. These studies when taken together have established a degree of construct validation for the SAPP.

Reliability

The Standards for Educational and Psychological Testing describes reliability as the "consistency of measurements when the testing procedure is repeated on a population of individuals or groups" (American Educational Research Association, American Psychological Association, & National Council on Measurement Education, 2004, p. 25). When referring to reliability of test measures, researchers discuss three different types of reliability: interrater reliability, internal consistency reliability, and test-retest reliability. The interrater reliability focuses on the degree of consensus in results between different raters. These raters must be independent of one another and unaware of the other rater's ratings. However, due to the nature of the measure allowing an examinee to measure their self-knowledge, the interrater reliability is not an appropriate gauge of reliability for the SAPP. Internal consistency reliability measures items within the test are measuring the same underlying construct. However, due to the derived and unitary nature of the SAPP this is also not an appropriate gauge of reliability. Test-retest reliability examines if the test's results are consistent over time. This can be measured by administering the test to the same participant with a significant period separating the testing sessions and comparing the results. Due to the nature of the SAPP, the test-retest reliability is the most useful indicator of reliability and stability.

The test-retest reliability of the SAPP has been examined in the past. Silva (2011) was first to study the test-retest reliability of the SAPP through the successive administration of the measure. Silva's hypothesis was that an individual's SAPP score would remain stable over a two-week period. From the initially distributed 100 packets to participants, 62 were returned, and the population largely consisted of psychology graduate students in the Southeastern United States. Each participant rated themselves on a 16PF profile sheet and completed the 16PF each trial. Using these two sources, individual SAPP scores were calculated. Silva's results indicated a significant correlation of .397. However, this was considered low for an acceptable level of reliability and Silva was unable to find an explanation for the low SAPP reliability results. Silva did indicate that the small sample size and lack of control testing setting may have played a role.

Other studies have also followed up this examination of the test-retest reliability of the SAPP. Three studies conducted in 2012 examined test-retest reliability of the SAPP utilizing different intervals of testing. The studies were conducted in a similar manner to the Silva 2011 study, with an initial target of 100 participants. However, these three studies also contained a methodological change, utilizing online versions of the 16PF profile sheet and the 16PF questionnaire to collect data. Hirsch's (2012) study utilized a two-week interval between testing trials. With 58 participants in the Hirsch's final dataset, a statistical analysis revealed a significant moderate correlation ($r = .566$, $p < .01$) between the two

derived SAPP scores. Sverdlova's (2012) study increased the interval from two weeks to four weeks. With 58 participants completing both trials, Sverdlova's statistical analysis revealed a significant correlation of .466. Elghossain's (2012) study examined the test-rest reliability using a six-week interval period. Elghossian's study had a final sample size of 47 participants and yielded a statistically strong Pearson correlation ($r = .772$, $p<.01$) between the two derived SAPP scores. All three studies yielded higher correlations than the initial Silva 2011 study, providing evidence for the test-retest reliability of the SAPP. However, to be further confident in the strength of these results, additional research is necessary.

The Importance of Measuring the Reliability of the SAPP

As previously mentioned, the reliability of a measure is an important component in the development of a new test or measure. If the reliability of a measure is questionable, the measure will lack viability as testing instrument due to the lack of consistency or meaning of the information produced. The aim of this study was to test the six-week test-retest reliability of the SAPP. Two other concurrent studies are underway to test the two-week and four-week intervals, respectively. If proven reliable, the SAPP may serve as a tool to guide treatment in mental context and for researchers to further explore the concept of self-knowledge.

Methods

Subjects:

Subjects of this study were asked to participate on a voluntary basis. One hundred packets were targeted for distribution to a nonrandomized sample from the United States during the initial testing trial. Participants who respond to the initial trial were tracked in a database file and six weeks later received a second testing trial.

Instruments:

Instruments that were utilized in this study included the 16PF Fifth Edition (electronic version), an electronic survey of 16PF Fifth Edition Individual Record Form, a demographic survey, and a letter of instruction.

