THE COMMENSURABILITY OF SELF-REPORTED PERSONALITY AND MOOD ASSESSMENTS

R.J. Wirth

A Thesis Submitted to the
University of North Carolina at Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Masters of Art

Department of Psychology

University of North Carolina at Wilmington

2001

Approved by

Advisory Committee

Dr. Dale Cohen   Dr. Robert Hakan

Dr. Len Lecci
Chair

Accepted by

Dr. Neil F. Hadley
Graduate Dean

TABLE OF CONTENTS

ABSTRACT

Two studies were conducted to explore the overlap between personality and mood measures. Study one focused explicitly on achieving greater congruence in the retrospective timeframes used to assess personality and mood. Instructional sets drawn from several measures were used in an attempt to quantify the retrospective timeframes individuals employ when being assessed with personality and affective measures utilizing similar instructional sets. Using the instructional sets derived in Study 1, Study 2 attempted to address the overlap between affect and personality measures through confirmatory factor analytic (CFA) techniques. Eight hundred and thirty-two participants were asked to complete one of two versions (trait or modified affect) of the NEO-FFI (a brief personality measure) or one of two versions (affect or modified trait) of the PANAS (Watson, Clark, & Tellegen, 1988). The PANAS was supplemented with 35 descriptor terms that were intended to measure the "Big Five" of personality. No CFA model tested yielded a good fit. Empirically derived models were also explored and tested with similar findings. Theoretical problems with the "Big Five" are discussed.

ACKKNOWLEDGMENTS

I would like to extend my deepest gratitude to Dr. Lecci for his guidance on this project. Over the past few years Dr. Lecci has repeatedly offered his advice, time, and friendship. I am very thankful for having the opportunity to work with such a talented researcher.

Others have also aided extensively in my education and have dedicated their time, energy, and guidance in helping me complete this project. I would like to specifically thank my committee, which consisted of Dr. Cohen, Dr. Hakan, and Dr. Lecci. Each individual has helped to provide me with knowledge, not just to complete this project, but to obtain my goals in life.

I would also like to thank the numerous people who have help in the data collection process. More specifically, Miss Darcy McMullin, Mrs. Stephanie Cohen, Dr. Caroline Clements, Dr. Patrick McKay, and Dr. James Johnson, among others, have donated their time to this project. Thank you.

Of course I would also like to thank my friends and family for their constant support. I would especially like to thank my mother, Sandra J. Rose, for her words of wisdom, love, and funding. I would also like to thank Darcy M. McMullin for her love, support and for just making me happy.

I would like to conclude by dedicating this to my little sister. Lisa has unfortunately taught me the most valuable lesson during this time: Never take things so seriously that you forget to enjoy the life you have.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Personality and mood have been linked throughout history. For instance, the Greeks, as exemplified by Galen's Humoral theory forwarded in 450 BC, explained personality types and emotional states in terms of bodily fluids. Many of the terms used during the time of the Greeks are still prevalent in modern language. For example, depression was called melancholia and was thought to be caused by excessive black bile. There are two important qualities to the theory proposed by Galen. First, Galen did not differentiate personality from mood. This provides evidence that the overlap of personality and affect has long been suspected. Second, an analysis of the terms used by the Greeks, as well as more modern terms used to describe personality and mood, has led to the current theories of personality and mood (Digman, 1990). Indeed, there has been remarkable consistency in the relevant characteristics and terms used to describe and define personality, as evidenced by the fact Humoral theory uses characterological descriptors very similar to those employed today (Pervin & Oliver, 1997).

This paper will delineate the relation between personality and mood as they are assessed by current self-report measures. To explore this topic, a number of areas need to be addressed, at the heart of which is the research exploring the structure of affect and personality. Because the factor analytic approach dominates the personality and affect literature, it is imperative to understand the use of factor analysis prior to reviewing the relevant literature.

Factor Analysis

At the heart of every theory is a construct. In the current paper it is the constructs of personality and mood. However, problems arise when the constructs that are hypothesized to gird the theories of personality and mood do not accurately reflect the theories. Factor analysis can address this issue.

With any assessment scale there are a number of items used to assess the construct(s) of interest. The items should broadly represent the domain of interest (e.g., Messick, 1992). Factor analysis allows for the testing of the construct validity of a scale by offering an empirical evaluation of the relation between the items. If a scale that was designed to measure five constructs of personality (i.e., extroversion, neuroticism, openness, agreeableness, conscientiousness) mathematically aligns into three categories, then it would suggest that the scale was not assessing what it purports to assess. That is, the scale and theory are incongruous.

The categories that emerge from a factor analysis are called factors. Furthermore, these factors can be orthogonal (independent) or oblique. The relation among the factors can be determined by the type of factor analysis performed (e.g., those using orthogonal vs. oblique rotations) or by the functional relatedness of the items entered into the analysis. Therefore, in order to evaluate a construct based on the analysis of items it is important that items be used which do not bias the outcome of the factor analysis. This would represent an empirical approach to evaluating construct validity. Alternatively, item selection and construct validity can be based on theoretical models.

Overview of Personality Theory

Throughout psychology's relatively short history a number of different theoretical perspectives have attempted to define personality (e.g., Allport & Allport, 1921; Erickson, 1950; Eysenck & Fulker, 1983; Freud, 1933; Gray, 1970; Hull 1943; Jung, 1939; Kelly, 1955; Masters, 1967; Skinner, 1948; Watson & Rayner, 1920). One approach to defining this construct is to focus on the elements that are common to all theoretical perspectives.

All personality theories are in agreement that personality is not defined by any transitory qualities. For example, Allport and Odbert (1936) defined personality as "generalized and personalized determining tendencies – consistent and stable modes of an individual's adjustment to his environment" (1936, p.26). In other words, an individual's personality consists of stable characteristics that should be approximately the same at age nine as they are at age seventy-nine. In contrast, states and activities are those aspects of a person that are temporary, brief, and caused by external circumstances (see also Cholin, John, & Goldberg, 1988).

A second common feature of every theoretical perspective is that the study of behavior is ultimately understood and described through language. Likewise, the understanding and description of an individual's personality (i.e., stable behavioral characteristics) is ultimately accomplished through the use of language. Thus, it is this analysis of language, as descriptors of behavior, which has lead to a better understanding and to the current method of personality description (Digman, 1990). More specifically, through its association with descriptor terms (i.e., descriptive language), personality was defined and assessed in terms of traits (i.e., categories of stable behaviors; Klages, 1926; Digman, 1990).

Trait Theories

Cattell (1943) developed one of the first and most comprehensive trait theories. Using factor analytic techniques, Cattell advocated a very complex personality model consisting of 16 factors. Although an actual description of Cattell's factors is beyond the scope of this project, it is important to note two things. First, the method by which Cattell developed his theory was through the use of factor analytic techniques. Second, many of the theories developed after Cattell's (1943) are based on a further factor analyses of Cattell's 16 factors (e.g., Fiske, 1949; Tupes & Christal, 1961; see Digman, 1990 for further discussion). In fact, many of the current trait theories are direct empirical descendants of Cattell's research (Digman, 1990).

Although the popularity of trait theory temporarily declined as a result of Mischel's (1968) strong criticism of traits, ultimately this critique sparked a resurgence of the trait perspective. In his book entitled Personality and Assessment, Walter Mischel (1968) argued that traits where useful summaries of people's behaviors, but that they lacked the discriminative ability to account for the differences in behavior across time and that they ignored the circumstantial influences on behavior. One reason, Mischel suggested, for the apparent inability of trait theories to explain the inconsistencies in behavior was their atheoretical development. In response to these and other criticisms, trait theorists began to accumulate research in order to better validate their approach to assessment as well as develop new trait-based theories of personality.

In an attempt to link trait descriptors with a theoretical model, Eysenck (1990) developed a hierarchical organization of personality. At the most basic level of this hierarchy is the specific response or behavior (e.g., introducing yourself to a new coworker).[1] If similar responses are consistently elicited in different contexts (e.g., introducing yourself to strangers at social gatherings, at the market, etc.) then they are considered to be a habitual response. A trait consists of a number of related habitual responses. For example, the trait of sociability may consist of the habitual responses mentioned above along with related habits such as initiating hand shaking when you meet someone new, volunteering your input at work and in social settings, or being the "life of the party." Thus, the trait represents a broader construct than the habitual single response. Furthermore, Eysenck (1990) proposed that personalities can be described in terms of superfactors, and these superfactors are comprised of several related traits. A person considered to be an extrovert (superfactor), for instance, may exhibit such traits as sociability, excitability, and activity. It is the stability of these traits and factors that is at the heart of personality theory. Thus, the term "trait" and the more superordinate term "factor" are central to the study of personality.

Cattell's (1965) 16 factor and Eysenck's (1970) three factor theories are two of the trait taxonomies most commonly found in psychology textbooks (Digman, 1990). However, as early as 1946, Fiske developed a five factor theory similar to current theories of personality (Digman, 1990). Since that time, "Big Five" theories of

---

[1] Eysenck's specific definition of personality and traits is not central to this research. However, because many of the definitions found in a variety of theories are similar, it provides a basic overview of a prototypical trait theory.

personality have become prominent in the personality literature (e.g., Borgatta, 1964; Costa & McCrae, 1985; Digman, 1990; Goldberg, 1981; John, 1989; Norman, 1963; Smith, 1967; Tupes & Christal, 1961). Researchers suggest that the "Big Five" (i.e., neuroticism, extroversion, openness, agreeableness, and contentiousness) or factors similar in name, are reliable descriptors for personality. This is also one reason why traits are so prevalent in the empirical literature, as they are theoretically easy to assess and even easier to reputably analyze.

The "Big Two" as a More Parsimonious Model

A close examination of the taxonomy of trait terms finds that the degree of "consensus" is, however, questionable. Rather than attempting to identify the largest number of factors where there is some degree of agreement, we can instead select common points that are evident in all models. Consensus can be found between a majority of personality theories on at least two factors. That is, there are two factors that are found within the "Big Five" theory of personality as well as in all other models (e.g., Cattell, 1943; Eysenck, 1970).

This minimum level of consensus can be defined by the "Big Two" factors of extroversion and neuroticism. As can be seen in Table 1, a multitude of researchers, those who have adopted a five factor model and those who have not, using very similar terms, have all agreed on the presence of these two factors. Moreover, the "Big Two" have been reported to consistently emerge as the two primary factors accounting for the greatest variance in the factor analysis of personality traits (Costa & McCrae, 1985). Table 1 provides a summary of the prominent models in personality psychology and their

Table 1

The consensus on the "Big Two"

Theorist	Extroversion	Neuroticism
Fiske (1949)	social adaptability	emotional control
Cattell (1957)	exvia	anxiety
Tupes & Chistal (1961)	surgency	emotionality
Norman (1963)	surgency	emotional
Borgatta (1964)	assertiveness	emotionality
Eysenck (1970)	extraversion	neuroticism
Guilford (1975)	social activity	emotional stability
Costa & McCrae (1985)	extraversion	neuroticism
Lorr (1986)	interpersonal involvement	emotional stability
Hogan (1986)	sociability & ambition	adjustment
Digman (1990)	extraversion	neuroticism

Table 1 is a partial replication of a table depicting the overlap between five factor theories as presented by Digman (1990).

fit with the "Big Two." This two-factor model demonstrates empirical consistency and, in the sections to come, I will delineate its theoretical consistency within both personality and the related field of mood/affect.

Interestingly, neuroticism pertains to a person's emotional stability (high neuroticism is defined by being emotionally unstable) and it has been found that individuals who score high on neuroticism also score high on negative affect scales (Gross, 1998; Wilson, 1999). It has also been found that individuals who score high on extroversion (e.g., high extroversion is defined by outgoing, active, sociable, etc.) also tend to score high on positive affect scales (Gross, 1998; Wilson, 1999). This suggests that the "Big Two" of personality may be the primary factors that overlap with the structure of mood. Obviously, affect alone may not sufficiently explain personality in all it's complexity. However, for the purpose of this study, the emotional aspects of personality will be the primary focus.

Affect

The terms affect, emotion, and mood are used in a multitude of contexts to describe individuals and their experiences. Examples can be found throughout English literature (e.g., "Mr. Doppler played on the vast organ of human emotions like a master musician, twittering on the Acquisitiveness stop as one possessed of an evil genius."; Shepherd, 1964, p.257) the scientific literature, with some journals specifically dedicated to the topic (e.g., Cognition and Emotion, Motivation and Emotion), and most commonly in everyday life (e.g., "I am never in a better mood than when I'm doing research."). Although the terms affect, emotion, and mood may in fact address slightly different aspects of experience, generally in the psychological literature they are used interchangeably.

Clearly, affect experiences consist of several elements. For example, they are comprised of physiological changes such as increased heart rate or blood pressure (Cannon, 1929). There has also been extensive research in the area of facial expressions and mood states (e.g., Eckman & Friesen, 1978; National Advisory Mental Health Council, 1995; Carroll & Russell, 1996). Facial movements such as pulling the lip corners upward to denote a smile have long been associated with a positive affective state (e.g., happiness), whereas facial movements such as pulling the lip corners down to denote a frown have long been associated with a negative affective state (e.g., sadness). Furthermore, much of Eckman and Friesen's (1978) research has shown that these facial cues exist cross-culturally and such universal manifestations would likewise be consistent with basic physiological processes. Cognitive components have also been associated with emotional experiences (Martin, 1990). For instance, The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychological Association, 1994) cites the cognitive construct of hopelessness as a central criterion for Disthymic Disorder. These examples exemplify the complexity with which affect is experienced, expressed, and defined.

The Structure of Affect

Schlosberg (1954), after struggling to integrate many theories of affect, proposed a three dimensional factor structure, which combined an activation continuum with two affect continuums. More important is that Schlosberg's (1954) research suggested the bipolarity of the affect and activation continuums such that high activation is not necessarily associated with either positive or negative affect but can be experienced with both positive and negative affect. Figure 1 provides a depiction of this model. Another important aspect of the dimension of affect suggested by Schosberg (1954) is that positive and negative affect are not intrinsically separate (i.e., orthogonal). This model however has not been widely supported.

More recently, the debate over the structure of affect has centered on the relation between positive and negative affect. Since the time of Schosberg (1954) many researchers have suggested that positive and negative affect are in fact orthogonal (e.g., Warr, Barter, & Brownbridge, 1983; Watson, Clark, & Tellegen, 1988). That is, high levels of one type of affect (e.g., positive affect) does not necessarily require low levels of the reciprocally labeled affect (e.g., negative affect). For example, Watson and Tellegen (1985) provided a two-dimensional structure of affect (i.e., positive and negative, refer to Figure 2) in which positive and negative affect make-up the two affect continuums with each factor ranging from low to high activation. This model was tested using data from several self-report mood studies (Watson & Tellegen; 1985). As reported by Watson and Tellegen, the first two factors with orthogonal rotations, to consistently emerge when analyzed using exploratory factor analysis are positive and negative affect. Although Watson and Tellegen's pictorial model has changed over the years, a cornerstone of their model remains the orthogonality of affect.

