

Development and Validation of the Temperament and Affectivity Inventory (TAI)

Assessment
2015, Vol. 22(5) 540–560
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191114557943
asm.sagepub.com



David Watson¹, Sara M. Stasik², Michael Chmielewski³,
and Kristin Naragon-Gainey⁴

Abstract

Trait affect scales have been a mainstay of the assessment literature for more than 50 years. These scales have demonstrated impressive construct validity, including substantial relations with personality, satisfaction, and psychopathology. However, the accumulating evidence has exposed several limitations, including (a) problems associated with retrospective biases, (b) lower temporal stability because of enhanced susceptibility to transient error, and (c) reduced self–other agreement. These limitations motivated the creation of the Temperament and Affectivity Inventory (TAI), which uses a traditional personality format (i.e., full sentences rather than single words or short phrases). The 12 TAI scales were created based on factor analyses in two samples and validated in four additional samples. The scales are internally consistent, highly stable over time, and show strong convergent, discriminant, and incremental validity in relation to self-report and interview-based measures of personality and psychopathology. Thus, the TAI provides a promising new approach to assessing trait affectivity.

Keywords

factor analysis, scale construction, personality, temperament, affectivity

Multi-affect mood questionnaires emerged in the 1950s, based on the pioneering factor-analytic work of Vincent Nowlis and Russel Green (for summaries of this seminal research, see Nowlis, 1965, 1970). Nowlis and Green initially created a large pool of 130 mood terms. Extensive factor analyses of these terms led to the creation of the Mood Adjective Checklist (MACL), which included 12 specific content scales (e.g., Aggression, Anxiety, Skepticism, Vigor). It is noteworthy, moreover, that their comprehensive pool of mood terms was the starting point for many later structural analyses and scale development projects in this domain (e.g., McNair, Lorr, & Droppleman, 1971; Stone, 1987; Thayer, 1967, 1978; Zuckerman, 1960).

Researchers quickly learned that standard mood questionnaires—which typically involve responding to a list of single words or short phrases—could be used with a number of different instructions to assess affect over varying time intervals (e.g., right now, today, over the past few weeks). A particularly important development occurred when Marvin Zuckerman published the Affect Adjective Check List (AACL) in 1960. The AACL was the first instrument to include parallel state and trait versions. In the state format, respondents were asked to describe “how you feel now—today,” whereas in the trait form they were asked to

rate “how you generally feel.” Zuckerman’s (1960) innovation proved to be quite popular: Several subsequent affect measures—including the Multiple Affect Adjective Checklist (MAACL; Zuckerman & Lubin, 1965), the Multiple Affect Adjective Checklist–Revised (MAACL-R; Zuckerman & Lubin, 1985), the Differential Emotions Scale IV (DES-IV; Izard, Libero, Putnam, & Haynes, 1993), the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), and the Expanded Form of the Positive and Negative Affect Schedule (PANAS-X; Watson & Clark, 1999)—contain alternative versions that permit one to assess either (a) short-term fluctuations in current mood or (b) long-term individual differences in trait affectivity.

¹University of Notre Dame, Notre Dame, IN, USA

²Knox College, Galesburg, IL, USA

³Southern Methodist University, Dallas, TX, USA

⁴University at Buffalo, State University of New York, Buffalo, NY, USA

Corresponding Author:

David Watson, Department of Psychology, University of Notre Dame,
118 Haggart Hall, Notre Dame, IN 46556, USA.
Email: db.watson@nd.edu

The Validity of Trait Affect Measures

Evidence Supporting Validity

These trait affect measures have been a mainstay of the assessment literature for more than 50 years. During this time, an impressive body of data has accrued to support their construct validity (see Watson & Tellegen, 1999; Watson & Vaidya, 2013). For example, these ratings show strong convergence with general traits of personality (e.g., Naragon-Gainey, Watson, & Markon, 2009; Watson & Clark, 1999). In a sample of 4,457 respondents, Watson, Wiese, Vaidya, and Tellegen (1999) reported that the PANAS-X General Negative Affect scale correlated .58 with Neuroticism, whereas the General Positive Affect scale correlated .51 with Extraversion (see also Beer, Watson, & McDade-Montez, 2013). Trait versions of other PANAS-X scales correlate strongly with measures of agreeableness and conscientiousness (Watson & Clark, 1992); indeed, among the Big Five personality traits, only openness does not display strong links to affectivity.