Procedure:

Each potential subject was asked to engage in two testing trials separated by an interval of six weeks. Test packets including typed instructions was given to the participants. Participants were instructed to complete the two surveys. The link for the second survey was provided at the end of the first survey

The initial trial email contained the test administration instructions. After completing the 16 PF, the subjects were required to fill out a blank 16PF Fifth

Edition Individual Record Form in an electronic survey format. This form contained the subject self-ratings on the continuums for the sixteen personality factors and the five global factors (see Appendix B). As in the original study by Miller (2000), the SAPP score was derived for each participant by summing the amount of the absolute differences between the predicted score and the obtained scores for each of the twenty-one scales. Scores on the SAPP range from 0, which indicates the poorest accuracy of self-prediction, to 189, which indicates the best predictive ability.

The participants' SAPP scores were archived during a six-week period. After this six-week period, subjects were sent an email with further instructions and a second activation code. This packet contained the same instructions and. The participants once again completed the online version of the 16PF and fill in a blank 16PF Fifth Edition Individual Record Form through an online survey, based on their evaluations and perceptions of themselves on each of the sixteen personality factors and five global factors. The second testing trial was scored in the same manner as the initial testing trial. The second trial SAPP score will be then compared to the participants SAPP score from the first trial and analyzed.

Data Analyses

In accordance with the previous analyses of the test-retest reliability of the SAPP, the main data analysis consisted of a Pearson correlation performed on the data set. In addition, descriptive demographic statistics and sten score results will be found and reported for each of the testing trials.

Hypothesis:

The hypothesis tested by this study is that a participant's SAPP score will remain stable over time, thus when the participant is tested twice, with a period of six weeks between the testing sessions, a statistically significant Pearson correlation coefficient should be produced. This hypothesis has been supported by previous studies and this study is being conduct to add to the statistical strength of this conclusion.

Results

Subjects for the study participated on a voluntary basis. The researcher solicited participation through various channels including social media and in-person requests. As a result, 59 people requested to participate in the study. From these initial volunteers, 34 participants completed just the 16PF and 33 participants completed the 16PF and the record form in a timely manner. After the second trial after a six-week interval, 25 participants completed just the 16PF, with 22 participants completing both the 16PF and the additional 16PF record form for the survey. These 22 data sets were used in the findings of the study.

The mean age of the 22 final participants was 37.59 years with a standard deviation of 16.41 (see Table 1). Most participants reported having a master's degree or higher (17 participants). The final participants included 13 self-identified males and 9 self-identified females. None of the participants indicated Hispanic origin. The majority of the participants were Caucasian (11 participants, 50.0%), with the remaining participants identifying as Black or African American (4 participants, 18.2%), Asian (4 participants, 18.2%) or Other (3 participants, 13.6%). Geographically, the participants were concentrated in the Southeast of the United States (11 participants, 50.0%). The remaining participants were located in the Northeast of the United States (7 participants, 31.8%), Western region of the United States (2 participants, 9.1%) and the Caribbean (2 participants, 9.1%). Further information on the study's demographics are presented in Table 1.

Table 1. Demographics

Demographic Category		Current Sample Percent (N=22)	SAPP Database Percent[a] (N=645)	Normative Sample Percent[b] (N=2500)
Gender	Male	59.1%	42.0%	48.7%
	Female	40.9%	58.0%	51.3%
Race[c]	African American/Black	18.2%	2.3%	12.1%
	Asian	18.2%	9.3%	2.9%
	Caucasian	50.0%	71%	80.2%
	Native American	0%	.2%	1.0%
	Other	13.6%	5.3%	3.8%
	Hispanic Origin	0%	11.9%	9.0%
Age Group	15 to 17	0%	1.2%	4.6%
	18 to 24	13.6%	51.3%	13.8%
	25 to 44	54.5%	34.4%	41.7%
	45 to 54	4.5%	5.3%	12.9%
	55 to 64	22.7%	5.9%	10.8%
	65 and older	4.5%	1.9%	16.2%
Education Level	≤12 years	0%	5.0%	61.5%
	12.5-16 years	22.7%	55.8%	22.7%
	16+ years	77.3%	39.2%	15.8%