Alternative models have also been developed. Much like Watson and Tellegen's (1985) model, Russell's (1980) two-dimensional model was also developed using primarily self-report methods. More specifically, Russell asked individuals to sort 28

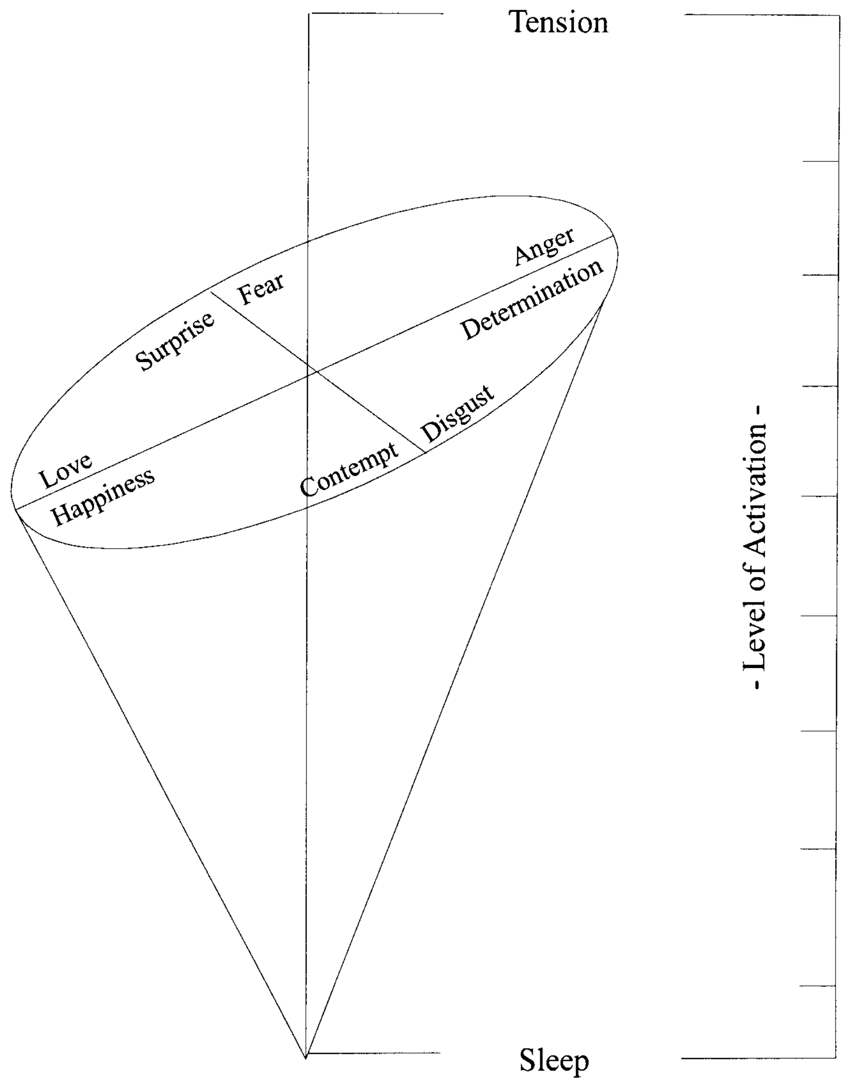

Figure 1. Replication of Schlosberg's (1954) three dimensional structure of affect.

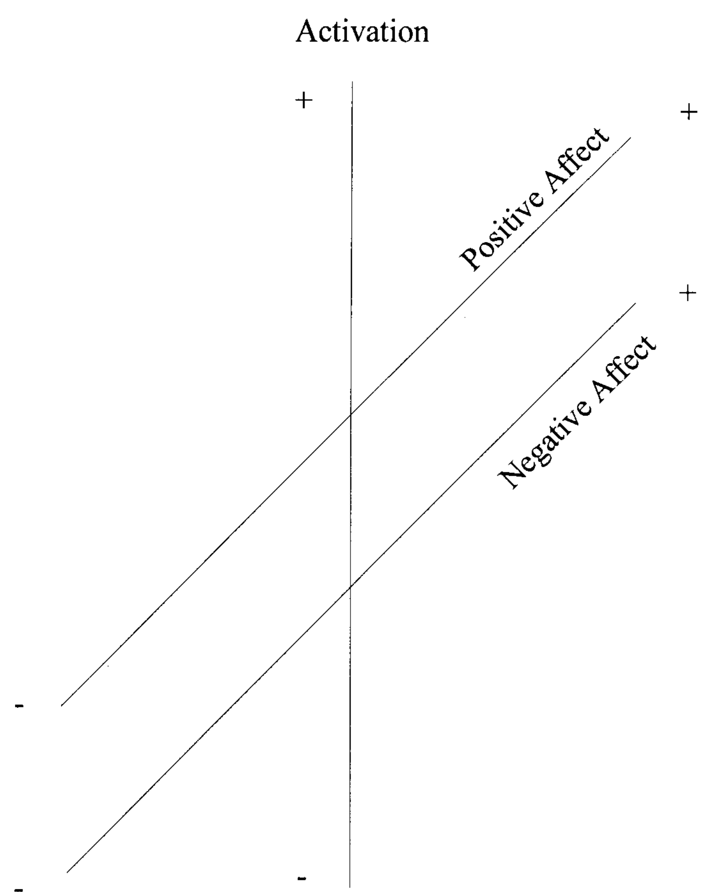

Figure 2. Replication of Watson & Tellegen's (1999) orthogonal model of affect.

words described as "words or phrases that people use to describe their moods, feelings, temporary states, affect, or emotions," into eight categories (p. 1164). The categories consisted of aroused, contented, depressed, distressed, excited, miserable, pleased, and sleepy. Subjects were then asked to order the eight categories in a circle by the way each category relates to one another. The more similar the terms the closer the terms should be placed on the circle and opposite terms should fall directly opposed to one another. Using this technique, Russell's circumplex model was developed which had one factor consisting of positive and negative affect (i.e., they are bipolar; refer to Figure 3) with a second factor that consists solely of arousal (activation).

Russell's (1980) theoretical model (i.e., the order of the eight categories) is shown in Figure 3. With the data obtained from this study the model was tested using the Guttman-Lingoes non-metric multidimensional scaling procedure SSA-1, producing a stress value of .001. That is, the all of subjects seemed to produce very similar models (Russell, 1980).

Regardless of the model of affect one adopts, there are several conclusions that are true in all cases. First, arousal is central to studying the structure of affect. Second, the language within the affect literature needs to be more consistent (i.e., a standardized language of affect; Russell & Carroll, 1999a; Watson & Tellegen, 1999). Indeed, Watson and Tellegen (1999) recently stated that the majority of the disagreement over the orthogonality of affect is due primarily to using the same language to describe two different results.[2] Therefore, any future research on the structure of affect or mood must

[2] Watson and Tellegen's (1999) argument and suggestions have not been fully articulated in the scientific literature and therefore, for the purpose of this study the current language will be used.

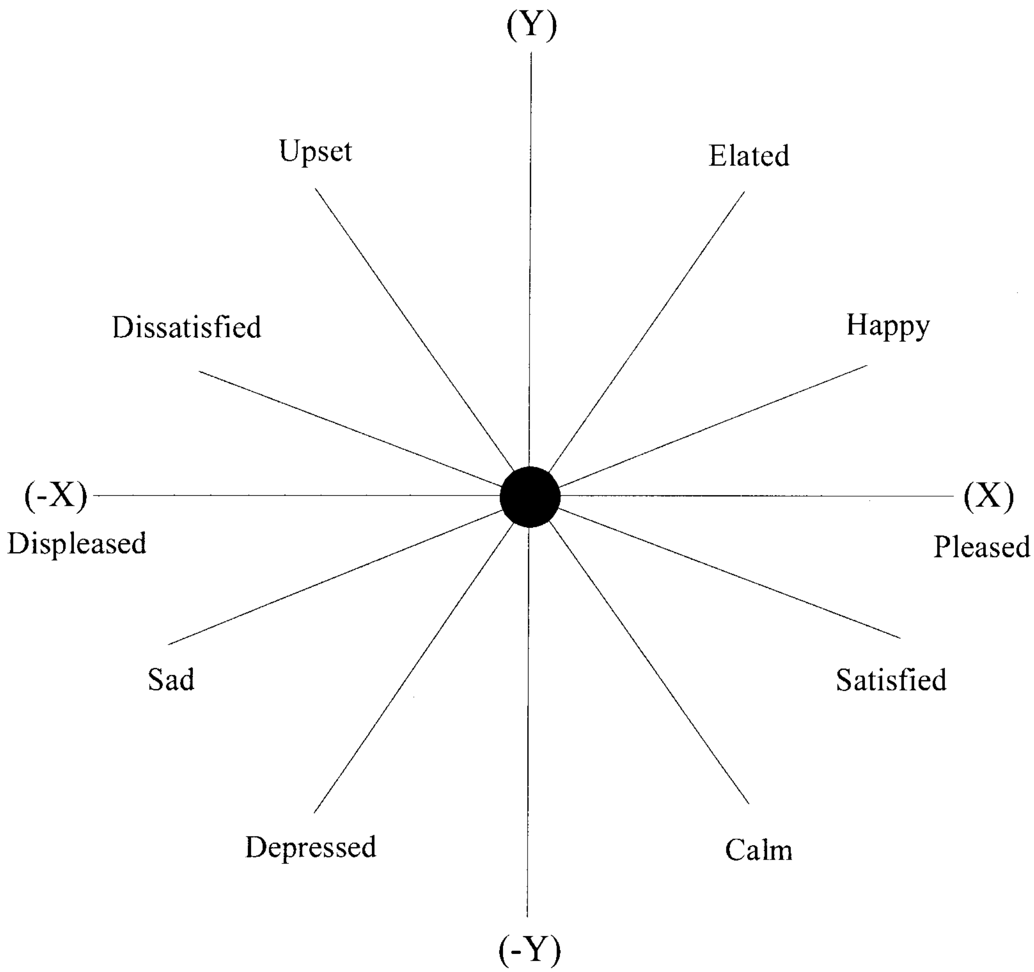

Figure 3. Replication of Russell & Carroll's (1999b) circumplex model of affect: X = Pleasure, Y = Activation.

attempt to control for the biasing influences of the descriptor terms used. For example, the use of primarily all high activation terms in a factor analysis would not allow for the emergence of an activation factor, if in fact one exists. A final point of convergence for the theories of affect is that individuals actively self-regulate their affective state (Carver & Scheier; 1990). This last point will serve as the mechanism for making a theoretical link between mood and personality.

Affective Self-regulation

The idea that self-regulating systems exist within the human body is nothing new. For example, the human autonomic nervous system has been shown to regulate itself, whereby activation of the sympathetic nervous system (increase in heart rate, blood pressure, and breathing) is followed by the activation of the parasympathetic nervous system (e.g., heart rate dropping below baseline; see Selye, 1974). This example, drawn from one of the most basic human responses, illustrates the presence of an automated, self-regulatory system.

Later researchers have shown that people actively work to maintain a baseline mood state (Carver & Scheier, 1990). In this model, the individual's "normal" mood is their baseline state (which need not be at the center of the affect continuum). As an individual deviates from this mood he/she has a tendency to regulate it, or correct it back to his/her normal emotional state (the process itself is known as affect self-regulation). For example, if a person becomes sad he/she will regulate him/herself back to a more positive affective state until he/she has reached his/her normal baseline mood. This process can occur at both an automated and volitional level.

More specifically, Carver and Scheier (1990) suggest that a person's mood is regulated by a series of feedback loops. Actions or behaviors are constantly being monitored. This monitoring is composed of comparisons between an individual's current action or behavior and the reference or goal behavior. When expectancies are in the desired direction (i.e., the individual is moving towards a goal), then the result is the experience of positive affect. Following the same line of reasoning, if expectancies are not being met (i.e., the individual is not moving towards the goal or is moving away from the goal), then the result is the experience of negative affect. Therefore, there is a constant attempt to minimize discrepancies. Although a positive discrepancy leads to the experience of positive affect, there is still a discrepancy and the individual will attempt to minimize all discrepancies (Carver & Scheier; 1990).

In addition, there is a second feedback loop that is responsible for monitoring the action feedback loop mentioned above (Carver & Scheier; 1990). This meta-monitoring loop surveys the efficiency of the action loop by monitoring the rate of change in the discrepancies found in the action loop. The size of the discrepancy is not important per se, but it is the perceived rate of progress towards the goal that is of interest. For instance, if the rate of progress towards the goal is too slow (producing the experience of negative affect), then an individual will increase the rate at which it is being approached (i.e., regulate) until the expected rate is maintained (Carver & Scheier; 1990).

It is important to note that Carver and Scheier (1990) suggest that the affect experience associated with a discrepancy between an individual's goal and the velocity at which an individual is moving towards that goal is a function of the individual's perception of the discrepancy. Assuming that failure to effectively pursue a goal can temporarily increase negative affect, the chronic failure to effectively pursue goals can therefore result in more stable experiences of high negative affect (i.e., neuroticism). Similarly, a consistent inability to increase the rate at which a goal is being approached (to reduce the experience of negative affect) could likewise result in stable experience of negative affect or neuroticism. As a final safeguard, Carver and Scheier (1990) propose the existence of a disengagement function. Specifically, if an individual is able to disengage from a goal that is unsuccessfully being pursued, they can reduce the experience of negative affect. This disengagement is a healthy adaptive function (Carver & Scheier, 1990; Lecci, Okun, & Karoly, 1994), and the failure to disengage from the action or goal may also result in the chronic experience of such psychological disorders as depression (e.g., Klinger, 1987; Kuhl & Helle, 1986).

In sum, a person's mood or affective state is comprised of a degree of positive and/or negative affect and some level of arousal. The most frequent occurring affective and arousal states will be defined as baseline states. When deviation from either of these baselines occurs, a self-regulating system returns the mood or affective state to its baseline level. This baseline is a person's "normal" mood. Furthermore, a person's personality is comprised of traits that are a person's normal or primary state. Intuitively, one might conclude that frequently occurring states (i.e., habitual states) are the building blocks for traits (see Eysenck, 1970). Thus, from a research standpoint, it becomes important to define the point at which a person's "normal" or "baseline" mood state becomes a trait. In other words, temporal stability is assumed for the constellations of personality while affective states are considered more transient (Cholin et al., 1988). Nevertheless, no clear temporal boundaries have been articulated in the literature. One way to begin addressing this issue is to examine the methods of assessment in the areas of personality and mood, and attempt to derive temporal specifications in this way.

The Relation Between Theory and the Assessment Tool

When developing an assessment tool, the researcher has a theory or model of the construct that he/she is attempting to assess. For example, the PANAS was developed after Watson and Tellegen (1985) developed an orthogonal two-factor theory of affect. Therefore, Watson et al. (1988) developed the PANAS to measure orthogonal factors of positive and negative affect. In fact, the terms used to develop the PANAS are considered by Watson et al. (1988) to be only high activation positive and negative terms. This ensures orthogonal results when using factor analytic techniques to define the factor structure of the PANAS. If a larger number of terms were introduced with varying activation levels, the factor structure that emerged with factor analytic procedures should in fact change. For example, the addition of lower activation terms may produce a factor structure more congruent with Russell's (1980) model. Furthermore, the addition of terms not commonly associated with positive or negative affect may in fact produce a more complex factor structure that may be more congruent with the "Big Five" theories of personality.

Therefore, an examination of personality and mood assessment tools as a method to derive temporal specifications or structural differences between mood and personality must be done with the understanding that the underlying theory upon which the assessments were based will influence the conclusions reached.

The Measurement of Affect and Personality

Differences between affect and personality are more complicated than mere differences in language. In fact, many assessment tools use similar, if not identical language to describe affect and personality. For instance, two versions of the Multiple Affect Adjective Check List - Revised (MAACL-R; Zuckerman & Lubin, 1965) have been widely used and both versions consist of the same list of 132 adjectives. Thus, when assessing affect or personality using a measure such as the MAACL-R or other measures such as the Positive Affect Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), the outcome is defined using the same descriptive terms. Following this line of reasoning, affect and personality are necessarily defined using the same terms or language. This similarity in language, in and of itself, assumes (and ensures) a certain degree of overlap between the structure of personality and mood. Therefore, the true difference between current (on-line) affect and personality must be defined outside the area of the descriptive terms employed, at least with regard to measures such as the MAACL-R and PANAS.

Differences in assessment (current affect vs. personality) may in fact be a function of the instructions used. For example, when assessing current levels of affect with the MAACL-R, individuals are instructed to check any adjectives that describe how they feel right now, while assessing personality with the MAACL-R individuals are instructed to check any adjectives that describe how they generally feel. Therefore, the MAACL-R distinguishes current affective levels from personality with the use of instructional sets that incorporate words such as "generally" (personality) and "right now" (mood).

As can be seen in Table 2, the MAACL-R is not the only assessment tool using temporal language in the instructions to make the distinction between current levels of affect and personality. In fact, some measures have a variety of instructional sets. The PANAS (Watson, Clark, & Tellegen; 1988) offers seven different sets of instructions to be used with its 60 item or its 20 item assessment.[3] The instructions range from asking individuals how they feel at the moment, today, the past few days, week, past few weeks, year, and in general. Although the PANAS offers a series of instructions with a variety of temporal distinctions, it still does not offer a clear distinction between affective experiences and personality. Thus, where the scale begins to measure personality and stops measuring affect is unclear.

One may consider addressing this issue through a systematic examination of the test-retest reliability of different instructional sets. Personality is considered to be stable across time, while an individuals' on-line affective experience is expected to show greater variability. Therefore, the test-retest reliability for a personality measure should be (significantly) higher than the test-retest reliability of an on-line measure of affect. Due to the sheer number of instructional sets available with the PANAS, it seems to be an appropriate scale for such a discussion. As can be seen in Table 3, the test-retest reliabilities for all instructional sets, with the exception of the "Moment" instruction, increase as the referenced time increases (as reported by Watson et al., 1988). Unfortunately, this does not offer a possible splitting point (i.e., a temporal definition)

---

[3] The Original PANAS consists of 60 items. However, a 20 item short form was developed that is capable of assessing levels of positive and negative affect for all of the instructional sets used with the 60 item version.