Moreover, trait affect ratings display substantial links to indicators of happiness, subjective well-being, and satisfaction across various life domains. For example, the general affect scales of the PANAS-X show moderate to strong associations with life satisfaction (Chang & Sanna, 2001; Heller, Judge, & Watson, 2002), marital/relationship satisfaction (Watson et al., 2004; Watson, Hubbard, & Wiese, 2000a), and job satisfaction (Connolly & Viswesvaran, 2000; Heller et al., 2002).

Trait affect ratings also show substantial links to psychopathology, particularly with symptoms and diagnoses of internalizing disorders. For instance, Watson and Walker (1996) reported that the general affect scales of the PANAS-X were significantly related to depression and anxiety symptoms that were assessed approximately 6 to 7 years later. Watson, Clark, and Stasik (2011) found that the positive and negative affect scales of the PANAS-X—including both the general affect scales and specific scales such as Fear, Sadness, and Joviality—were significantly related to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (*DSM-IV-TR*; American Psychiatric Association, 2000) diagnoses of major depression, generalized anxiety disorder (GAD), posttraumatic stress disorder (PTSD), panic disorder, social phobia, and obsessive-compulsive disorder (OCD).

Finally, trait affect measures converge well with aggregated state ratings of mood. Watson and Tellegen (1999) reported correlations between the general affect scales of the PANAS-X and aggregated scores in three different samples in which participants rated their mood “today” ($N = 455$; total observations = 20,367), “during the past few days” ($N = 61$; total observations = 2,567), or “during the past week” ($N = 195$; total observations = 2,483). The

General Negative Affect scale had correlations ranging from .53 to .65 with mean state negative affect; in parallel fashion, the PANAS-X General Positive Affect scale had correlations ranging from .61 to .64 with mean state positive affect.

Problems With Trait Affect Scales

Retrospective Biases. Despite this impressive evidence, it now is clear that trait affect measures also suffer from a variety of problems that lessen their construct validity. Many of these problems arise from the retrospective nature of these global ratings, which require respondents (a) to recall their past experiences and then (b) draw inferences from them. This process is subject to at least three significant problems. First, Fredrickson and Kahneman (1993) demonstrated that global ratings suffer from *duration neglect*, that is, from an insensitivity to the actual amount of time that an affect was experienced (see also Kahneman, 1999; Russell & Carroll, 1999). Second, several studies have shown that general affect ratings are influenced by *transient error*, that is, the respondents' mood at the time of assessment (e.g., Schwarz & Clore, 1983; Schwarz & Strack, 1999; Stone, Shiffman, & DeVries, 1999). Third, retrospective ratings are subject to *recency effects*, such that more recent experiences have a greater influence than more distant ones (Schwarz & Sudman, 1994; Stone et al., 1999).

Reduced Dependability/Temporal Stability. Other evidence indicates that scales assessing the Big Five are significantly more stable than measures of trait affectivity (Vaidya, Gray, Haig, Mroczek, & Watson, 2008; Vaidya, Gray, Haig, & Watson, 2002; Watson & Humrichouse, 2006), despite theoretical similarities and strong empirical correlations between these constructs (Vaidya et al., 2002; Watson et al., 1999). Using data from the Iowa Longitudinal Personality Project, Vaidya et al. (2002, 2008) found that the Big Five Inventory scales (BFI; John & Srivastava, 1999) were significantly more stable than comparable PANAS-X scores over a 2.5-year period. The fact that BFI Neuroticism was significantly more stable than the PANAS-X negative affectivity scales was especially surprising given that these scales are strongly correlated (e.g., Vaidya et al., 2008, reported a median correlation of .59 between BFI Neuroticism and the General Negative Affect Scale) and all are highly affective in nature (Beer et al., 2013; Pytlik Zillig, Hemenover, & Dienstbier, 2002). Moreover, these differences in stability could not be fully explained by content differences, the differential impact of life experiences, or different developmental trajectories in the stability of personality and trait affect (Vaidya et al., 2002, 2008; Watson, 2004).