Marital Status	Single	66.7%	72.9%
	Married	31.8%	20.6%
	Divorced	4.5%	4.8%
	Separated	0%	1.1%
	Widowed	0%	.6%
Geographic Location	Southeast	50.0%	78.9%
	Southwest	0%	3.8%
	Northeast	31.8%	13.1%
	Midwest	0%	4.0%
	West	9.1%	0%
	Canada	0%	.2%
	Caribbean	9.1%	0%

[a] From the SAPP Database, which is an accumulation of data from multiple studies on the SAPP.

[b] From "Characteristics of the Norm Sample" by S.R. Conn & M.L. Rieke, in press. In S.R. Conn & M. L. Rieke (Eds.). *The 16PF Fifth Edition Technical Manual,* Champaign, IL: Institute for Personality and Ability Testing, Inc. Please note that information regarding marital status and geographic location is not available.

[c] 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.

[d] Totals may be less than 100% since participants of the present study were allowed to choose not to answer demographic information.

Test-retest reliability of the SAPP was examined by deriving SAPP scores for all subjects during their initial testing trial and again 6 weeks later during their second testing trial. The initial testing trial contained a mean SAPP score of 145.72 with a standard deviation 10.15, and an inclusive range from 125.00 to 163.00. The secondary trial resulted in a mean SAPP score of 145.18, with a standard deviation 9.32 and a range from 128.00 to 163.00. A Pearson correlation was performed resulting in an r = .572, which was significant at the $p<0.01$ level indicating a

statistically significant correlation. This value is considered of statistically moderate strength. This result supports the hypothesis, that SAPP is a reliable measure as measured by test-retest reliability. The means and standard deviations of the obtained primary factor scores from this study and the corresponding 16PF normative data are reported in Table 2. The means and standard deviations obtained from the currently sample are similar to those in the 16PF normative sample. The largest difference between the primary factor means obtained from the current study and the normative sample was on Primary Factor A (warmth) with a 1.76 difference between the sten scores. The smallest difference between primary factor means occurred between the current study and normative sample means for Factor O (apprehension) with a difference of 0.02. Table 3 presents the test-retest reliability data of the current study along with the test-retest reliability data from the 16PF technical manual. The test-retest correlation of the current study (with a six-week interval), were in general, comparable to or better than the reliability correlations of normative sample for the two month interval. However, the reliability correlations for primary factors B (reasoning), G (rule-consciousness), and Q3 (perfectionism) from the currently were lower than they corresponding values from the 16PF technical manual.

Table 2. Primary Scale Sten Score Means and Standard Deviations

Primary Factor	Mean	S.D	Mean*	S.D.*
A	4.05	1.53	5.81	1.79
B	7.00	1.41	5.58	1.89
C	4.63	1.65	5.34	1.79
E	4.41	2.06	5.23	1.70
F	4.73	1.78	5.64	1.85

G	4.27	1.45	5.54	1.80
H	4.95	1.81	5.63	1.96
I	5.59	1.65	5.36	1.82
L	5.82	1.65	5.67	1.90
M	5.73	1.91	5.50	1.76
N	6.36	1.81	5.36	1.82
O	5.64	2.08	5.66	1.77
Q1	6.09	1.66	5.67	1.77
Q2	7.09	1.69	5.52	1.83
Q3	4.73	1.45	5.43	1.84
Q4	5.41	1.33	5.30	1.66

Note The Means and Standard Deviations on the left side of the table correspond to the scores obtained during Trial 1 of the present study. Those in the starred columns are reported directly from the 16PF Fifth Edition Norm Supplement, Release 2002 by Catherine C. Maraist and Mary T. Russell.