Table 2

Instructional differences between affect and personality assessment

Assessment Tool	Instructions	
	Affect	Personality
Affect Grid (Russell, Weiss, & Mendelsohn, 1980)	"Right Now"	- none-
Beck's Depression Inventory (Beck & Streer, 1987)	"Now"	- none –
NEO – PI – R (Costa & McCrae, 1985)	- none -	-no specific instructions-
MAACL-R (Zuckerman & Lubin, 1965)	"Right Now"	"In General"
PANAS (Watson, Clark, & Tellegen, 1988)	"Moment"	"In General"

between affect and personality. In fact, the incremental increase in test-retest reliability as a function of time suggests that the two are more intertwined, at least within the context of assessment, than has been suggested previously. As can be seen in Table 3, the assessment period of past few weeks does not differ significantly from the general assessment period. Moreover, the moment (presumably on-line mood) assessment does not differ significantly from the general (presumably personality) assessment period.

Some personality assessment tools avoid this issue by using more complex statements. Costa and McCrae (1992) have developed one of the most widely used questionnaires measuring personality factors (NEO-Personality Inventory-Revised; NEO-PI-R). The NEO-PI-R offers two clear advantages over other measures. First, the NEO-PI-R is consistent with many current trait theories (i.e., it measures five orthogonal factors). Furthermore, the factors measured by the NEO-PI-R are similar to many of the factors found in other non-five-factor models. For example, the extraversion and neuroticism factors of the NEO-PI-R map onto Eysneck's (1970) extroversion and emotional instability superfactors. Second, the NEO-PI-R offers a contrast to the assessment tools using descriptor adjectives (e.g., PANAS), as it employs more complex descriptor statements. Although it was originally designed to measure only three factors (extroversion, neuroticism, and openness), two factors were later added conforming to the popular "Big Five" model (agreeableness and conscientiousness) put forth by other researchers of the time (e.g., Goldberg, 1992).

The NEO-PI-R attempts to assess personality by having individuals rate their agreement with 240 items or statements using a 0 (strongly agree) to 4 (strongly disagree)

Table 3

Test-retest reliability of the different instructional sets for the PANAS.

Time Instructions	Average Reliability	Retest Interval	Significance of Difference
Moment	.50	8 weeks	AB
Today	.43	8 weeks	A
Past few days	.45	8 weeks	A
Past week	.47	8 weeks	A
Past few weeks	.53	8 weeks	AB
Year	.62	8 weeks	AB
General	.70	8 weeks	B

Note: N=101. Coefficients not sharing the same letter under the "Significance of Difference" heading are significantly different from one another at the $p<.05$ level (two-tailed, Bonferroni correction for multiple comparisons).

scale. This procedure eliminates the need for temporal instructions because the statements themselves imply a retrospective account of behavior and feelings using such terms as "often", "sometimes", "seldom", and "rarely." Indeed, a term like "seldom" as in "I'm seldom sad or depressed" implies behavioral consistency through the consistent absence of the target behavior. For example, strong disagreement with the item implies frequent bouts of sadness or depression. As can be seen in Table 4, many of the statements from the NEO-PI-R contain a high degree of overlap with other items found on single adjective assessments such as the PANAS.

Throughout this paper the extensive overlap between affect and personality has been documented. Interestingly, although most researchers now agree that, at least in some respect, affect is bipolar, most would agree that personality (and especially, the "Big Two") is orthogonal. Assuming that affect and personality are intimately related, it seems inconsistent that the structure of one is bipolar (i.e., affect) while the other is orthogonal (i.e., personality). This, in fact, may be an artifact of the assessment tools themselves. More specifically, it has been shown that the clearest difference in assessing affect and personality are the instructions used. Affect is most often assessed on-line and encompasses a brief retrospective period of time (e.g., right now) while personality is assessed using longer retrospective accounts (e.g., in general). This retrospective assessment may ultimately influence the structure of affect and personality.

Retrospective Assessments

The temporal nature of the assessment seems the most important issue for distinguishing the difference between personality and affect and therefore the most

Table 4

An illustration of the convergence of NEO-PI-R and PANAS terminology

NEO-PI-R Statement	PANAS Descriptor
"I am seldom sad or depressed" (reverse coded)	"Sad"
"Sometimes I bubble with happiness"	"Happy"
"I often crave excitement"	"Excited"
"I'd rather not talk about myself and my achievement"	"Bashful"
"Once I start a project, I almost always finish it"	"Determined"

appropriate place to attempt an unraveling of this conundrum. As was mentioned above, most of the data used to formulate these modern models of affect and personality have been self-report. Furthermore, personality is most often assessed using long retrospective accounts. The instructions in Table 3 provide examples of how an individual is asked to recall or summarize their affective experience over time. This retrospective assessment may allow for a more complex view when labeling an affective experience.

Alternatively, some researchers have questioned an individual's accuracy recalling emotional information (i.e., intensity and frequency). In a study conducted by Thomas and Diener (1990) individuals' on-line emotional state was assessed four times a day for a period of 3 weeks and then at the end of the day for an additional 6 weeks. Retrospective assessments were then taken 1 week after the conclusion of the daily assessments. Mean on-line scores (daily levels of affect) for level of positive and negative affect as well as intensity were then statistically compared to the retrospective mean scores (recalled mean levels of affect). Interestingly, participants significantly overestimated the intensity of both the positive and negative affect experienced. Furthermore, participants significantly underestimated the frequency of positive affect experienced (Thomas & Diener, 1990). These results suggest inadequacies in the recall of affect experiences.

In a related vein, it is clear that current (on-line) mood influences self-report measures, and this would be especially true if the construct being assessed was closely related to mood (e.g., personality). That is, if currently sad, one's rating of neuroticism (the predisposition to interpret events negatively) may be biased high. Likewise, personality influences how individuals subjectively interpret on-line experiences of mood. That is, someone scoring high on neuroticism is more likely to over-report negative mood experience (Gross, 1998). Both of these scenarios reflect construct overlap as a consequence of methodological problems with self-report inventories. Although these are important points, current assessment techniques offer few viable options to circumvent this confound.

Current Research on the Factor Structure of Mood Measures

In a study by Deiner and Emmons (1984) evidence was found to suggest that the orthogonality of positive and negative affect was in part dependent upon the time frames used for assessment.[4] More specifically, as the time frame increased (e.g., from "moment" to "general") positive and negative affect appeared more orthogonal.[5] Therefore, longer assessment time frames appear to more closely resemble the structure of personality. This is noteworthy because the first two factors to consistently emerge in the personality literature are extroversion (positive affect) and neuroticism (negative affect) and they are considered orthogonal (Costa & McCrae, 1985).

The findings of Deiner and Emmons (1984) conflict with those of a later study conducted by Watson (1988) in which he examined the psychometric properties of the PANAS by evaluating the orthogonality of positive and negative affect for each

[4] The orthogonality of any two factors or items is in part dependent upon the items used in the factor analysis as well as the rating scale used for the assessment of each item. For an in depth review of the importance and influence of rating scales on orthogonality see Russell and Carroll (1999) and for a review of factor analytic techniques refer to Nunnally and Bernstein (1994).

[5] Although throughout this paper the PANAS is often said to employ "moment" instructions, the PANAS actually incorporates both "moment" and "right now" into the instructional set.

instructional set (i.e., ranging from the "Moment" to "General"). A total of 3248 subjects were asked to rate 60 or 65 affect descriptors, drawn from three mood measures, on a five point scale ranging from very slightly or not at all to extremely (i.e., the same rating system as the PANAS) to identify how much they have experienced each mood state during the instructional time frame. Although a majority of the participants were asked to complete only one of the instructional sets, some did complete multiple temporal instruction sets. Using primarily between subject exploratory factor analysis, Watson (1988) found that the different instructional sets of the PANAS had little affect on the apparent orthogonality of positive and negative affect. However, it is important to remember that the PANAS was created with terminology and ratings that bias responses towards orthogonality. For example, the lowest rating (i.e., 1 = slightly or not at all) used by Watson (1988) does not explicitly allow for the non-experience of that mood. The addition of a lower rating allowing for the absence of a feeling (e.g., 0 = not at all) creates an ambiguous response format allowing for orthogonal and oblique factor structures. However, with the PANAS lacking this type of ambiguous response format it is not surprising that the factors remain orthogonal regardless of the timeframe used. Although the different scales diverged with regard to the orthogonality of positive and negative affect, Watson reported a high degree of converging correlations between the different scales of positive and negative affect, suggesting the measurement of a single underlying construct.

The previous points are important for several reasons. First, the primary purpose of the current study is to explore the commensurability of self-reported personality and mood assessments. That is, the structural overlap of affect and personality as measured by the most prominent assessment tools in the literature. The tools are largely self-report in nature and many self-report measurements theoretically assess both affect and personality. Second, although there is disagreement over the orthogonality of positive and negative affect between the different scales, there is agreement on the existence of both positive and negative affect. This allows for greater confidence in the measurement of positive and negative affect regardless of which scale is used. Indeed, virtually any scale could be used for the measurement of affect while studying the structural relations between affect and personality. However, it is possible that the scale could establish a floor on the simplicity of the emergent factor structure. For instance, if using the PANAS (orthogonal factors of affect), one would expect a minimum of two factors to emerge (i.e., positive and negative affect). On the other hand, if using a scale such as the one developed by Larsen and Diener (1985) one would expect to see the two factors of hedonic level as well as an intensity factor (i.e., a minimum of three factors). Importantly, regardless as to which mood measure is selected, a two-factor structure should be the minimum factor structure to emerge.

The number of factors to emerge is also dependent upon the type of factor analytic procedure used. For instance, Hunsley (1990) found that the MAACL-R's factor structure varied from a two factor (positive and negative affect) to a five factor structure (sensation seeking, hostility, depression, positive affect, anxiety) depending on the specifications regarding orthogonality. Thus, it is important to select factor analytic techniques that allow for both orthogonal and oblique (non-orthogonal) rotations.

Meyer and Shack (1989) attempted to further investigate the relation between affect and personality. Specifically, a total of 231 participants were asked to complete a state mood measure (i.e., a list of descriptor terms with the instructions to rate on a 4-point Likert scale how closely each term described him/her “in the past day”), a trait mood measure (i.e., a list of descriptor terms with the instructions to rate on a 4-point Likert scale how closely each term described how him/her “generally feel”), and a personality measure (i.e., Eysenck Personality Questionnaire subscales of neuroticism and extroversion; Eysenck & Eysenck, 1975). Sixty-nine subjects were then retested twice over a five-week period using the mood measure and again two weeks later on the trait and personality measures.

The results showed structural agreement between current mood, trait, and the personality facets of neuroticism and extroversion. More specifically, two-dimensional, orthogonal structures consistently emerged with similar item loadings. This provides further evidence of a structural relation between some of the basic constructs in affect and personality. However, Meyer and Shack (1989) did point out that because the assessments were administered on the same day, the similar loadings may in fact be an artifact of “conceptual carryover effects” (p. 698). Therefore, in future studies it would be important to minimize the overlap in the timeframes being assessed. More specifically, assessments using moment instructions on one occasion may be compared to assessments using in general instructions on another occasion in order to better assess the stability over time (Meyer & Shack, 1989). It is important to note that by using a number of assessments and numerous timeframes, the ability to adequately study the temporal component of the structural convergence of affect and personality increases.

Current Research on the Validity of the NEO-PI-R and NEO-FFI Factor Structure

Costa and McCrae have published numerous studies describing the creation and validation of the NEO-PI (e.g., Costa & McCrae, 1985), NEO-FFI (e.g., Costa & McCrae, 1989), and NEO-PI-R (e.g., McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). The primary methodology employed for both the development and validation of the NEO measures has been exploratory factor analysis (McCrae et al., 1996). Although the validity of a measure is in part dependent upon the methodology used during the creation of the measure, for the purpose of the following studies, the validity and replicatability of the NEO measures are of primary interest.

As was previously stated, the primary methodology employed by Costa and McCrae was exploratory factor analysis. In an attempt to remain consistent with the theoretical orthogonality of personality, strictly orthogonal rotations (i.e., varimax, validimax, and procrustes) were applied throughout the development and validation of the NEO measures (McCrae et al., 1996). Exploratory factor analysis of the NEO-PI with varimax rotations have consistently produced five-factor solutions similar to those propose by Costa and McCrae to be employed when scoring the NEO-PI (McCrae & Costa, 1989).

In an attempt to further validate the structure of the NEO-PI using exploratory factor analysis, McCrae and Costa (1989) proposed and tested a new weighted orthogonal rotation technique. The validimax procedure was derived from Schonemann's (1966) orthogonal Procrustes procedure and was intended to maximize the external validity of the measure (McCrae & Costa, 1989). This is done by weighing the factor loadings based on external criteria or correlates. That is, the rotations are not dependent upon the factor loadings themselves, but on the weights assigned to particular loadings and the weights assigned to their respective alternative loadings (McCrae & Costa, 1989). Validimax rotations have been shown to produce similar factor structures as the more excepted varimax rotations (McCrae & Costa, 1989). Thus, suggesting high construct validity.

Although orthogonal rotations were used to stay consistent with the five theoretical factors of personality, it is important to remember that these procedures force orthogonality. That is, varimax, validimax, and procrustes rotations do not allow for oblique factors to emerge if they in fact exist. This confound, associated with the use of orthogonal rotations, can effect the ability to recover the factor structure using confirmatory factor analysis.

The potential problems that this confound could produce were highlighted in a study conducted by McCrae et al. (1996) exploring the different findings produced by confirmatory factor analysis and procrustes rotations. In this study McCrae and associates, using the procrustes rotation with the NEO-PI-R, produced a similar factor structure to that proposed during the development and earlier validation of the NEO-PI and NEO-PI-R. However, less success was found when exploring the factor structure of the NEO-PI-R using confirmatory factor analysis.

An advantage of confirmatory factor analysis is that it allows for an empirical evaluation of a model. This evaluation is most often conducted using a series of fit-indices whereby a fit index above .90 is considered a satisfactory fit and a fit index above .95 is considered good fit. However, when McCrae and associates (1996) tested the fit of the NEO-PI-R's simple structure using orthogonal confirmatory analysis, they reported fit-indices of less than .50. The highest fit index reported for this model was an adjusted goodness of fit index of .57. Clearly, the fit of the NEO-PI-R model relative to their data was poor. McCrae and associates were not the only individuals to report such findings. In fact, a number of other researchers have reported similar findings with regard to the NEO-PI-R (e.g., Borkenau & Ostendorf, 1990; Church & Burke, 1994; Panter, Tanaka & Hoyle, 1994) and related findings with the NEO-FFI (e.g., Egan, Deary & Austin, 2000; Ferguson & Patterson, 1998; Mooradian & Nezlek, 1996).

The NEO-FFI is of particular interest to the present studies. Unfortunately the number of studies examining the factor structure of the NEO-FFI are limited and, at times, fail to address the fit of the standard NEO-FFI model. For example, Ferguson and Patterson (1998) examined an alternative structure for the NEO-FFI. More specifically, they were interested in a two-factor model where neuroticism, extroversion, agreeableness, and conscientiousness make-up a single factor and the second factor is comprised of openness. Although Ferguson and Patterson reported fit indices for their two-factor orthogonal model ranging from a Bentler and Bonett Index score of .88 to a Delta2 score of .98 thereby suggesting a good fit, the method used to calculate the fit indices is unclear. It appears that the standard NEO-FFI model's chi-square was used in the calculation as opposed to the null model's chi-square. This may have inflated the fit-indices. Furthermore, the standard NEO-FFI model's fit was never tested thereby not allowing for a comparison between the two models.

In another study conducted by Mooradian ad Nezlek (1996) the standard NEO-FFI's model was tested using both principle component analysis and confirmatory factor analysis. Although principle component analysis produced solutions similar to those proposed by McCrae and Costa (1995), the confirmatory factor analysis produce a Bentler and Bonett index score of .66. Similar to the findings with the NEO-PI-R this fit index suggests a poor fit with the data.

This finding is somewhat surprising. The NEO-FFI was developed using the highest loading items from the original NEO-PI factors. One would expect that the elimination of the lowest loading items would produce a better fit. In fact, this is done when attempting to produce an empirically-derived model from an existing model. With this knowledge and the limited number of studies exploring the NEO-FFI's factor structure, the evidence is equivocal at best.

The Current Studies

The following studies attempt to address the relation between mood and personality measures with an emphasis on the temporal instructions. Although past research has attempted to study the structures of mood and personality, no study has attempted to study the factors of mood and personality using methods typically employed in mood and personality research (i.e., alternative instructional sets to assess both mood and the five factors of personality). The instructional sets that are to be used for this purpose will first be examined in Study 1.