Furthermore, the lower long-term temporal stability of trait affect ratings is due—at least in part—to the fact that

they are more susceptible to transient error than other types of personality scales (Watson & Vaidya, 2013). Chmielewski and Watson (2009) demonstrated this effect directly by comparing short-term dependability correlations for PANAS-X and Big Five scores obtained over (a) 2-week and (b) 2-month retest intervals. Dependability can be defined “as the correlation between two administrations of the same test *when the lapse of time is insufficient for people themselves to change* with respect to what is being measured” (Cattell, Eber, & Tatsuoka, 1970, p. 30). Thus, in contrast to longer term stability correlations—which can be influenced by both measurement error and by true change—dependability coefficients provide a highly sensitive index of the effects of transient error (Gnambs, 2014). Chmielewski and Watson (2009) found that in all 17 comparisons, dependability coefficients for neuroticism were significantly higher than those for the PANAS-X negative affectivity scales (see their table 5).

Lower Self–Other Agreement. Finally, trait affect scores show substantially poorer self–other agreement than other types of personality scales, even strongly related measures with very similar content (Beer et al., 2013; Watson, Hubbard, & Wiese, 2000b). Beer et al. (2013) reported particularly striking evidence along these lines. They began by replicating the familiar finding that Neuroticism was strongly correlated with the PANAS-X General Negative Affect scale in both self-ratings (mean $r = .59$) and other-ratings (mean $r = .69$) across six dyadic samples (overall $N = 1,852$). Nevertheless, neuroticism scales showed substantially better self–other convergence (mean agreement $r = .51$) than did General Negative Affect (mean $r = .28$) in these samples. Furthermore, this same pattern emerged in a seventh sample of married couples ($N = 381$): Once again, two different measures of neuroticism showed significantly better self–other agreement ($r_s = .62$ and $.50$) than did General Negative Affect ($r = .34$).

The Temperament and Emotion Questionnaire

Development of the Temperament and Emotion Questionnaire

Consequently, we now have extensive data indicating that trait affect scales show significantly poorer psychometric properties than highly correlated measures with very similar item content (e.g., indicators of neuroticism). Given that these differences cannot simply be attributed to content, it seems likely that the format and presentation of these trait affect measures (such as the nature of the instructions given to respondents) is problematic in some way.

To investigate the effects of format and instructions, Watson (2004) created the TEQ (see also Chmielewski &

Watson, 2009). The TEQ consists of statements that are answered using a 5-point agree–disagree format (*strongly disagree, disagree, neutral or cannot decide, agree, strongly agree*). Watson (2004) created the TEQ items by taking individual PANAS-X terms and embedding them in complete sentences. For example, *cheerful* became “I am a cheerful person,” *sad* became “I often feel a bit sad,” and *active* became “I lead a full and active life.” This process ensured that the original content of the PANAS-X basically remained unaltered, although some minor differences inevitably occurred in the transition from simple descriptors to full sentences. The TEQ contains the same scales as the PANAS-X, except that it does not include a measure of Surprise (which shows poor validity as a dispositional construct; see Watson & Vaidya, 2013).

Chmielewski and Watson (2009) compared the dependability correlations for parallel TEQ and PANAS-X scales in the 2-week and 2-month retest data discussed earlier (results for the Shyness, Fatigue, and Serenity scales were only available in the 2-week sample, however). It is noteworthy that the TEQ negative affect scales yielded significantly higher dependability coefficients in 6 of 10 comparisons (3 in each time interval); the TEQ Shyness and Fatigue scales also showed superior dependability in the 2-week data. In contrast, the positive affect scales did not yield any significant effects across instruments. Overall, the TEQ scales showed significantly greater dependability in 8 of 21 comparisons (38.1%). These results provided the first systematic demonstration that retest correlations can be significantly influenced by simple changes in wording and format, even while maintaining the same basic item content.