Table 3. Test-Retest Reliability Data of the 16PF, Fifth Edition

		Test-Retest Interval		
Primary Factor		**Six-Week (N=22)**	**Two-Week[a] (N=204)**	**Two Month[a] (N=159)**
A	Warmth	.76	.83	.77
B	Reasoning	.61	.69	.65
C	Emotional Stability	.77	.75	.67
E	Dominance	.84	.77	.69
F	Liveliness	.91	.82	.69
G	Rule-Consciousness	.72	.80	.76
H	Social Boldness	.89	.87	.79
I	Sensitivity	.93	.82	.76
L	Vigilance	.77	.76	.56
M	Abstractedness	.87	.84	.67
N	Privateness	.81	.77	.70
O	Apprehension	.77	.79	.64

Q1	Openness to Change	.87	.83	.70
Q2	Self-Reliance	.81	.86	.69
Q3	Perfectionism	.67	.80	.77
Q4	Tension	.79	.78	.68
SAPP		.572		
Global Factor				
Extraversion		.85	.91	.80
Anxiety		.73	.84	.70
Tough- Mindedness		.90	.87	.82
Independence		.89	.84	.81
Self-Control		.80	.87	.79

[a]From "Comparison of the 16PF Fifth Edition and Form A (Fourth Edition)" by S.R. Conn, in press. In S.R. Conn & M.L. Rieke (Eds.), *The 16PF Fifth Edition Technical Manual.* Champaign, IL: Institute for Personality and Ability Test, Inc.

Discussion

This study sought to replicate the work of Elghossain's 2012 study examining the test-retest reliability of the SAPP with a six-week interval between trials. To study this, a Pearson correlation was derived from the SAPP scores for the initial and the six-week test Administration. A correlation of r=.572 was found which is lower than the results found in the initial six-week study. However, the result was still statistically significant and of moderate strength, indicating that the SAPP is a reliable measure of self-knowledge. The overall possible strength of the test-retest reliability of the SAPP is potentially limited by the test-retest reliability of the 16PF scales themselves. As a derived measured from the 16PF, the variability of the SAPP can be impacted the variability from 16PF's individual scales, and we would not expect the SAPP to demonstrate a greater reliability than the 16PF. The test-retest reliability of the sample on the individual scales along with the test-retest reliability data from the 16PF manual is presented in Table 3. The test-retest performance of the sample was in line with or higher than the correlations present in the original 16PF standardization sample studies.

The 2012 study was itself a study to build on the initial test-retest reliability study as performed by Silva in 2011. Silva's results found a low though still significant test-retest reliability (r = .397). Elghossain's study indicated a statistically stronger correlation (r = .772, $p<.01$). The current result lies in

between the results of the Elghossain and Silva studies. Taken as a whole, this data further suggests the SAPP appears to be a reliable measure.

The Silva and Elghossain studies identified small sample size as a primary weakness. Silva attributed the low strength of her test-retest value to her small sample size of 62 participants. Elghossain's study yielded a strong and significant correlation, however she also commented "it is important to caution the possibility of large effect size, which may have increased the probability of a false positive result" (Elghossain, 2012, p. 39). This study was conducted in order to address this, yet unfortunately this study has also experienced this limitation. This study concluded with a small final sample size of 22, which is smaller than the preceding studies. While the correlation value is statistically significant, it still experiences the possibility that small sample size has enlarged the effect. However, in combination with the previous studies, the general evidence suggests that these results are not merely statistically aberrations and further indicates the SAPP is a stable and reliable measure. However, a study with a larger sample would more definitively confirm the stability of test-retest reliability of the SAPP.

Sample size may have been affected multiple factors including, the time commitment of the study, the long delay interval, and the complexity of the task. A few participants indicated that they were interested in the study, however then proceeded to decline after the time commitment. Similarly, the six-week interval also can potentially add to the attrition rate. Unfortunately, both the time commitment and six-week delay are unavoidable aspects to the nature of the

measure and of the nature of the test-retest reliability question. The impact of these factors could be potentially reduced with a captive participant base, where structured time was set aside to complete the 16PF and the individual record form. This may occur in a classroom or computer lab. The complexity of the study was impacted by the chosen delivery method. With the support of PSI/PAN, the current owners of the 16PF, this study utilized online test administration for 16PF. This online administration would allow for remote data collection and a wider potential participant base. However, the complexity of the task was increased by this data collection method. Due to the nature of the PSI's test administration system, the researcher was unable to integrate the 16PF individual record form with the administration of the 16PF. As a result, the researcher recreated the 16PF individual record form on an external survey site, and a link to this survey was included at the end of the 16PF. There are data that indicated that a few participants (N=4) completed the 16PF but did not continue to the record form portion of the study. If future studies were able to streamline to process and better integrate the 16PF and the 16PF individual record form into the same system, this would potentially reduce the attrition rate.