In Study 2 it is expected that mood assessments (i.e., shorter retrospective timeframes) will produce a two-factor solution similar to those suggested by Watson and Tellegen (1985) and consisting of two orthogonal factors of positive and negative affect.[6]

[6] Due to the development of the measures used in the following studies orthogonal structures as opposed to oblique structures are expected.

One the other hand, a five-factor structure is expected to emerge with longer retrospective timeframes.

STUDY 1

The primary purpose of Study 1 was to examine the retrospective time periods taken into consideration by individuals when their mood or personality is being assessed using the instructional terms moment, right now, today, and in general. More specifically, Study 1 attempts to empirically define the retrospective period used to answer such questions as "Are you happy in general?" and "Are you sad right now?" Traditionally, when employing the term "in general" researchers assumed that individuals were reporting an average state, that is, how he/she most often feels. Conversely, when using the term "right now" researchers assumed that the individual was reporting an on-line mood state that implies a transient experience of a shorter duration.

This investigation is important because the above terms are used extensively to aid in the assessment of mood and personality (e.g., Costa & McCrae, 1985; Russell, Weiss, & Mendelsohn, 1980; Watson, Clark, & Tellegen, 1988; Zuckerman & Lubin, 1965). However, the actual retrospective period taken into consideration has not been defined. Unfortunately, with this lack of quantification, both methodological and theoretical problems arise.

Methodologically a number of problems are evident. Primarily, a majority of recent research in the areas of mood and personality has been preformed using similar timeframe instructional sets. However, with no empirical knowledge of the effects of these instructional sets (i.e., the independent variable), confidence in the results necessarily decreases. It would be synonymous to basing psychopharmacology theory on the effects of a few drugs with no knowledge of the constituents of the drugs being used in the research. Moreover, the lack of knowledge of the independent variable can directly effect the power of the analyses being preformed. A major problem in psychological research is the lack of statistical power (Cohen, 1990). Knowledge of the timeframes being utilized by individuals can increase the statistical power in the current research by allowing researchers to utilize the instructional set that will produce the greatest effect size (for a comprehensive review of power analysis refer to Cohen, 1992).

Theoretical models are intrinsically linked to research and these methodological pitfalls can directly effect our understanding of a phenomenon. This is of great importance when studying the relation between mood and personality. Specifically, mood (i.e., a short retrospective timeframe) is defined by a two-factor model while personality (i.e., a long retrospective timeframe) is defined by a five-factor model. However, if these factor structures are based on a misunderstanding of the retrospective timeframe being assessed then the differences between mood and personality may not be as clear cut as much of the research suggests.

A pilot study was performed using a simple four question measure to address the retrospective timeframe issue. Seventy-eight individuals were asked how much time they take into consideration when asked how they felt at this moment, right now, today, and in general. As can be seen in Figure 4, the results of the pilot study suggested that when the moment and today instructions are utilized, individuals take the shortest retrospective timeframe into consideration. The today retrospective timeframe was not significantly shorter than the retrospective timeframe for the right now instructional set. And, the three shortest retrospective timeframes (i.e., moment, right now, & today) were significantly shorter than the retrospective timeframe produced by the in general instruction.

The two most important findings from the pilot study were the length of the retrospective timeframes and their extensive variability. It was found that the retrospective timeframes reported were longer than would have been expected based on the extant literature. For example, individuals reported taking an average of 3.35 days into consideration when asked how they felt right now. Intuitively, one might except this retrospective timeframe to be only a few seconds. Moreover, the extensive variability in the individuals' responses suggests that there is no standard retrospective timeframe.

It is also noteworthy that a number of individuals did report having difficulty understanding the instructions in the pilot study. Specifically, many individuals reported that the instructions led them to believe that they were to report how long it took them to answer the question. This was thought to have contributed to the variability. Therefore, the current study attempted to better define the retrospective timeframes, and these results will be used in Study 2 to provide greater power when attempting to study the robustness of the factor structures of mood and personality.

METHOD

Participants

Participants consisted of 57 undergraduates enrolled in a 100 level introduction to psychology course. Participation was voluntary, and they received class credit for their cooperation.

Retrospective Timeframe per Instructional Set

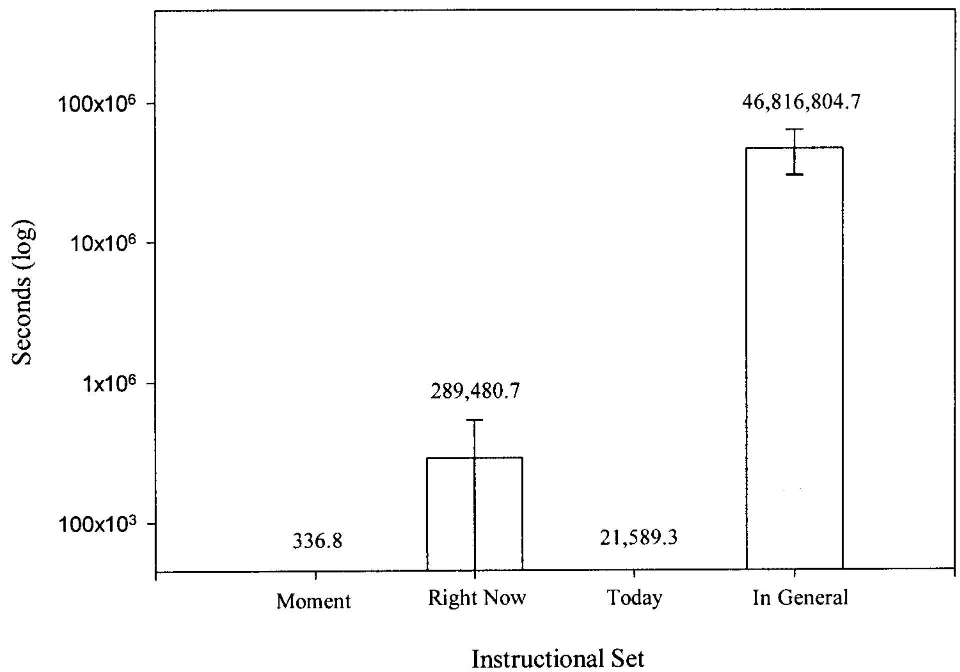

Figure 4. Pilot study retrospective timeframe (mean log seconds) and standard error for each instructional set.

Measures

Retrospective Timeframe Assessment: A questionnaire was developed to assess the retrospective timeframes employed when evaluating either positive or negative affect. The retrospective timeframe assessment consists of eight questions. The measure first requires a yes/no answer to a question addressing a mood or trait. For example, a participant would be asked, "Are you happy right now?" or , "Are you sad right now?" This was done for each of the four instructional sets. These questions were intended to operate as examples for the following question: "How much of your life did you take into consideration in your answer?" Participants were then asked to report a numerical value (e.g., 1, 2, 3, etc.) and then circle the appropriate timeframe (e.g., second(s), minute(s), hour(s), etc.). This was done for each of the four instructional timeframes.

Procedure

The measure was completed in a laboratory setting at the conclusion of an unrelated study. Half of the participants were asked to complete a measure assessing positive affect while the other half were asked to complete a measure assessing negative affect. Each measure consisted of the four instructional timeframes presented in a different order to minimize possible order effects. In an attempt to counterbalance for all possible orders, 96 different versions of the questionnaire were developed. Because there were 57 participants and 96 different versions of the questionnaire, each participant received a different version of the measure.

Data Analysis

All retrospective timeframes reported by the participants were first converted to a common measurement of seconds (e.g., 1 hour = 3600 sec.). Although the means and standard errors are reported in seconds, due to the nature of the data set (i.e., largely positively skewed), the data was further transformed using a log function prior to analysis (i.e., log(timeframe reported in seconds)). All comparisons between retrospective timeframes were done using a within subjects ANOVA on the transformed data. A Tukey's post hoc comparison was utilized to test for significant differences between the retrospective timeframes for each instructional set.

RESULTS

The retrospective timeframes utilized by individuals under the different instructional sets varied immensely. Most prominent, the in general instructional set had a range of 23 years. The smallest range (2.35 days) was associated with the today instructional set. An outlier analysis was performed prior to any further data analysis. No data point fell beyond three standard deviations from the mean and therefore no participants were removed from the following analyses.

A one-way within subjects ANOVA revealed a significant effect of instructional set; $F(59,165) = 7.61$, $p < .001$. As can be seen in Figure 5, post hoc analysis revealed that significant mean differences exist in the retrospective timeframes for the in general instructions compared to the today ($M_{diff} = 7.55$, $p < .05$), right now ($M_{diff} = 8.84$, $p < .05$), and moment ($M_{diff} = 11.02$, $p < .05$) instructional sets. Post hoc analyses also revealed that a significant difference exists between the moment and today instructions

Retrospective Timeframes per Instructional Set

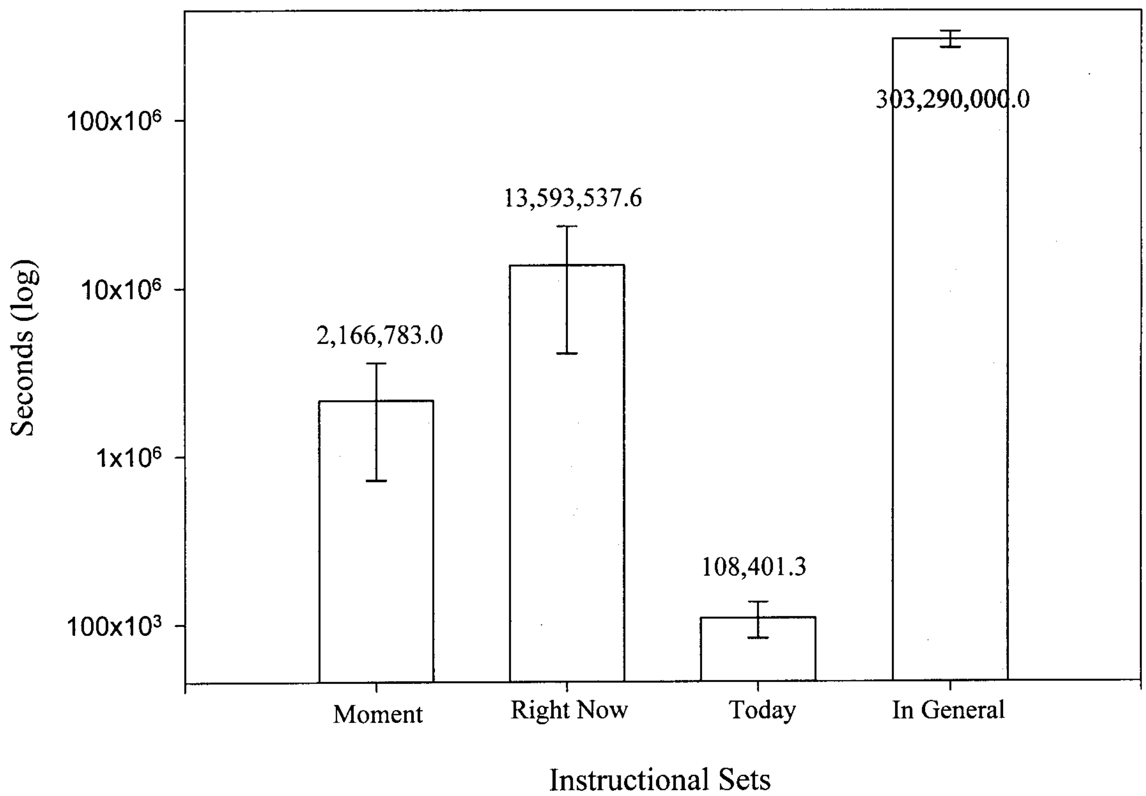

Figure 5. Retrospective timeframe (mean log seconds) and standard error for each instructional set.

($M_{diff} = 3.47$, $p < .05$). However, no significant differences were found between the moment and right now instructional sets or the right now and today instructional sets.

Due to concerns over the skewness of the distributions under investigation, the median and mode are also listed in order to provide another description of the retrospective timeframes utilized by individuals. As is apparent in Table 5, a trend similar to that found with the mean values was observed (although no statistical analysis was preformed). Specifically, the moment instructional set produced the shortest median (10 min.) and mode (1 sec.) while the in general instructions produced the longest median (15.5 yrs.) and mode (18 yrs.). Conversely, the today and right now instructions had equal modes (1 day) while differing with respect to their medians (1 day vs. 0.15 days respectively).

DISCUSSION

The results suggest that instructional sets most commonly employed in the personality and affect literature may be not be as intuitive as one may have believed. A seemingly standard tool (i.e., the instructions) appears to produce very different responses between individuals. In fact, the extensive variability within any particular instructional set raises questions as to the instruction's efficacy. More importantly, the average retrospective timeframes were found to be longer than any intuitively derived values. For example, studies of on-line mood states often utilize moment and/or right now instructional sets. However, these instructional sets appear to elicit responses corresponding to mood states experienced over days as opposed to the on-line experiences that might span several minutes or hours.

Table 5

Mean, median, and mode of the retrospective timeframes in seconds for each instructional set.

Instructional set	Mean	Median	Mode
Moment	2,166,783.0 (3.6 weeks)	600.0 (10 minutes)	1.0 (1 second)
Right now	13,935,537.6 (5.6 months)	12,600.0 (3.5 hours)	86,400.0 (1 day)
Today	108,401.3 (1.3 days)	86,400.0 (1 day)	86,400.0 (1 day)
In general	303,287,143.1 (10.5 years)	449,971,200.0 (15.5 years)	522,550,000.0 (18 years)

Note: Parenthesized timeframes were provided for greater conceptual understanding.

The differences that exist between the actual retrospective timeframes utilized by individuals and the theoretical timeframes believed to be assessed can influence our understanding of both mood and personality. For example, our understanding of the orthogonality of affect could be greatly influenced by the retrospective timeframes used by individuals when being assessed with measures employing similar instructional sets. Indeed, the apparent simultaneous experience of positive and negative affect may be a byproduct of the reporting timeframe, such that, affective experiences of the past few days are summarized as opposed to actual on-line affect experiences. Furthermore, the test-retest reliability of the reporting timeframes is unclear. Individuals may employ different retrospective timeframes at different times (i.e., different assessment periods). This may further explain the findings reported in the literature supporting both the orthogonal and oblique models of affect.

Although the results indicate that the mean retrospective timeframe used by individuals when assessed with the moment instructional set is approximately 25 days, it should be understood that the true retrospective timeframe may differ from the reported retrospective timeframe. That is, the current study attempts to better quantify self-reported retrospective timeframes, but does not measure the actual retrospective timeframes utilized by individuals. Indeed, attempting to assess the retrospective timeframes utilized is not an easy task due to the difficulties in formulating a question that is easily understood by individuals. Further studies addressing this issue are required before determining that the variability in the responses is genuine as opposed to attributing the variability to problems in the assessment. However, these findings along with those of the pilot study do suggest the existence of a disparity between the theoretical and actual retrospective timeframes employed by individuals.

STUDY 2

This study examined the structure of affect and personality as a function of the retrospective timeframe. With the support of the aforementioned literature, the following hypothesis was developed.

1 As the retrospective timeframe taken into consideration for the assessment of affective terms increases, the structure of affect that emerges will become more complex. More specifically, as the retrospective timeframe increases so will the number of factors that emerge. For example, when assessing mood (how you feel right now) a simplified factor structure should emerge. However, when assessing personality (how you feel in general) a more complex factor structure should emerge that approximates the well-established five-factor model of personality.

Although many other experimental questions may have arisen throughout the above discussion, the current study will focus primarily on the above hypothesis.

METHOD

Participants

The participants consist of a convenience sample of 832 undergraduate students. The participants consisted of students enrolled in introduction and advanced courses in psychology. Individuals enrolled in Introduction to Psychology received credit for their participation. A portion of the students in the advanced psychology courses received extra credit for their participation. All other participants completed the study voluntarily and without compensation. None of the participants in Study 1 took part in Study 2.

Measures

NEO-FFI (Costa & McCrae, 1989). The NEO-FFI is a shortened (60-item) form of the NEO Personality Inventory that assess neuroticism, extroversion, openness, agreeableness, and conscientiousness. The internal reliabilities range from r = .75 for conscientiousness to r = .89 for neuroticism. This scale has also been validated against the NEO-PI accounting for as much as 75% of the total NEO-PI variance.

Moment NEO-FFI. The Moment NEO-FFI is a revised version of the standard NEO-FFI. Each item was modified to begin with the phrase "At the moment." The instructions "At the moment" was chosen based on the findings of Study 1. Each item was then further modified so as to be expressed in the present tense. For example, item 1 on the standard NEO-FFI states, "I am not a worrier." Item 1 on the moment NEO-FFI states, "At the moment, I am not worried".