In a related vein, Beer et al. (2013) showed that the TEQ General Negative Affect scale showed better self-spouse agreement ($r = .46$) than did its PANAS-X counterpart ($r = .34$; $z = 2.24$, $p < .05$). Paralleling the dependability data, however, the agreement correlations for the General Positive Affect scales of the TEQ ($r = .43$) and PANAS-X ($r = .36$) did not differ significantly from one another ($z = 1.33$, *ns*). These results therefore provide mixed support for the idea that a sentence-based approach yields a more reliable and valid assessment of trait affectivity.

Limitations of the Temperament and Emotion Questionnaire

Findings obtained with the TEQ have been promising and provide support for the argument that the format of traditional trait affect measures is problematic. Nevertheless, the TEQ itself is limited in three important ways. First, its construction was purely rational in nature; that is, it was created simply by modifying existing PANAS-X items, rather than through rigorous structural analyses. Second, as we have noted, whereas the TEQ negative affect scales display psychometric properties that are superior to their PANAS-X

counterparts, the positive affect scales do not (Beer et al., 2013; Chmielewski & Watson, 2009). Third, although the TEQ negative affect scales outperform those on the PANAS-X, they still compare unfavorably with standard neuroticism scales. For example, Chmielewski and Watson (2009) reported that the dependability coefficients for neuroticism were significantly higher than those for the TEQ negative affectivity scales in 14 of 17 comparisons. Similarly, Beer et al. (2013) found that neuroticism scales showed significantly better self-spouse agreement than did TEQ General Negative Affect in one of two comparisons.

The Current Research: Development of the Temperament and Affectivity Inventory

In light of these limitations, we decided to develop a comprehensive new trait affect instrument—the Temperament and Affectivity Inventory (TAI)—from scratch, using a standard personality format (i.e., full sentences) rather than single words or phrases. In developing the initial TAI item pool, we included multiple markers for each of the targeted trait dimensions, rationally organizing them into *homogeneous item composites* (HICs; Hogan, 1983; Watson et al., 2007). The results of structural analyses inevitably reflect the set of indicators that are included in them (the old “garbage in, garbage out” problem). The creation of HICs ensures that all targeted constructs have a reasonable chance to emerge in subsequent factor analyses (e.g., the inclusion of multiple anger items makes it possible for an Anger factor to emerge). Note, however, that the construction of these HICs does not force a corresponding factor to emerge; indeed, as we will see, many of our HICs failed to define distinctive dimensions and, therefore, were not used to create scales.

We developed a total of 16 HICs (235 items overall), drawing on the content included in prominent multi-affect mood questionnaires such as the PANAS-X, MACL, MAACL-R, DES-IV, and Profile of Mood States 2 (POMS 2; Heuchert & McNair, 2012). Nine HICs targeted various negative affects. Anxiety (21 items; e.g., “I am more anxious than most people”) contained content similar to that included in the PANAS-X Fear, MACL Anxiety, MAACL-R Anxiety, POMS-2 Tension-Anxiety, and DES-IV Fear scales. Depression (19 items; e.g., “I often feel sad”) targeted content included in the PANAS-X Sadness, MACL Sadness, MAACL-R Depression, DES-IV Sadness, and POMS-2 Depression-Dejection scales. Anger (19 items; e.g., “I can be a bit grouchy at times”) tapped content related to the PANAS-X Hostility, MACL Aggression, MAACL-R Hostility, DES-IV Anger, and POMS-2 Anger-Hostility scales. Shame/Guilt (15 items; e.g., “I’ve made a lot of mistakes in my life”) targeted content related to the PANAS-X Guilt, DES-IV Guilt, DES-IV Shame, and DES-IV

Inner-Directed Hostility scales. Shyness (12 items; e.g., “It is difficult for me to talk to others”) assessed content corresponding to the PANAS-X Shyness and DES-IV Shyness scales. Lassitude (12 items; e.g., “I’m tired much of the time”) captured content related to the PANAS-X Fatigue, MACL Fatigue, and POMS-2 Fatigue-Inertia scales. Mistrust (13 items; e.g., “I am skeptical of others”) included content similar to the MACL Skepticism scale. Finally, we created two additional HICs on an exploratory basis to see if they might define distinct scales: Self-Doubt (14 items; e.g., “I wish I had more self-confidence”) and Frustration (8 items; e.g., “I have a low tolerance for frustration”).