The final sample was largely unrepresentative of the general population. Individuals of higher educational achievement were overrepresented. The majority of the participants had obtained a master's degree and none of the participants reported a high school diploma as a terminal degree. In the sample, Caucasians were underrepresented compared to the general population of the United States.

Additionally, the sample included a couple of participants who are citizens of other countries. The initial studies did not include participants from this region of the world, and this could potentially serve as an area for future study. While this sample may have been unrepresentative of the general population, it still contained ethnic diversity with half of the participants indicating an ethnicity other than Caucasian. While the sample was not representative of the population, it does suggest further promising areas of study.

One potential future study would be to examine the impact of educational achievements with scores on the SAPP. This study contained a sample that was positively skewed with regards to education level. A potential hypothesis would be that with higher educational achievements there would be higher self-knowledge and as a result the scores on the SAPP would be greater than those with less educational achievement. An additional study would potentially focus more on the generalizability of the measure. This study did include participants currently living in the Caribbean and participants who have grown up in the Caribbean. The 16PF is available in 17 additional languages other than English. A comparison study that examines SAPP scores among different populations not located in the United States using the localized versions of the 16PF is a potential area for future research. Zeng's 2015 study is similar, in that it compared a sample containing an Asian population with a random sample from the existing database. Non-statistical difference was found, which is a promising result for a future comparison study and would speak to the cross-cultural generalizability of the SAPP.

While they were limitations with this current study, the data gathered contributes evidence to the reliability of the SAPP as a measure. A significant correlation was found indicating adequate test-retest reliability for the SAPP. This indicates that the SAPP scale is a consistent and stable measure of self-knowledge. Taken together with the data from previous studies and results regarding the validity and reliability it suggests that the scale is appropriate for its intended use. The SAPP as a measure would allow clinicians to gather important data about the state of self-knowledge of their clients which would aid in areas such as treatment planning. Additionally, as a validity and reliable measure, the SAPP can serve as a platform for exciting directions for additional research.

References

Allport, G.W. (1961). *Pattern and Growth in personality.* New York: Holt, Rinehart, and Winston.

Anderson, A. M. (2002). A validation study of the Scale of Accurate Personality Prediction using the Self-Monitoring Scale. Unpublished master's thesis, Florida Institute of Technology.

Blankemeier, C.E. (2007). A validation of the Scale of Accurate Personality Prediction using self versus family member comparisons. Unpublished master's thesis, Florida Institute of Technology.

Brown, J. D. (1991). Accuracy and bias in self-knowledge. In C. R. Snyder & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective. Pergamon general psychology series* (Vol. 162, pp. 158–178). Elmsford, NY: Pergamon.

Cattell, H. B. (1989). *The 16PF: personality in depth*. Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R.B., Eber, H.W., & Tatsuoka, M. M. (1970). *Handbook for the 16 personality factor questionnaire.* Champaign, IL: IPAT

Cattell, R.B., & Krug S. E. (1986). The number of factors in the 16PF: A review of the evidence with special emphasis on methodological problems. *Educational and Psychological Measurement, 46,* 509-522

Conn, S. R., & Rieke, M. L. (1998). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 52, 177–193.

Elghossain, R. (2012). Examining the test-retest reliability of the Scale of Accurate Personality Prediction. Unpublished master's thesis. Florida Institute of Technology.

Ellis, A., Abrams, M., & Abrams, L. (2009). *Personality theories: critical perspectives*. Los Angeles: SAGE Publications.