PANAS (Watson, Clark, & Tellegen, 1988). There were two versions of the PANAS used, both based on the short (20-item) form of the scale. The PANAS is intended to measure levels of positive and negative affect over a variety of time frames. Two different instructional sets were used (in general and moment). The internal reliabilities range from r = .84 for negative affect ("year" instructions) to r = .90 for positive affect ("today" instructions). This scale has also been validated against a number of other assessment tools such as the Beck Depression Inventory (Beck, 1961).

Supplemental PANAS items: The PANAS was also supplemented with 35 personality items drawn from the NEO-FFI Manual (Costa & McCrae, 1985). Items that were chosen for the supplement had to meet two criteria. First, the item had to be no more than two words (e.g., self-disciplined) or easily translated into two or fewer words as opposed to a phrase (e.g., avoids over-stimulation vs. seeking-excitement). This was done to insure consistency with the original PANAS items. Second, the items were required to be translatable into adjective form. For example, the supplemental term "full of ideas" was originally reported in the NEO-FFI manual as "imaginative".

Procedure

All participants were tested in a classroom setting. Participants were told that the purpose of the study was to explore the structure of a number of measures. They then completed one of the four measures. The measure that was completed was determined by a number of factors. The amount of time allocated to the researcher was the primary determinant. More specifically, when the researcher was only allowed the remaining few minutes (approximately 10 – 15 minutes) of class for data collection, participants received one of the two versions of the PANAS with the supplemental terms. When the researchers were allocated more than 15 minutes for data collection the participants were asked to complete one of the two versions of the NEO-FFI.

Data Analysis

Confirmatory factor analysis (CFA): This procedure was used in the present research to empirically evaluate the structure of the constructs under examination.

Unlike exploratory factor analysis (EFA), CFA allows one to specify a model and evaluate how well the data reflects that model. Moreover, CFA allows for a direct comparison of the fit of several models using the chi-square difference. A significant difference between the chi-square value of two models suggests a significant improvement in the fit of the model. If the chi-square values do not differ significantly, then the simplest model (the one using the fewest degrees of freedom) will be accepted as the best fitting model following the law of parsimony. The initial fit of the models directed the next step of the analysis.

Directions for analyses: If the predicted models for mood and personality fit the data, then the models listed below are compared using the corresponding measures. Unfortunately, there is currently no validated method for comparing identical models (i.e., models with equal degrees of freedom). Therefore, the relationship between the models must be explored somewhat subjectively. More specifically, this entails comparing each model's fit indices and the RMSRs. The model with the higher fit indices and lower RMSR is considered the better of the two models.

If the models fail to fit the data (i.e., they achieve a mean fit index less than .95), then a series of empirically – informed modifications will be explored using the CFA procedure. Specifically, each factor of a measure will be tested individually. Single items are removed from each factor based on the analysis of the item's standardized residuals, t-scores, and loadings. The chi-square of the model is recalculated after the removal of each item. If the chi-square difference is significant, the factor, minus the removed item or items, is reanalyzed followed by the removal of the next item. Although this is preformed until a significant chi-square difference no longer exists, the first model that reaches a mean fit index of .95 is considered the most appropriate model for further analysis[7]. The fit of the empirically-derived model is then quantified using CFA and chi-square difference is used to quantify its superiority over the original model.

If the empirically – defined models deviate substantially from the original, theoretical models, then comparisons between the mood and personality factor structures may not be possible. In this case, any comparisons that are possible with the original, theoretical models, are preformed on those models.

Personality and affect model comparisons: CFA was used to explore the relationship between the structure of affect and personality.

Three structural models will be tested for each instructional set:

1. Two orthogonal factors: Positive affect/extroversion and negative affect/neuroticism.
2. One bipolar factor: Positive affect/extroversion and negative affect/neuroticism all loading on a single factor with one set of items loading positively and the remaining items loading negatively.
3. Five orthogonal factors: Factors similar to the traditional "Big Five" of personality.

Measuring the fit of a model: The fit of the CFA models will be evaluated with the Root Mean Squared Residual (RMSR) and the following five fit indices (refer to Appendix A for fit index formulas): 1) Adjusted Goodness of Fit Index (AGFI; Bentler, 1983), 2) Bentler and Bonett Index (BBI; Bentler & Bonett, 1980), 3) Comparative Fit

[7] This was done to minimize the number of items removed from the theoretical model and limit possible confounds associated with item removal such as decreased reliability and validity.

Index (CFI; Bentler, 1989), 4) Delta2 (Bolen, 1989), 5) Tucker-Lewis Index (TLI; Tucker & Lewis, 1973). A mean of all of the fit indices will also be computed (MFI).

RESULTS AND DISCUSSION

NEO-FFI Five-Factor Model

The original 60 item NEO-FFI's five-factor model put forth by Costa and McCrae (1989) was tested with the scores provided by 237 participants using nested modeling CFA. As can be seen in Table 6, the over all fit of the five-factor model was poor with fit-indices ranging from a BBI score of .483 to a Delta2 score of .69 with a mean fit index of .625.

The relatively poor fit of the NEO-FFI has been found elsewhere in literature. In fact, in a study conducted by Mooradian and Nezlek (1995) exploring the five-factor structure of the NEO-FFI, they obtained a similar fit with a reported BBI of .66. The consistency of results between the two studies suggests that the poor fit of the NEO-FFI was caused by something other than random chance or sample-specific problems with the data (e.g., a non-representative sample).

NEO-FFI One-Factor Model: Law of Parsimony

The five-factor structure may be too complex. Although the five-factor structure is supported theoretically by the Big Five theory of personality, the Law of Parsimony suggests that, all things being equal, the simplest solution should be adopted as the correct solution. Therefore, due to the poor fit of the five-factor model a one-factor model was used as a test of a more parsimonious solution.

Table 6

Chi-square and fit indices for possible facorial models for the standard NEO-FFI

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Standard NEO-FFI									
Five factor	3818.46	1700	0.062	0.668	0.483	0.661	0.666	0.647	0.625
Two factor									
NEAC-O	4355.83	1709	0.081	0.497	0.303	0.409	0.417	0.388	0.403
N-E	578.68	251	0.074	0.794	0.695	0.798	0.801	0.777	0.773
One factor									
NEOAC	4835.18	1710	0.088	0.457	0.226	0.302	0.311	0.277	0.315
NE	949.23	252	0.108	0.592	0.499	0.569	0.576	0.528	0.553
Null									
Five factor	6246.56	1770							
Two Factor (NE)	1894.72	276							

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
** Mean fit index (MFI).

In this one-factor model, all 60 NEO-FFI items were positioned to load on a single factor and this factor structure was compared to the null model. As can be seen in Table 6, the fit of the one-factor model was also poor with the fit indices ranging from a BBI score of .226 to an AGFI of .457 and a mean fit index of .315. Furthermore, the χ^2 difference between the five-factor model and the one-factor model revealed that the one-factor model had a significantly poorer fit with the data than Costa and McCrae's five-factor model; $\chi^2_{diff} = 1616.72$, $df_{diff} = 10$, $p < .001$.

These results suggest that the structure of personality is more complex than a single factor. However, it also appears that the five-factor structure is inadequate. This inadequacy was explored in a number of ways. First two-factor models were tested to explore possible alternative factor structures for personality. Specifically, models that mirror the two-factor model of affect were tested. Second, the inadequacies were explored using a series of CFAs comparing each of the 12 item one-factor models individually against their respective null models. This analysis provided information regarding the location of potential problems (e.g., extremely poor fits for only a couple of factors vs. poor fitting models for all of the factors). Furthermore, if the fits for the single factor models are poor, then it should not be surprising that the five-factor model fails to replicate.

Exploring the NEO-FFI: Testing Two-Factor Models

A two-factor model was tested using a nested modeling CFA. This model was a replication of Ferguson and Patterson's (1997) earlier study. Recall, Ferguson and Patterson explored a two-factor model consisting of neuroticism, extroversion, agreeableness, and conscientiousness (48 items in all) on one factor, and openness to new experience (12 items) on the second factor. The rationale being that openness represents intelligence while the other four factors reflect personality.

The results obtained did not replicate the findings of Ferguson and Patterson's study. Specifically, the fit indices for this two-factor model ranged from a BBI score of .303 to an AGFI of .497 with a mean fit index of .403 (refer to Table 6 for a complete listing of fit indices). Not surprisingly, the current two-factor model resulted in significantly poorer fit than the five-factor model ($\chi^2_{diff} = 1016.72$, $df_{diff} = 10$, $p < .001$) as well as the corresponding one-factor model; $\chi^2_{diff} = 479.35$, $df_{diff} = 1$, $p < .001$.

A second two-factor model of personality focusing on neuroticism and extroversion was tested using nested modeling CFA. This model consisted of the 24 items that make-up the extroversion and neuroticism factors of the NEO-FFI. As can be seen in Table 6, although the fit indices for the two-factor model were higher than the five-factor model, they still did not reach the 0.90 level. Specifically, the fit indices ranged from a BBI score of .695 to a Delta2 score of .801 with a mean fit index of .773. However, due to the exploratory nature of the analysis of the NEO-FFI it becomes increasing important to test the new models against the five-factor model as opposed to relying entirely on the fit-indices themselves. That is, the relative fit of the model becomes central in the current research. With this in mind, it was found that the two-factor model of neuroticism and extroversion fit the data significantly better than the five-factor model; $\chi^2_{diff} = 3239.78$, $df_{diff} = 1449$, $p < .001$.

Consistent with the affect literature, a single factor consisting of the 24 items associated with neuroticism and extroversion was also tested. This model was then compared to the corresponding null model as well as the two-factor model described above. The neuroticism and extroversion one-factor model was found to have a poor fit, with fit indices ranging from a BBI score of .499 to an AGFI of .592 with a mean fit index of .553. As would be expected from the fit indices, the neuroticism and extroversion 24 item two-factor model was a significantly better fit than the neuroticism and extroversion 24 item one-factor model; $\chi^2_{diff} = 370.25$, $df_{diff} = 1$, $p < .001$. Although these findings are informative and are similar to the factor models of affect, they fail to address all of the components of the five-factor model.

Exploring the NEO-FFI: Testing Five One-Factor Models

The one-factor model for each of the five personality factors was compared to their respective null model using CFA. Table 7 provides the relevant information for each of the five models. The overall fit of each individual factor was better than the five-factor model, with the lowest fit index of any of the single factor model never dropping below the BBI score of .762 (for Agreeableness) and having a mean fit index of .85. However, it is also apparent that the fit of the single factors are not adequate. In fact, although the AGFI's for neuroticism, openness, and agreeableness were above the .90 level (.909, .907, .930 respectively), none of the other fit indices for the single factor models reach the .90 criterion denoting a good fitting model.

Although the indices of fit are relatively low, they do provide important information. Specifically, neuroticism (associated with negative affect) provided the highest and most stable fit-indices ranging from a BBI score of .827 to an AGFI score of .909 with a mean fit index of .874. Neuroticism was followed by openness with fit-

Table 7

Chi-square and fit indices for the standard NEO-FFI single facorial models

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Single Factor									
Neuroticism	138.57	54	0.065	0.909	0.827	0.885	0.887	0.860	0.8736
Null	802.48	66							
Extroversion	195.25	54	0.063	0.871	0.762	0.813	0.816	0.771	0.8066
Null	821.06	66							
Openness	135.78	54	0.059	0.907	0.803	0.869	0.871	0.840	0.858
Null	690.15	66							
Agreeableness	109.71	54	0.053	0.930	0.762	0.861	0.866	0.831	0.85
Null	453.35	66							
Conscientiousness	215.79	54	0.066	0.866	0.781	0.824	0.826	0.785	0.8164
Null	983.53	66							

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).

indices ranging from a BBI score of .803 to an AGFI score of .907 with a mean fit index of .858. The other three single factors of extroversion, agreeableness, and conscientiousness had slightly lower fit indices with the exception of the AGFI for openness. Moreover, extroversion provided the lowest fit-indices of the five single factors with scores ranging from .762 for the BBI to an AGFI score of .871 with a mean fit index of .807.

The extroversion one-factor model was one of the poorer fitting models that was tested. This was surprising as, historically, neuroticism and extroversion (i.e., the Big Two) have been reported to be the first two and most reliable factors to emerge when exploring the factor structure of personality (Costa & McCrae, 1985). Although extroversion was not the best fitting one-factor model, because of its association with positive affect, it was used along with neuroticism as a starting point to further explore the structure of personality.

An Empirical Attempt to Recover the Big Five using the NEO-FFI

Using a series of CFA's, the fit of the five individual factors of the NEO-FFI were explored and items extracted until there was no longer a significant improvement in the chi-square value between successive models. More specifically, single items were deleted from each model and tested against the previous model until the removal of an item no longer produced a significant change in the chi-square value (refer to the Method section for the criteria for item deletion). Although the models were tested until a nonsignificant chi-square difference was found, the first model to reach a mean fit index (i.e., (BBI+CFI+ Delta2+TLI+AGFI)/5) of .95 was considered the most appropriate model for further analyses. This also helps preserve a larger number of items which should, in turn, improve the scales' reliability and validity.

As can be seen in Table 8, the deletion of three neuroticism items (items 16, 21, and 46 from the NEO-FFI) produced a single factor with fit indices ranging from a BBI score of .920 to a Delta2 and CFI score of .980 with a mean fit index of .961. The empirically derived Neuroticism model provided a significantly better fit than the original Neuroticism model; $\chi^2_{diff} = 106.3$, $df_{diff} = 27$, $p < .001$.

The Extroversion scale required the deletion of five items before reaching the critical mean fit index (see Table 9). More specifically, items 7, 17, 22, 27, and 47 were deleted from the NEO-FFI extroversion scale to produce fit indices ranging from a BBI score of .961 to a Delta2 and CFI score of .990 with a mean fit index of .974 for the Extroversion scale (refer to Table RE for all fit indices for the Extroversion factor). The empirically derived Extroversion model was a significantly better fit than the original Extroversion model; $\chi^2_{diff} = 178.41$, $df_{diff} = 40$, $p < .001$.

The Openness factor of the NEO-FFI required the deletion of three items before the fit indices reached the critical level. As can be seen in Table 10, the removal of items 38, 43, and 53 produced fit indices ranging from a BBI score of .913 to Delta2 and CFI scores of .994 with a mean fit index of .97. The empirically derived Openness model was a significantly better fit than the original Openness model; $\chi^2_{diff} = 106.9$, $df_{diff} = 27$, $p < .001$.

As can be seen in Table 11, a total of two items required removal from the Agreeableness factor before the mean fit index reached the critical level. The removal of

Table 8

Chi-square, chi-square difference, and fit indices for the standard neuroticism factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	Fit Indices: BBI	Fit Indices: CFI	Fit Indices: Delta2	Fit Indices: TLI	Fit Indices: MFI**
12	-	138.57	54	-	-	-	0.081	0.868	0.827	0.885	0.887	0.860	0.865
11	16	74.78	44	63.70	10	0.001	0.054	0.921	0.884	0.948	0.949	0.934	0.927
10	46	56.77	35	18.11	9	0.05	0.051	0.931	0.896	0.956	0.957	0.944	0.937
9	21	32.27	27	21.51	8	0.01	0.036	0.950	0.920	0.980	0.980	0.973	0.961
8	1	22.03	20	13.22	7	0.10	0.021	0.960	0.943	0.994	0.994	0.992	0.977
7	6	11.68	14	10.35	6	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 9

Chi-square, chi-square difference, and fit indices for the standard extroversion factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	195.25	54	-	-	-	0.104	0.813	0.762	0.813	0.816	0.771	0.795
11	47	135.32	44	59.93	10	0.001	0.093	0.851	0.809	0.860	0.862	0.825	0.841
10	22	86.37	35	48.95	9	0.001	0.078	0.894	0.853	0.906	0.907	0.879	0.888
9	27	57.87	27	28.50	8	0.001	0.069	0.916	0.886	0.935	0.936	0.913	0.917
8	17	35.11	20	22.76	7	0.005	0.056	0.936	0.911	0.959	0.960	0.942	0.942
7	7	16.84	14	18.27	6	0.025	0.029	0.961	0.946	0.990	0.990	0.985	0.974
6	52	4.16	9	12.68	5	.025	0.000	0.987	0.982	1.023	1.022	1.038	1.010
5	2	1.47	5	2.69	4	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.

**Mean fit index (MFI).

Null models used in calculation of the fit indices are not reported in Table.