The seven remaining HICs assessed various positive affects. Joy (14 items; e.g., “Little things can make me very happy”) targeted content included in the PANAS-X Joviality, MACL Elation, MAACL-R Positive Affect, and DES-IV Enjoyment scales. Energy (13 items; e.g., “I am always on the go”) tapped content related to the PANAS-X Joviality, MACL Vigor, MAACL-R Sensation Seeking, and POMS-2 Vigor scales. Interest/Excitement (17 items; e.g., “I lead an interesting life”) assessed content related to the PANAS-X General Positive Affect and DES-IV Interest scales. Experience Seeking (15 items; e.g., “People would describe me as daring”) captured content contained in the PANAS-X Self-Assurance and MAACL-R Sensation Seeking scales. Attentiveness (14 items; e.g., “It is easy for me to concentrate for long periods of time”) tapped content similar to the PANAS-X Attentiveness and MACL Concentration scales. Friendliness (15 items; e.g., “People would describe me as warmhearted”) assessed content related to the MACL Social Affection, MAACL-R Positive Affect, and POMS-2 Friendliness scales. Finally, Serenity (14 items; e.g., “I have a relaxed approach to things”) included content corresponding to the PANAS-X Serenity and MACL Nonchalance scales.

Phase I: Scale Development

Method

Participants and Procedure

Sample 1. The participants were 562 undergraduate students enrolled in an introductory psychology course at the University of Iowa. They completed an online battery of self-report questionnaires in partial fulfillment of a course research exposure requirement. The sample (mean age = 19.0 years) consisted of 187 men and 375 women and was predominantly White (92.5%).

Sample 2. The participants were 374 community-dwelling adults living in Eastern Iowa who completed a battery of questionnaires online. Respondents were compensated \$15 for their participation. The sample (age range = 21–74 years, mean age = 36.7 years) consisted of 97 men and 276

women (the gender of one participant was unspecified) and was predominantly White (94.1%).

Measures. All participants completed the initial pool of 235 items that was described earlier. They responded to each item using a 5-point scale (*strongly disagree, disagree, neutral or cannot decide, agree, strongly agree*). The participants also completed additional measures for validation purposes; these will be described in Phase 2.

Results and Discussion

Basic Analytic Approach. We conducted a separate series of noniterated principal factor analyses (using squared multiple correlations as the initial communality estimates) in each sample. All factors were rotated using both varimax (which constrains them to be orthogonal) and promax (which allows them to be correlated; power = 3 in these analyses). In selecting dimensions as targets for scale development, we were guided by two basic principles. First, our goal was to identify the maximum number of factors (and scales) that were psychologically meaningful in order to differentiate specific types of affect as precisely as possible. Second, we only were interested in dimensions that were robust and generalizable; consequently, our final scales reflect factors that emerged clearly in both data sets. Specifically, to be retained for scale development, a dimension had to have at least five consistent markers across both samples; a marker was defined as a variable that (a) loaded at least $|\cdot40|$ on the target factor and (b) had loadings below $|\cdot30|$ on all other factors.

We selected items for the final scales based on two key considerations. First, we retained items that were the purest factor markers (i.e., had high loadings on that dimension and very low loadings on all other factors) and, therefore, maximized the discriminant validity of the scale vis-à-vis the others. Second, we minimized redundancy as much as possible and retained maximally distinct and informative items.

Scale Development Analyses

Preliminary analyses. We began by checking for highly redundant items that potentially could distort the subsequent structural analyses. We identified nine redundant item pairs that correlated $>|\cdot70|$ in both samples. We eliminated one item from each pair, thereby yielding a reduced pool of 226 items.

Structural analyses: Round 1. We then conducted principal factor analyses of the remaining items in each sample; these analyses revealed eight clear, replicable dimensions. Two of these factors were large, nonspecific dimensions representing Negative Affect (defined primarily by items from the Anxiety, Depression, Shame/Guilt, and Self-Doubt HICs)

and Positive Affect (marked primarily by items from the Joy, Energy, Interest/Excitement, and Friendliness HICs).