Engler, B. (2003). *Personality theories: an introduction* (6th ed.). Boston: Houghton Mifflin.

Graham, J.R. & Lilly, R.S. (1984). Psychological Testing. Englewood Cliffs/ NJ: Prentice Hall.

Grossenbacher, A. S. (2006). A validation of the Scale of Accurate Personality Prediction using a priori comparison. Unpublished master's thesis, Florida Institute of Technology.

Glywasky., K. E. (2003). Re-evaluation of the Scale of Accurate Personality Prediction. Unpublished master's thesis, Florida Institute of Technology.

Hart, D., Matsuba., M., K., (2012) The Development of Self-Knowledge. In Vazire, S., & Wilson, T. D. (Eds.). *Handbook of self-knowledge*. New York: Guilford Press.

Hickey, J. (2005). A validation of the Scale of Accurate Personality Prediction using self versus family comparisons. Unpublished master's thesis, Florida Institute of Technology.

Hirsch, S. (2012). Test-retest reliability of the Scale of Accurate Personality Prediction. Unpublished master's thesis. Florida Institute of Technology.

Hood, C.A. (2001). A measure of the accuracy of self-predicted personality traits. Unpublished master's thesis, Florida Institute of Technology.

Hjelle, L. A., & Ziegler, D. J. (1992). *Personality theories: basic assumptions, research, and applications* (3rd ed.). New York: McGraw-Hill.

Karson, M., Karson, S., & O'Dell, J. W. (1997). *16PF interpretation in clinical practice: a guide to the fifth edition*. Champaign, IL: Institute for Personality and Ability Testing.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology,77*(6), 1121-1134. doi:10.1037//0022-3514.77.6.1121

Layton, E.E. (2005). A validation of the Scale of Accurate Personality Prediction using self versus other comparisons. Unpublished master's thesis, Florida Institute of Technology.

Maddi, S. R. (1980). *Personality theories: a comparative analysis*. Homewood, IL: Dorsey Press.

Mazur. L. A. (2015). An Examination of 16 PF Primary Factors as Predictors of the Scale of Accurate Personality Prediction (SAPP). Unpublished Doctoral thesis, Florida Institute of Technology

McElligott, M. (2014). Standardization Efforts of the 16-PF Based Scale of Accurate Personality Prediction. Unpublished Doctoral thesis, Florida Institute of Technology

Miller, K.L. (2000). A measure of the accuracy of self-predicted personality traits. Unpublished master's thesis, Florida Institute of Technology.

Piotrowski, C., Keller, J.W. (1989) Psychological testing in outpatient mental health facilities: A National study. *Professional Psychology' Research and Practice, 20*(6), 423-425.

Silva, M. (2011). Test re-test reliability of the Scale of Accurate Personality Prediction. Unpublished master's thesis. Florida Institute of Technology.

Standards for Educational and Psychological Testing. (2014). Washington, USA: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Sverdlova, M. (2013). Test-retest reliability of the Scale of Accurate Personality Prediction. Unpublished master's thesis. Florida Institute of Technology.

Vogt, D. S., & Colvin, C. R. (2005). Assessment of Accurate Self-Knowledge. *Journal of Personality Assessment,84*(3), 239-251. doi:10.1207/s15327752jpa8403_03

Wolf, A. T. (2006). A Re-validation of the Scale of Accurate Personality Prediction Using Self Versus Peer Comparisons. Unpublished Master's Thesis. Florida Institute of Technology.

Winter, S.N. (2002). A measure of construct validity of the Scale of Accurate Personality Prediction. Unpublished master's thesis, Florida Institute of Technology.

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 six 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 when prompted. You will use the Trial Two ID Code six weeks later. If this is your second trial, please enter the Trial Two ID Code.
3. The 16PF should take approximately 30 minutes to complete. Once you have completed the 16PF, click on the link at the end. This link will redirect you to take to 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 six-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 six 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 via email, 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.