Table 10

Chi-square, chi-square difference, and fit indices for the standard openness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	135.78	54	-	-	-	0.079	0.865	0.803	0.869	0.871	0.840	0.850
11	43	69.23	44	66.55	10	0.001	0.049	0.923	0.852	0.939	0.940	0.923	0.915
10	53	43.24	35	25.99	9	0.005	0.031	0.942	0.882	0.974	0.975	0.967	0.948
9	38	28.88	27	14.36	8	0.01	0.017	0.957	0.913	0.994	0.994	0.992	0.970
8	13	16.67	20	12.21	7	0.01	0.000	0.969	0.917	1.019	1.018	1.027	0.990
7	18	7.77	14	8.90	6	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 11

Chi-square, chi-square difference, and fit indices for the standard agreeableness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	109.71	54	-	-	-	0.066	0.899	0.758	0.856	0.860	0.824	0.839
11	29	74.14	44	35.57	10	0.001	0.534	0.923	0.817	0.914	0.917	0.893	0.893
10	34	40.35	35	33.79	9	0.001	0.025	0.949	0.879	0.981	0.982	0.976	0.953
9	49	20.78	27	19.57	8	0.025	0.000	0.969	0.926	1.026	1.025	1.034	0.996
8	54	9.46	20	11.32	7	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

items 29 and 34 from the NEO-FFI produced fit indices ranging from a BBI of .879 to a Delta2 of .982 with a mean fit index of .953. The empirically derived Agreeableness model of the NEO-FFI was a significantly better fit than the original Agreeableness model of the NEO-FFI; χ^2_{diff} =69.35, df_{diff} = 19, $p < .001$.

Similar to the Extroversion factor, the Conscientiousness factor of the NEO-FFI required the removal of five items before the fit indices reached the critical value. As can be seen in Table 12, the removal of items 5, 25, 30, 35, and 40 produced fit indices ranging from a BBI score of .938 to a Delta2 score of .977 with a mean fit index of .961. The empirically derived Conscientiousness model yeilded a significantly better fit than the original Conscientiousness model; χ^2_{diff} = 194.04, df_{diff} = 40, $p < .001$.

The fit of five-factor model was again tested using the five empirically derived single factors. As can be seen in Table 13, the overall fit of the empirically derived CFA model was poor. Specifically the fit indices for the empirically derived model ranged in value from a BBI score of .514 to a Delta2 score of .705 with a mean fit index of .668. However, there was a significant decrease in the chi-square value of the empirically derived model compared to the original model; χ^2_{diff} = 1767.35, df_{diff} = 891, $p < .001$. Comparison of the mean fit indices for the empirically derived and original models suggests only a slight increase in fit with neither of the mean fit indices approaching the .90 criterion for a good fit.

Recall, the primary purpose of this study was to explore the effects of instructional changes (i.e., from in general to this moment) on the factor structure of a common personality assessment. More specifically, it was hypothesized that a better fit would be found for models similar to the models of affect found in the literature when

Table 12

Chi-square, chi-square difference, and fit indices for the standard conscientiousness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	215.79	54	-	-	-	0.112	0.806	0.781	0.824	0.826	0.785	0.804
11	5	161.42	44	54.37	10	0.001	0.106	0.831	0.815	0.857	0.859	0.821	0.837
10	40	114.10	35	47.32	9	0.001	0.097	0.856	0.854	0.893	0.894	0.862	0.872
9	35	58.26	27	55.84	8	0.001	0.070	0.918	0.897	0.941	0.942	0.921	0.924
8	30	37.67	20	20.59	7	0.005	0.061	0.932	0.923	0.962	0.962	0.946	0.945
7	25	21.75	14	15.92	6	0.025	0.048	0.951	0.938	0.976	0.977	0.965	0.961
6	10	8.88	9	12.87	5	0.05	0.000	0.972	0.966	1.001	1.001	1.001	0.988
5	60	3.6	5	5.28	4	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 13

Chi-square and fit indices for original and empirically derived models for the standard and moment NEO-FFI

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Standard NEO-FFI									
Original Five factor	3818.46	1700	0.062	0.668	0.483	0.661	0.666	0.647	0.625
Empirical Five factor	1451.11	809	0.058	0.746	0.514	0.698	0.705	0.679	0.668
Moment NEO-FFI									
Original Five factor	3038.59	1700	0.060	0.661	0.463	0.656	0.662	0.642	0.617
Empirical Five factor	1061.73	692	0.049	0.783	0.575	0.790	0.795	0.775	0.744

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
** Mean fit index (MFI).

moment instructions were employed. Furthermore, it was hypothesized that the personality models tested (i.e., the standard five-factor model) would provide a better fit than the affect models when the in general instructions were employed. Although the fit of the original NEO-FFI was extremely poor, it was important to test the moment NEO-FFI's model. This allowed for a number of comparisons to be preformed to address the aforementioned hypotheses.

Moment NEO-FFI Five-Factor Model

The revised NEO-FFI using the moment instruction was tested with the scores provided by 192 participants using nested modeling CFA. Specifically, the NEO-FFI theoretical model was evaluated relative to the null model. As can be seen in Table 14, similar to the original NEO-FFI's fit, the five-factor model produced fit indices ranging from a BBI score of .463 to a Delta2 score of .662 with a mean fit index of .617. As a result of the moment instructions providing an assessment of current mood state (as opposed to personality), the poor fit of the five-factor model may suggest, much like it did with the original NEO-FFI, that the factor structure might consist of fewer factors. As with the original NEO-FFI's five-factor structure, we subjected the moment NEO-FFI to the same series of single factor tests.

Moment NEO-FFI One Factor Model: Law of Parsimony

CFA was used to test a one-factor model in which all 60 items were hypothesized to relate to a single construct. The null model was also used in the calculation of the fit indices. As can be seen in Table 15, the fit of the one-factor model was poor.

Table 14

Chi-square and fit indices for possible facorial models for the moment NEO-FFI

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Moment NEO-FFI									
Five factor	3038.59	1700	0.060	0.661	0.463	0.656	0.662	0.642	0.617
Two factor									
NEAC-O	3598.79	1709	0.701	0.571	0.364	0.514	0.522	0.497	0.494
N-E	538.61	251	0.072	0.791	0.714	0.821	0.824	0.804	0.791
One factor									
NEOAC	3866.78	1710	0.076	0.539	0.317	0.445	0.454	0.426	0.436
NE	671.42	252	0.087	0.736	0.644	0.740	0.743	0.715	0.716
Null									
Five factor	5659.56	1770							
Two facor (NE)	1885.09	276							

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).

Specifically, fit indices for the one-factor moment NEO-FFI ranged from a BBI score of .317 to an AGFI of .539 with a mean fit index of .436. Furthermore, the chi-square difference between the five-factor model and the one-factor model achieved statistical significance such that the five-factor model offered a significantly better fit with the data ($\chi^2_{diff} = 828.19$, $df_{diff} = 10$, $p < .001$). Supporting the extant literature, these results suggest that the structure of affect is more complex than a single factor.

Exploring the Moment NEO-FFI: Testing Two-Factor Models

Using nested modeling CFA a 24 item two-factor (i.e., neuroticism and extroversion) model of affect was tested against the respective null and one-factor models. As can be seen in Table 14, similar results to the original NEO-FFI were found with the moment NEO-FFI. The fit-indices were consistently higher for the two-factor model compared to the five-factor model however, the values did not reach the .90 level. Furthermore, it was found that the two-factor model of neuroticism and extroversion fit the data significantly better than the five-factor model; $\chi^2_{diff} = 2499.98$, $df_{diff} = 1449$, $p < .001$.

Using a nested modeling CFA this model was further compared to the corresponding null and one-factor model. Results (see Table 14) suggest that the overall fit of the one-factor model was relatively poor, with fit indices ranging from a BBI score of .644 to a Delta2 score of .743 with a mean fit index of .716. Furthermore, a significant chi-square difference was found between the one and two-factor models ($\chi^2_{diff} = 132.82$, $df_{diff} = 1$, $p < .001$) with the two factor model providing a better fit with the data. Although the two-factor model with the moment instructions is similar to the theoretical structure of affect found in the literature, the fit of the model fails to reach the critical level of .90. Due to this failure, additional testing for each of the one-factor models was conducted.

Exploring the Moment NEO-FFI: Testing Five Single Factor Models

Confirmatory factor analysis was employed to test the five one-factor models and these were compared to their respective null models. As can be seen in Table 15, the overall fit of each individual factor was better than the moment NEO-FFI's fit-indices. Interestingly, Agreeableness had both the lowest fit index (BBI = .696) and the highest fit index (AGFI = .891). With respect to the other four single factors, their fit indices ranged from a BBI score of .716 for openness to a Delta2 score of .873 for neuroticism. These fit index ratings suggest that they are, again, not adequate as they all fall well below the .90 threshold.

Interestingly, the two factors that provide the highest and most stable fit-indices are neuroticism (BBI = .822 to Delta2 = .873, mean fit index = .850) and extroversion (BBI = .784 to AGFI = .866, mean fit index = .837) factors. This finding supports the claim that the moment NEO-FFI was assessing affect more than personality, as the two factors most associated with affect are also the two factors that appear to fit the data the best. Similar to the previous analyses using the original NEO-FFI, extroversion (positive affect) and neuroticism (negative affect) were used as a starting point to further explore the structure of the revised moment NEO-FFI.

Table 15

Chi-square and fit indices for the moment NEO-FFI single facorial models

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Single Factor									
Neuroticism	164.43	54	0.035	0.840	0.822	0.871	0.873	0.842	0.8496
Null	921.18	66							
Extroversion	135.18	54	0.034	0.866	0.784	0.855	0.858	0.823	0.8372
Null	626.26	66							
Openness	146.08	54	0.038	0.848	0.716	0.794	0.800	0.749	0.7814
Null	513.86	66							
Agreeableness	107.00	54	0.033	0.891	0.696	0.815	0.822	0.774	0.7996
Null	352.21	66							
Conscientiousness	158.82	54	0.034	0.849	0.785	0.845	0.847	0.810	0.8272
Null	739.99	66							

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).

An Empirical Attempt to Recover the Big Five in the Moment NEO-FFI

The fit of the five individual factors of the Moment NEO-FFI was explored using a series of nested CFA's. More specifically, items were extracted until there was no longer a significant chi-square difference between models. Similar to the fit evaluation of the original NEO-FFI, the first model to reach a mean fit index of .95 was considered the most appropriate model for further analysis.

Table 16 shows that the removal of neuroticism items 21, 31, 46, and 56 produced a single factor with fit indices ranging from a BBI score of .93 to Delta2 and CFI scores of .98 with a mean fit index of .96. The empirically derived moment Neuroticism model was a significantly better fit than the original moment Neuroticism model; χ^2_{diff} = 135.47, df_{diff} = 34, $p < .001$.

The moment Extroversion scale also required the removal of four items (12, 22, 37, & 47) before reaching the critical mean fit index. As can be seen in Table 17, the removal of these four items produced fit indices ranging from a BBI score of .922 to a Delta2 score of .981 with a mean fit index of .961. The empirically derived moment Extroversion model was a significantly better fit than the original moment Extroversion model; χ^2_{diff} = 109.11, df_{diff} = 34, $p < .001$.

The critical mean fit index was also reach after the removal of four items (3, 8, 23, & 53) from the moment Openness factor. Table 18 shows fit indices ranging from a BBI score of .896 to Delta2 and CFI scores of .982 with a mean fit index of .958. The empirically derived moment Openness model was a significantly better fit than the original moment Openness model; χ^2_{diff} = 122.44, df_{diff} = 34, $p < .001$.

Table 16

Chi-square, chi-square difference, and fit indices for the moment neuroticism factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference difference	chi-square difference df	p-value	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	164.43	54	-	-	-	0.095	.840	0.822	0.871	0.873	0.842	0.850
11	10	114.59	44	49.84	10	0.001	0.084	.867	0.834	0.889	0.891	0.861	0.868
10	12	67.29	35	47.3	9	0.001	0.064	.907	0.884	0.939	0.941	0.922	0.919
9	7	46.09	27	21.2	8	0.01	0.056	.927	0.903	0.956	0.957	0.942	0.937
8	5	28.96	20	17.13	7	0.025	0.045	.948	0.929	0.977	0.977	0.967	0.960
7	1	14.42	14	14.54	6	0.025	0.012	.966	0.960	0.999	0.999	0.998	0.984
5	11	6.79	9	7.63	5	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 17

Chi-square, chi-square difference, and fit indices for the moment extroversion factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	135.18	54	-	-	-	0.082	0.866	0.784	0.855	0.858	0.823	0.837
11	8	93.92	44	41.26	10	0.001	0.071	0.895	0.802	0.881	0.884	0.852	0.863
10	10	60.94	35	32.98	9	0.001	0.057	0.920	0.859	0.933	0.935	0.914	0.912
9	3	42.84	27	18.10	8	0.025	0.051	0.931	0.892	0.956	0.957	0.941	0.935
8	5	26.07	20	16.77	7	0.025	0.037	0.950	0.922	0.980	0.981	0.972	0.961
5	11	17.35	5	8.72	4	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 18

Chi-square, chi-square difference, and fit indices for the moment openness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	146.08	54	-	-	-	0.087	0.848	0.716	0.794	0.800	0.749	0.781
11	2	98.90	44	47.18	10	0.001	0.074	0.883	0.788	0.866	0.870	0.833	0.848
10	5	66.64	35	32.26	9	0.001	0.063	0.912	0.796	0.888	0.892	0.856	0.869
9	11	39.22	27	27.42	8	0.001	0.045	0.939	0.848	0.945	0.947	0.927	0.921
8	1	23.64	20	15.58	7	0.05	0.028	0.957	0.896	0.982	0.982	0.975	0.958
7	10	9.74	14	13.90	6	0.05	0.000	0.977	0.937	1.032	1.030	1.048	1.005
6	8	3.91	9	5.83	5	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

As can be seen in Table 19, the moment Agreeableness factor required the removal of items 9, 29, 49, and 59 in order to reach the critical fit index level of .95. The removal of these four items from the moment Agreeableness factor produced a factor with fit indices ranging from a BBI score of .881 to Delta2, CFI, and TLI scores of approximately 1.0 with a mean fit index of .968. The empirically derived moment Agreeableness model was a significantly better fit than the original moment Agreeableness model; $\chi^2_{diff} = 86.13$, $df_{diff} = 34$, $p < .001$.

The moment Conscientiousness factor required the removal of four items (10, 35, 40, & 50) to reach the critical mean fit index. The new moment Conscientiousness model's fit indices ranged from a BBI score of .903 to a Delta2 score of .973 with a mean fit index of .951 (refer to Table 20). The empirically derived moment Conscientiousness model was a significantly better fit than the original moment Conscientiousness model; $\chi^2_{diff} = 131.77$, $df_{diff} = 34$, $p < .001$.

The five new moment factors were then combined in order to retest the moment five-factor structure using the 30 remaining items. Similar to findings using the original NEO-FFI, the empirically derived moment model's fit was poor. As can be seen in Table 13, the fit indices range from a BBI score of .575 to a Delta2 score of .80 with a mean fit index of .745. However, a significant difference was found between the empirically derived moment model and the moment model; $\chi^2_{diff} = 1976.86$, $df_{diff} = 1008$, $p < .001$. The difference between the two model's mean fit indices suggests that the overall improvement was relatively small. Furthermore, neither of the mean fit indices approaches the .90 level of significant.

Table 19

Chi-square, chi-square difference, and fit indices for the moment agreeableness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	107.00	54	-	-	-	0.066	0.891	0.967	0.815	0.822	0.774	0.854
11	11	77.28	44	29.72	10	0.001	0.058	0.910	0.733	0.856	0.864	0.822	0.837
10	10	51.87	35	25.41	9	0.005	0.046	0.935	0.765	0.904	0.909	0.877	0.878
9	2	35.04	27	16.83	8	0.05	0.036	0.947	0.813	0.947	0.950	0.929	0.917
8	6	20.13	20	14.91	7	0.05	0.005	0.963	0.881	0.999	0.999	0.999	0.968
7	7	9.99	14	10.14	6	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
** Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Table 20

Chi-square, chi-square difference, and fit indices for the moment conscientiousness factor on the NEO-FFI.