Regards,

Jared Barrow, M.S.
Doctoral Candidate
Florida Institute of Technology
jbarrow2014@my.fit.edu

Appendix B

Primary Factors

Factor	Left Meaning	Standard Ten Scores (STEN)	Right Meaning
A: Warmth	Reserved, Impersonal, Distant	1…2…3…4…5…6…7…8…9…10	Warm, Outgoing, Attentive to Others
B: Reasoning	Concrete	1…2…3…4…5…6…7…8…9…10	Abstract
C: Emotional Stability	Reactive, Emotionally Changeable	1…2…3…4…5…6…7…8…9…10	Emotionally Stable, Adaptive, Mature
E: Dominance	Deferential, Cooperative, Avoids Conflict	1…2…3…4…5…6…7…8…9…10	Dominant, Forceful, Assertive
F: Liveliness	Serious, Restrained, Careful	1…2…3…4…5…6…7…8…9…10	Lively, Animated, Spontaneous
G: Rule-Consciousness	Expedient, Nonconforming	1…2…3…4…5…6…7…8…9…10	Rule-conscious, Dutiful
H: Social	Shy, Threat-Sensitive, Timid	1…2…3…4…5…6…7…8…9…10	Socially Bold, Venturesome, Thick-skinned
I: Sensitivity	Utilitarian,	1…2…3…4…5…6…7…8…9…10	Sensitive,

	Objective, Unsentimental		Aesthetic, Sentimental
L: Vigilance	Trusting, Unsuspecting, Accepting	1…2…3…4…5…6…7…8…9…10	Vigilant, Suspicious, Skeptical, Wary
M: Abstractedness	Grounded, Practical, Solution-Focused	1…2…3…4…5…6…7…8…9…10	Abstracted Imaginative, Idea-Oriented
N: Privateness	Forthright, Genuine, Artless	1…2…3…4…5…6…7…8…9…10	Private, Discreet, Non-disclosing
O: Apprehension	Self-assured, Unworried, Complacent	1…2…3…4…5…6…7…8…9…10	Apprehensive, Self-Doubting, Worried
Q1: Open to Change	Traditional, Attached to Familiar	1…2…3…4…5…6…7…8…9…10	Open to Change, Experimenting
Q2: Self-Reliance	Group-oriented, Affiliative	1…2…3…4…5…6…7…8…9…10	Self-Reliant, Solitary, Individualistic
Q3: Perfectionism	Tolerates Disorder, Unexacting Flexible	1…2…3…4…5…6…7…8…9…10	Perfectionistic, Organized, Controlled
Q4: Tension	Relaxed, Placid, Patient	1…2…3…4…5…6…7…8…9…10	Tense, High Energy, Impatient, Driven

Global Factors

Factor	Left Meaning	Standard Ten Scores (STEN)	Right Meaning
EX: Extraversion	Introverted, Socially Inhibited	1…2…3…4…5…6…7…8…9…10	Extraverted, Socially Participating
AX: Anxiety	Low Anxiety, Unperturbed	1…2…3…4…5…6…7…8…9…10	High Anxiety, Perturbable
TM: Tough-Mindedness	Receptive, Open-Minded Intuitive	1…2…3…4…5…6…7…8…9…10	Tough-Minded, Resolute, Unempathetic
IN: Independence	Accommodating, Agreeable, Selflessness	1…2…3…4…5…6…7…8…9…10	Independent, Persuasive, Willful
SC: Self-Control	Unrestrained, Follows Urges	1…2…3…4…5…6…7…8…9…10	Self-Controlled, Inhibits Urges

Appendix C

Table 4. SAPP Ranges, Means, and Standard Deviations

Sample	Range	Minimum	Maximum	Mean	Standard Deviation
Current Sample Trial One (N=22)	8.00	125.00	163.00	145.73	10.15
Current Sample Trial Two (N=22)	35.00	128.00	163.00	145.18	9.32
Current Sample Both Trials (N=22)	38.00	125.00	163.00	145.45	9.63
SAPP Database[a] (N=643)	69.30	101.30	170.60	147.04	13.05

[a]From the SAPP Database, which is an accumulation of data from multiple studies on the SAPP.