Number of items used in model	Item deleted from model	CFA Model χ^2	df	chi-square difference: difference	chi-square difference: df	p-value	RMSR*	Fit Indices: AGFI	BBI	CFI	Delta2	TLI	MFI**
12	-	158.82	54	-	-	-	0.093	0.849	.0785	0.844	0.847	0.810	0.686
11	7	105.58	44	53.24	10	0.001	0.079	0.877	0.814	0.880	0.883	0.850	0.861
10	2	61.86	35	43.72	9	0.001	0.058	0.917	0.867	0.936	0.937	0.918	0.915
9	8	40.67	27	21.19	8	0.01	0.047	0.935	0.894	0.961	0.962	0.948	0.940
8	10	27.05	20	13.62	7	0.10	0.040	0.948	0.903	0.972	0.973	0.961	0.951
7	1	14.31	14	12.74	6	0.05	0.010	0.964	.943	0.999	0.999	0.998	0.981
6	3	3.04	9	11.27	5	0.05	0.000	0.990	0.986	1.030	1.030	1.050	1.017
5	4	1.20	5	1.84	4	ns	-	-	-	-	-	-	-

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).
Null models used in calculation of the fit indices are not reported in Table.

Comparing the Original and Moment NEO-FFIs

The extremely poor fit of the original NEO-FFI limits the specific questions that can be addressed in the present investigation. Although it was possible to empirically derive new models for the comparisons, the number of confounds associated with performing comparisons between the empirically derived affect and personality models would make such comparisons less helpful. For example, when comparing the original models the same items could be utilized in the comparison. The empirical exploration of the in general and the moment NEO-FFIs created two new factor structures comprised of two different sets of items. Furthermore, although the empirically derived models produced a better fit than the original models, the fit of the models was still poor. More importantly, the comparison of the empirically derived models should follow a validation of the models. With the validity of the empirically derived scales unknown, comparisons between the original models were deemed more appropriate.

It is essential to remember that the poor fit of the original NEO-FFI limits the conclusions that can be drawn from these results. Furthermore, there is currently no accepted statistical procedure to test the difference between factor structures of two different assessments. Therefore, the most appropriate study of differences between the two models consists of comparing the mean fit indices, the variability of the fit indices, as well as the RMSR.

As can be seen in Table 21, the differences between the standard and moment five-factor models are minimal. The standard NEO-FFI provides only a slightly better mean fit index (.625) compared to the moment NEO-FFI (.617). More importantly, the root mean squared residuals are almost identical. This finding is difficult to interpret.

Table 21

Chi-square and fit indices for the standard and moment NEO-FFI five, two, and one factor models

Model	CFA Model χ^2	df	RMSR*	Fit Indices AGFI	BBI	CFI	Delta2	TLI	MFI**
Five Factor Model									
Standard	3818.46	1700	0.062	0.668	0.483	0.661	0.666	0.647	0.625
Moment	3038.59	1700	0.060	0.661	0.463	0.656	0.662	0.642	0.617
Two Factor Model									
Standard	578.68	251	0.074	0.794	0.695	0.798	0.801	0.777	0.773
Moment	538.61	251	0.072	0.791	0.714	0.821	0.824	0.804	0.791
One Factor Model									
Standard	949.23	252	0.108	0.592	0.499	0.569	0.576	0.528	0.553
Moment	671.42	252	0.087	0.736	0.644	0.740	0.743	0.715	0.716

*Unlike other measures of fit, the RMSR suggests a better fit as the value approaches zero.
**Mean fit index (MFI).

The moment two-factor model provided a better fit than the standard two-factor model (refer to Table 21). However, similar to the five-factor model, the difference between the two-factor models was marginal. Specifically, a difference of .018 was found between the mean fit indices with approximately identical RMSRs. A larger difference was found between the one-factor models. Table 21 shows that the mean fit index for the moment one-factor model was larger (.716) than the standard one-factor model (.553). Importantly, the RMSRs also showed a difference with the moment one-factor model yielding a smaller RMSR (.087) than the standard one-factor model (.108).

Confirmatory and Exploratory Factor Analysis of the Revised PANAS

Using confirmatory factor analysis the five-factor models for both the in general and the moment instructional sets for the PANAS failed to converge. That is, the model's fit was so poor that the computer could not calculate the appropriate statistics based on the variance-covariance matrix. To provide more information on the factor structure that underlies the revised PANAS, exploratory factor analysis using both principle component analysis with orthogonal (i.e., "varimax") rotations and unweighted least squares factor analysis using oblique (i.e., "oblimin") rotations were preformed on both (i.e., "moment" and "in general") versions of the revised PANAS.

Interestingly, regardless of the exploratory factor analytic procedure employed 13 factors (i.e., eigenvalues > 1.0) accounting for over 60 percent of the variance were extracted when exploring both versions of the revised PANAS. Clearly, this is a much more complex factor structure than the two or five-factor solution suggested by the literature. However, a more appropriate examination of the results may be to look for a clear break in the eigenvalues obtained in each exploratory analysis (Reise, Waller, & Comery, 2000). Specifically, a more reliable solution (i.e., number of factors) can be obtained by accepting the factors that emerge with the highest eigenvalues and disregarding the factors that emerge after a substantial drop in the eigenvalues. Indeed, factors with small eigenvalues are more likely to be spurious factors that will emerge even when factor analyzing sets of random numbers (Reise et al., 2000).

When employing the above mentioned method the split between factors becomes obvious. Once again, all versions of the revised PANAS, independent of the rotation or extraction utilized, produced similar factor structures such that two factors yielded substantial eigenvalues. Both factors, which appear to represent PA and NA, consistently emerged with eigenvalues greater than five while the remaining 11 factors never yielded eigenvalues over three. Furthermore, these two factors consistently account for over 30 percent of the variance.

This finding lends some support to both the "Big Two" of personality as well as the current theories of affect. More importantly, the similar findings with the right now and in general instructional sets suggest a close relationship between the factor structure of affect and personality. Using an independent sample, a confirmatory factor analysis must be preformed on the factor structure extracted from this exploratory analysis before drawing any definitive conclusions.

GENERAL DISCUSSION

The comparison of the original NEO-FFI's five one-factor models to the five-factor model was intriguing. Although the one-factor models appear to fit the data better than the five-factor models, none of the single factors could be considered a good fit for their respective constructs. Recall, the NEO-FFI was developed using the highest loading items from the NEO-PI. Following this, one would expect that the items which make up the single factors of the NEO-FFI would all load highly, thereby producing a good fit. According to the results of the current study, this is not the case. In fact, a number of items (approximately 33%) required removal from each factor in order to produce satisfactory fit indices for the empirically derived single factors. Given this finding, it is not surprising that when all five factors were analyzed together the five-factor model produced poor fit indices.

In some instances, poor fit indices may be the result of selecting an index that is inappropriate for the sample characteristics. For example, the Delta2 has been shown to be a good estimator for small sample sizes compared to the Goodness of Fit Index that is considered less accurate with small data sets (Boland, 1989). Although many of the fit indices are calculated in a similar way, each is considered to have its strengths and weaknesses. Using sample size as an example again, the Tucker-Lewis Index is known to be less influenced by sample size than other fit indices. However, the Tucker-Lewis Index is also considered to be less accurate than other fit indices (Ferguson & Patterson, 1998). Given that the current study employed a number of fit indices with various strengths and weaknesses, the overall pattern of the fit indices and the mean fit index should have provided an accurate appraisal of the overall fit of the models. Therefore, the well defined pattern of fit indices and the acceptance of confirmatory factor analysis as an appropriate methodology to explore personality data (e.g., Borkenau & Ostendorf, 1990; Panter, Tanaka, & Hoyle, 1994) suggest that another variable must be responsible for the poor fit of the model.

The poor fit indices that were obtained when analyzing the empirically derived five-factor model is particularly interesting. Since the empirically derived single factors could be considered a good fit to the data, the poor fit of the empirically derived five-factor model suggests that there are a number of cross loading items. Many of these cross loadings have been documented. For example, McCrae and Costa (1989) report that hostility loads highly on both neuroticism and agreeableness. Furthermore, McCrae and Costa (1989) have suggested that many of the descriptive terms used in personality assessment fall in between two factors when the terms are presented in a circular model. Recall, Russell (1980) developed one of the more prominent theories of affect using the circumplex technique. Within his theory, positive and negative experience constitute a single factor. Unfortunately, the circumplex methodology has been underutilized and at times disregarded within the personality literature (e.g., McCrae & Costa, 1995) leaving researchers only two options. First, researchers can speculate as to what employing such techniques may suggest about the true structure of personality. For instance, the circumplex method may suggest that personality is not consistent with the orthogonal structure that it is now widely accepted. In fact, such a technique may suggest a structure very similar to that of affect (i.e., similar to Russell's, 1980, model of affect) with personality being comprised of an average level of affect as well as an average level of arousal. However, until such research has been preformed current theorist must rely upon the second option; the more acceptable methods of factor analysis and analysis of fit.

The Orthogonality of Personality

Theoretically, personality consists of a number of orthogonal factors. The empirical literature on the structure of personality suggests otherwise. McCrae et al. (1996) reported that when the NEO-PI-R was analyzed with orthogonal confirmatory factor analysis that allowed for cross loadings greater than .20, the fit indices increased. Specifically, the CFI score elevated from a .55 for the simple structure model to a .79 for modest loadings model (i.e., >.2). Similar findings were found when the NEO-PI-R was analyzed using confirmatory factor analysis on oblique models (McCrae et al., 1996).

Although the oblique model as well as models allowing for cross loadings did not produce a good fit, the fit of each model is clearly better than the original orthogonal five-factor model. The fit of the models in the current project may also have been affected by similar cross loadings. While these alternative models were not directly tested in the current study, the number of items that were removed from the individual factors in order to achieve a good fit (recall large residuals and low loadings were a requirement for removal), suggest that these items may have loaded higher on alternative factors (i.e., cross loading).

If the cross-loading of items is not an artifact (i.e., not due to a methodological confound such as an inappropriate scaling system), it could be explained by the fact that the NEO-FFI could be assessing a construct more similar to the structure of affect as opposed to the theoretical structure of personality. Therefore, an orthogonal five-factor model would be unsuitable and produce low fit indices. This finding persisted when the empirically derived standard NEO-FFI was explored. Thus suggesting that the standard

NEO-FFI, even with empirical modifications, does not support the orthogonality of the five-factor model.

There are possible alternative explanations for the reported findings. For example, McCrae et al. (1996) suggest that confirmatory factor analysis may not be an appropriate tool for analyzing personality data. Although McCrae and associates are correct in reporting that the "CFA users are generally cautioned to test only clearly specified models," (p.553) they seemingly fail to believe that item analysis of a presumably well validated measure falls within this definition. Furthermore, McCrae and associates (1996) suggest that exploratory factor analysis with Procrustes rotations along with Monte Carlo simulations provide better evidence than the well-accepted confirmatory factor analysis and the use of indices of fit. Specifically, McCrae et al. suggest that the repeated exploratory findings of the five-factor model lends enough support for the model so as to off-set the poor fit indices that emerge when using confirmatory analysis.

There are a number of problems associated with McCrae et al.'s (1996) position on this point. Most problematic is their argument that confirmatory factor analysis is an improper to evaluate the factor structure of personality. In fact, confirmatory factor analysis has been readily used by quantitative psychologist to explore this issue (e.g., Panter, Tanaka, & Hoyle, 1994). Moreover, personality was originally considered to consist of as many as 16 factors (see Cattell, 1965). As the precision of factor analytic techniques increased, researchers found that the 16 factor model was inadequate (Digman, 1990). The "Big Five" theory of personality was derived in this way (i.e., from Cattell's original 16 factor structure). The poor fit of the "Big Five" model may not be due to confirmatory factor analysis being an inappropriate method but due to the increase in precision of current factor analytic techniques. More specifically, the same problems (e.g., poor fit) that were associated with the 16 factor model may, with current techniques, also be found with the five factor model.

McCrae et al. (1996) made another major mistake in their logic. Specifically, the five-factor model that is reported to continuously emerges in exploratory factor analysis does so with specific factor rotations (i.e., Validimax, Varimax rotations). The repeated forcing of orthogonal factors does not necessarily suggest a valid factor structure. If the same orthogonal five-factor structure emerged with oblique rotations, the theory would find more support. However, this alone would not validate the model.

Evidence from Study 1 suggests the existence of other possible problems with the NEO-FFI. For example, the results of Study 1 suggests that large variability exists in the retrospective timeframe used by individuals when assessed with in general instructional sets. Although the NEO-FFI does not specifically employ a timeframe instructional set, it does use general questions to assess the individual's average state. Given that the NEO-FFI's questions are less specific with regard to the retrospective timeframe than the in general instructional set tested in Study 1, and that each item utilizes a different retrospective timeframe (i.e., there is not a standard timeframe instruction), these items may produce greater variability with respect to the retrospective timeframe utilized by individuals. Following this logic, the probability of individuals using relatively short retrospective timeframes would increase when being assessed with the NEO-FFI, thereby decreasing the possibility of recalling both positive (i.e., extroversion) and negative (i.e., neuroticism) affectivity charged memories simultaneously. The inability to recall both positive and negative memories simultaneously would produce a more bipolar model (e.g., extroversion-neuroticism continuum). In turn, this would increase the probability of an oblique model emerging when analyzed using exploratory factor analysis. Naturally, when analyzing a true oblique model using orthogonal confirmatory factor analysis, the results will produce a poor fit (as is evidenced in Study 2).

The Revised PANAS as Evidence of a Poorly Constructed NEO-FFI

The PANAS data produced worse findings than the NEO-FFI. Unfortunately, with the models failing to converge, the required statistics to calculate the fit indices could not be recovered. Therefore, no specific empirical evaluation of the PANAS could be preformed. The failure of the PANAS to converge may suggest something intrinsically wrong with NEO-FFI. If the descriptive terms employed by the NEO-FFI were consistent with the construct of personality, one would expect the method by which individuals rated the descriptive terms, as long as the retrospective timeframe was consistent, would not substantially effect the emergence of the five-factor model. This however does not appear to be the case.

Assuming that the retrospective timeframes were consistent between the NEO-FFI and the PANAS[8], only two differences existed. First, the scale differed slightly in the language used to label each integer on the scale. However, both measures do employ a five-point scale. The measures also differed on the presentation of the terms used to describe the "Big Five." It appears that simply removing the terms from the sentence format employed by the NEO-FFI and presenting them in the single item format utilized by the PANAS affected the factor structure to such an extent as to not allow for the convergence of the "Big Five" factor structure.

The NEO-FFI's Poor Fit and Other Personality Measures

Although the poor fit of the NEO-FFI may be due to the relations among the factors (i.e., the factors do not relate to one another in a way that is consistent with the theory), until this option is examined in depth, the current findings raise concerns over the validity of the NEO-FFI. That is, the poor fit suggests that the NEO-FFI may not be measuring what it was designed and intended to measure. Furthermore, the NEO-FFI has been used in the validation of numerous other measures (e.g., California Psychology Inventory, Hogan Personality Inventory, Interpersonal Adjective Scales). This of course also raises concern over the validity of these alternative measures. Whereas, a high correlation between an alternative measure and the NEO-FFI does suggest that the measures are assessing similar constructs, it does not address whether or not they are assessing the correct construct (i.e., the theoretically-defined construct).

The results from the confirmatory factor analyses preformed in the current study would suggest that the NEO-FFI is not assessing what it was intended to assess thereby suggesting that the alternative measures are also failing to assess the construct of interest. This problem could arise from two very different circumstances. First, the NEO-FFI

---

[8] It is not clear whether or not this was the case between the PANAS's *in general* instructional set and the standard NEO-FFI, however it seems likely that both *moment* instructional sets would have produced similar retrospective timeframes.

could be an accurate measure of personality developed by haphazardly following an incorrect theoretical model. This solution seems unlikely. Alternatively, the NEO-FFI may fail to measure the correct theoretical model. This seems more likely. Of course, considering the results from the current factor analyses, those conducted previously in the literature, and the results found in Study 1 regarding the retrospective timeframes utilize by individuals, it may be a combination of a poorly constructed measure based on an inaccurate model.

The Relation Between Affect and Personality

Clearly the poor fit of the models tested in the current study did not allow for a more empirical exploration of the affect and personality relationship. Although problems with the five-factor model existed, these problems were not fatal to the current project. That is to say, knowledge regarding the affect and personality relationship was obtained. For instance, the similarity between the empirically derived standard and moment NEO-FFI's single factor structures suggests the assessment of a similar latent construct. More specifically, given that many of the same items were retained, regardless of the retrospective timeframe employed in the empirically derived models suggests the measurement of a similar if not identical construct.

These results in no way suggest that the latent variables associated with these factors are on-line mood states or aspects of the "Big Five." Furthermore, the poor fit associated with these factors once combined (i.e., the five-factor model) suggests that the relationship between these different single factors are in no way orthogonal (as proposed by current personality theory). However, the retention of many of the same items in both the standard and moment versions of the NEO-FFI does suggest that the constructs are closely related. In fact, any variation in items retained from the single factors may in part be due to random error. Clearly, no definitive conclusions can be reached as to the role of random error in the current analyses without cross-validating the CFA findings in an independent sample.

The current findings further suggest great similarity between affect and personality. Such that, regardless of the retrospective timeframe utilized by individuals, when the same terms and behaviors are rated, similar single factors are empirically derived. Furthermore, when the empirically derived single factors are analyzed together in five-factor models the fit of the models drop substantially. Thereby, suggesting that both the affect and personality five-factor models likely involved extensive item cross loadings.

Assessment of Affect and Personality

Given the similarity in fit and empirically derived models in Study 2, the results of Study 1 seem to suggest the existence of a bigger problem with the assessment of affect and personality than was initially thought. More specifically, provided that the terminology used to assess affect and personality is in many respects identical, the only method by which to discriminate between the two constructs is by defining the experiential timeframe (i.e., retrospective timeframe) associated with the two constructs. Although this appears easy on a theoretical level, Study 1 provides evidence that this is not easily accomplished.

The introduction to the current study discussed in detail the similarities between the descriptive terms used by many measures to assess both affect and personality. The current study went a step further by examining the retrospective timeframes employed with these measures. Undeniably, the variation within the moment and right now retrospective timeframes causes the greatest concern with respect to the assessment of affect. If the theoretical definition of an on-line affective experience is in part defined by the affective experiences occurring at this very instance (i.e., this second), then the current timeframe instructional sets fail to clearly assess the appropriate retrospective timeframe. In fact, if the appropriate on-line timeframe is approximately 1 second (or shorter) then the mean retrospective timeframes elicited by the moment and right now instructional sets are too long to assess on-line mood.

Theoretically, the temporal definition of personality is the majority of an individual's adult lifespan. Therefore, there appears to be less of an issue with regard to the in general instructional set. Although there was great variability associated with the retrospective timeframes reported by individuals in Study 1 when using the in general instructional set, the mean retrospective timeframe encompassed a majority of the individuals' life span. This is not to say that the variation associated with the distribution of retrospective timeframes is not a concern. In fact, many individuals reported relatively short retrospective timeframes (e.g., 1 min.). However, the point at which a retrospective timeframe is considered too short to define personality (i.e., deviates from the individual's actual age) is unclear.

Limitations and Future Directions

Clearly, as with any study, there are limitations. The current exploration is no exception. With regard to the hypotheses put forth in the current study, the poor fit of the measures limited the comparisons that could be preformed. More specifically, the NEO-FFI did not serve as a suitable measure for the comparison of affect and personality. Although initially the use of confirmatory factor analysis was only considered a strength of the current study, the poor fit of the models suggest that further exploratory factor analysis may be a more appropriate starting point.

A logical next step in the exploration of personality may be to explore the structure of personality using oblique rotations. Importantly, this method would still allow for an orthogonal structure to emerge if in fact an orthogonal structure existed. However, the results of the current study suggest that the opposite may in fact be true. That is, personality may consist of an oblique factor structure. To further maximize the likelihood that personality is being assessed as opposed to affect, and to increase the likelihood of an orthogonal factor structure emerging with oblique rotations, the instructional timeframe utilized should be standardized and allow for the greatest retrospective timeframe.

Of course, as was obvious in Study 1, communicating the appropriate instructions (e.g., retrospective timeframe) can be difficult. Questions still exist as to individuals' understanding of the instructions employed in Study1. Therefore, questions must still exist regarding the instructional sets utilized in Study 2. Future research is needed to better define the retrospective timeframes utilized by individuals being assessed.

Likewise, research exploring individuals' understanding of the assessment tools in general may shed light on many of the contradictory findings in the literature.

In fact, the age of the participants may be related to their understanding and/or interpretation of an assessment tool. For example, older individuals maintain a greater number of memories to draw upon when completing a personality measure. This may influence their interpretation of such terms as "sometimes" or "generally." Unfortunately, the current study's sample was comprised primarily of college age individuals.

The understanding of a construct can also be greatly influenced by the method used to measure the construct. Although the current study was primarily interested in self-report methods, there are alternative methods of assessing affect and personality (e.g., behavioral, case study, physiological changes, etc.). The incorporation of various methods of assessment would provide a superior understanding of affect and personality. Moreover, multiple measures may provide new insights as to the relationship between affect and personality.

REFERENCES

Allport, F. H., & Allport, G. W. (1921). Personality traits: Their classification and measurement. Journal of Abnormal and Social Psychology, 16, 1 – 40.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. Psychological Monographs, 47, (Whole No. 21).

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4^{th} ed.). Washington, DC: APA.

Bandura, A. (1999). Social cognitive theory: An agentic perspective. Asian Journal of Social Psychology, 2, 21 – 41.

Beck, A. T., & Steer, R. A., (1987). BDI: Beck depression inventory manual. USA: Harcourt Brace.

Bentler, P., & Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588 – 606.

Bollen, K. (1989). A new incremental fit index for general structural equation models. Sociological Methods and Research, 17, 303 – 316.

Borgotta, E. F. (1964). The structure of personality. Behavioral Science, 12, 9 – 17.

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. Personality and Individual Differences, 11, 515 – 524.

Cannon, W. B. (1929). Bodily changes in pain, hunger, fear, and rage. New York: Branford.

Carroll, J.M., & Russell, J.A. (1996). Do facial expressions signal specific emotions? Judging the face in context. Journal of Personality and Social Psychology, 70, 205-218.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. Psychological Review, 97, 19 – 35.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. Journal of Abnormal and Social Psychology, 38, 476 – 506.

Cattell, R. B. (1957). Personality and motivation structure and measurement. New York: World Book.

Cattell, R. B. (1965). The scientific analysis of personality. London: Penguin.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. Journal of Personality and Social Psychology, 66, 93 – 114.

Cohen, Jacob (1990). Things I have learned (so far). American Psychologist, 45, 1304 – 1312.

Cohen, Jacob (1992). A power primer. Psychological Bulletin, 112, 155 – 159.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. Journal of Consulting and Clinical Psychology, 56, 754 – 761.

Costa, P. T., Jr., & McCrea, R. R. (1980). Influence of extroversion and neuroticism on subjective well-being: Happy and unhappy people. Journal of Personality and Social Psychology, 38, 668 – 678.

Costa, P. T., Jr., & McCrea, R. R. (1985). The NEO Personality Inventory manual. Odessa, Fla.: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrea, R. R. (1989). The NEO PI/FFI manual supplement. Odessa, Fla.: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrea, R. R. (1992). The NEO-PI-R: Professional manual. Odessa, Fla.: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrea, R. R. (1995). Primary traits of Eysenck's P-E-N system: Three- and five- factor solutions. Journal of Personality and Social Psychology, 69, 308 – 317.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. Journal of Personality and Social Psychology, 47, 1105 – 1117.

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. Journal of Personality and Social Psychology, 69, 130 – 141.

Digman, J. M. (1990). Personality structure: Emergence of the five factor model. Annual Review of Psychology, 41, 417 – 440.

Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A and C are more reliable than O and E. Personality and Individual Differences, 29, 907 – 920.

Ekman, P., & Friesen, W. V. (1978). Facial Affect Coding System. San Francisco, Ca.: University of California, San Francisco.

Erikson, E. (1950). Childhood and society. New York: Norton.

Eysenck, H. J. (1967). The biological basis of personality. Springfield, IL: Charles C. Thomas.

Eysenck, H. J. (1970). National differences in personality as related to ABO blood groups polymorphism. Psychological Reports, 41, 1257 – 1258.

Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Eds.), Handbook of personality: Theory and research. (pp.244 – 276). New York: Guilford Press.

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck personality questionnaire. San Diego, CA: Hodder & Staughton/DIGITS.

Eysenck, H. J., & Fulker, D. (1983). The components of Type A behavior and its genetic determinants. Personality and Individual Differences, 4, 499 – 505.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specifications on structural equation modeling fit indexes. Structural Equation Modeling, 6, 56 – 83.

Ferguson, E., & Patterson, F. (1998). The five factor model of personality: Openness a distinct but related construct. Personality and Individual Differences, 24, 789 – 796.

Fiske, D. W. (1949). Consistency of the factorial structure of personality ratings from different sources. Journal of Abnormal and Social Psychology, 44, 329 – 344.

Freud, S. (1933). New introductory lectures on psychoanalysis. New York: Norton.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Eds.), Review of personality and social psychology (pp. 141 – 165). Beverly Hills, Calif.: Sage.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26 – 42.

Gray, J. A., (1970). The psychophysiological basis of introversion-extroversion. Behavior Research and Tharapy, 8, 249 – 266.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. Journal of Personality and Social Psychology, 64, 1029 – 1041.

Greenburg, M. A., & Stone, A. A. (1992). Emotional disclosure about traumas and its relation to health: Effects of previous disclosure and trauma severity. Journal of Personality and Social Psychology, 63, 75 – 84.

Gross, J. J. (1998). Antecedent- and response-focused emotional regulation: Divergent consequences for experience, expression, and physiology. Journal of Personality and Social Psychology, 74, 224 – 237.

Guilford, J. P. (1975). Factors and factors of personality. Psychological Bulletin, 82, 802 – 814.

Hogan, R. (1986). Hogan personality inventory. Minneapolis, MN: National Computer Systems.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structural modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3, 424 – 453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structural analysis. Structural Equation Modeling, 6, 1 – 55.

Hull, C. L. (1943). Principles of behavior. New York: Appleton.

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. Journal of Personality and Social Psychology, 61, 932 – 945.

Hunsley, J. (1990). Dimensionality of the Multiple Affect Adjective Check List-Revised: A comparison of factor analytic procedures. Journal of Psychopathology and Behavioral Assessment, 12, 81 – 91.

John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. Buss & N. Cantor (Eds.), Personality Psychology: Recent Trends and Emerging Directions. New York: Springer-Verlag.

Jung, C. G. (1939). The integration of the personality. New York: Farrar & Rinehart.

Kelly, G. A. (1955). The psychology of personal constructs. New York: Norton.

Klages, L. (1927) Persoenlichkeit, Einfuehrung in die Charakterkunde. Washington, US: Unknown.

Klinger, E. (1987). Current concerns and disengagement from incentives. In F. Halisch & J. Kuhl (Eds.), Motivation, intention, and volition (pp. 95 – 112). Berlin: Springer-Verlag.

Kuhl, J., & Helle, P. (1986). Motivation and volitional determinants of depression: The degeneration-intention hypothesis. Journal of Abnormal Psychology, 95, 247 – 251.

Larson, R. L., & Diener, E. (1985). A multitrait-multimethod examination of affect structure: Hedonic level and emotional intensity. Personality and Individual Differences, 6, 631 – 636.

Lecci, L., Okun, M. A., & Karoly, P. (1994). Life regrets and current goals as predictors of psychological adjustment. Journal of Personality and Social Psychology, 66, 731 – 741.

Lorr, M. (1986). Interpersonal style inventory: Manual. Los Angeles: Western Psychological Services.

Marsh, H., Balla, J., & MacDonald, R. (1988). Goodness-of fit in confirmatory factor analysis: The effects of sample size. Psychological Bulletin, 103, 391 – 410.

Martin, M. (1990). On the induction of mood. Clinical Psychology Review, 10, 669 – 697.

Masters, A. B. (1967). The distribution of blood groups in psychiatric illness. British Journal of Psychiatry, 113, 1309 – 1315.

McCrae, R. R., & Costa, P. T. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. Multivariate Behavioral Research, 24, 107 – 124.

McCrae, R., Zonderman, A., Costa, P., Bond, M., & Paunonen, S. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. Journal of Personality and Social Psychology, 70, 552 – 566.

Meyer, G. J., & Shack, J. R. (1989). Structural convergence of mood and personality: Evidence for old and new directions. Journal of Personality and Social Psychology, 57, 691 – 706.

Mischel, W. (1968). Personality and assessment. New York: Wiley.

Mooradian, T. A., & Nezlek, J. B. (1996). Comparing the NEO-FFI and the Saucier's Mini-Markers as measures of the big five. Personality and Individual Differences, 21, 213 – 215.

Mulaik, S., James, L., Van Alstine, J., Bennett, J., Lind, S., & Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430 – 445.

Norman, W. T., (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574 – 583.

Nunnally, J. C., & Berstein, I. H. (1994). Psychometric Theory (3th ed.). New York, NY: McGraw-Hill Inc.

Pervin, L. A., & Oliver, P. J. (Eds.). (1997). Personality: Theory and research (7th ed.). New York, NY: Wiley.

Reise, S. P., Waller, N. G., & Comery, A. L. (2000). Factor analysis and scale revision. Psychological Assessment, 12, 287 – 297.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality of Social Psychology, 39, 1161 – 1178.

Russell, J. A. & Carroll, J. M. (1999a). The phoenix of bipolarity: Reply to Watson and Tellegen (1999). Psychological Bulletin, 125, 611 – 617.

Russell, J. A. & Carroll, J. M. (1999b). On the bipolarity of positive and negative affect. Psychological Bulletin, 125, 3 – 30.

Russell, J. A., Wiess, A., & Mendelson, G. A. (1980). Affect Grid: A scale of pleasure and arousal. Journal of Personality and Social Psychology, 57, 493 – 502.

Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks. Psychological Bulletin, 124, 165 – 196.

Rusting, C. L. (1999). Interactive effects of personality and mood emotion-congruent memory and judgment. Journal of Personality and Social Psychology, 77, 1073 – 1086.

Schlosberg, H. (1954). Three dimensions of emotion. Psychological Review, 61, 81 – 88.

Schonemann, P. H. (1968). On two-sided orthogonal Procrustes problem. Psychometrika, 33, 19 – 34.

Selye, H. (1974). Stress without distress. Philadelphia: Lippincott.

Shepherd, J. (1966). In God we trust: All others pay cash. New York: Doubleday.

Skinner, B. F. (1948). Walden two. New York: Macmillan.

Smith, G. M. (1967). Usefulness of peer ratings of personality in educational research. Journal of Education and Personality Measurement, 27, 967 – 984.

Thomas, D., & Diener, E. (1990). Memory accuracy in the recall of emotions. Journal of Personality and Social Psychology, 59, 291 – 297.

Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1 – 10.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on the trait ratings. USAF ASD Technology Report. 61 – 97.

Warr, P., Berter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. Journal of Personality and Social Psychology, 55, 128 – 141.

Watson, D. (1988). The vicisstude of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. Journal of Personality and Social Psychology, 55, 128 – 141.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063 – 1070.

Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. Psychological Bulletin, 98, 219 – 235.

Watson, D. & Tellegen, A. (1999). Issues in the dimensional structure of affect – effects of descriptors, measurement error, and response formats: Comments on Russell and Carroll (1999). Psychological Bulletin, 125, 601 – 610.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. Journal of Experimental Psychology, 3, 1 –14.

Yadama, G. N., & Pandey, S. (1995). Effects of sample size on goodness-of-fit indices in structural equation models. Journal of Social Service Research, 20, 49 – 70.

Zuckerman, M. & Lubin, B. (1965). Multiple Affect Adjective Check List. San Diego, Ca.: Edits.

APPENDIX A

Formulas for Fit Indices

BBI/NFI = $(\chi^2_{null} - \chi^2_{model}) / \chi^2_{null}$

CFI = $1 - (\chi^2_{model} - df_{model}) / (\chi^2_{null} - df_{null})$

Delta2 = $(\chi^2_{null} - \chi^2_{model}) / (\chi^2_{null} - df_{model})$

RMSCA = $\sqrt{(}\chi^2_{model} - df_{model}) / (df_{model} \text{ x } N-1)$

TLI = $\{(\chi^2_{null} / df_{null}) - (\chi^2_{model} / df_{model})\} / \{(\chi^2_{model} - \chi^2_{null}) - 1\}$

Biographical Sketch

R.J. Wirth was born in Kingston, New York, on September 21, 1973. In 1999, he graduated Magna Cum Laude with Honors in Psychology from the University of North Carolina at Wilmington with a B.A. in psychology. That same year, he entered the graduate program in psychology at the University of North Carolina at Wilmington, where he worked under the direction of Dr. Len Lecci. During August of 2001 he will begin doctoral work, under the direction of Dr. Patrick Curran, at the University of North Carolina at Chapel Hill where he intends to pursue his interest in quantitative psychology and personality. He is a member of the American Psychological Society and the Society for Personality and Social Psychology and has presented his research at American Psychological Society, American Psychological Association, and South Eastern Psychological Society annual meetings. He has also published in the Journal of Behavioral Therapy and Experimental Psychiatry.