The remaining six factors defined specific content dimensions and were used to create corresponding scales. Two scales—Attentiveness (six items) and Mistrust (nine items)—consist solely of items from the corresponding HICs. Anger (12 items) contains 9 items from the Anger HIC and 3 items from the Frustration HIC (frustration did not define a distinct affect in our data). Lassitude (seven items) includes six items from the corresponding HIC, plus an Energy item (“I sometimes wish I had more energy”). Shyness (nine items) contains seven items from the Shyness HIC, plus single items from Shame/Guilt (“I worry about looking foolish in public”) and Sociability (“I’m a bit of a loner”). Finally, Experience Seeking includes eight items from the corresponding HIC, plus an item from Anxiety (“I don’t scare easily”).

Structural analyses: Round 2. Next, we conducted a series of analyses to see if we could identify replicable content dimensions within the large Negative Affect factor that emerged in Round 1. Specifically, we subjected the 45 top markers of this factor to separate principal factor analyses in each sample. Four clear, replicable factors emerged and were used to create corresponding scales. All four scales contained items from a single HIC: Anxiety (eight items), Depression (seven items), Self-Doubt (six items), and Regret (five items, all from the Shame/Guilt HIC).

Structural analyses: Round 3. Finally, we conducted parallel analyses of the 38 best markers of the large Positive Affect factor that emerged in Round 1. These analyses revealed two replicable content dimensions that were used to create corresponding scales. Geniality essentially represents a fusion of Joy (four items) and Sociability (five items), plus a single item from Interest/Excitement (“The world is a fascinating place”). Vigor (five items) includes three items from Energy, plus single items from Interest/Excitement (“I lead a full and active life”) and Experience Seeking (“My life is fast-paced”).

The final version of the TAI includes 12 scales and a total of 93 items. It is noteworthy that the scales contain relatively few reverse-keyed items (10 overall, or 10.8% of the total). Specifically, Anger (e.g., “I am a very patient person”) and Mistrust (e.g., “Most people can be trusted”) each have three reverse-keyed items: Attentiveness contains two (e.g., “I have difficulty concentrating”); and Lassitude (“I wake up feeling rested and refreshed”) and Depression (“I rarely feel sad”) both include one. The composition of the TAI scales is consistent with the broader affect literature, as the major multi-affect mood questionnaires all contain very few reverse-keyed items. In fact, none of the PANAS-X (Watson & Clark, 1992) and DES-IV (Izard et al., 1993) scales contains any reverse-keyed items. Similarly, only

four of the 70 MAACL-R scale items (5.7%) are reverse-keyed (Zuckerman & Lubin, 1985). Thus, the largely unipolar nature of the TAI scales reflects the basic underlying structure of affective experience.

Phase 2: Scale Validation

Method

Participants and Procedure

Sample 3. In addition to the two samples described in Phase 1, we recruited four new samples for scale validation purposes. Sample 3 participants were 219 outpatients living in the greater South Bend metropolitan area who were recruited from the Oaklawn Psychiatric Center, and from listservs, newsletters, and mass e-mails sent to University of Notre Dame staff, faculty, and graduate students. All participants were run in small group sessions at the research facility of the Center for Advanced Measurement of Personality and Psychopathology (CAMPP). Patients were compensated \$40 for their participation. The sample (age range = 18-68 years, mean age = 43.2 years) consisted of 90 men and 129 women; it was 63.5% White, 28.8% African American, and 7.7% multiracial or other.

Sample 4. The participants were 341 undergraduate students enrolled in various psychology courses at the University of Notre Dame. They completed an online battery of self-report questionnaires in partial fulfillment of a course research exposure requirement or for extra course credit. The sample (mean age = 19.6 years) consisted of 146 men and 195 women; it was 80.6% White, 7.3% Asian, 3.5% African American, and 8.6% multiracial or other.

Sample 5. The participants were 495 undergraduate students enrolled in various psychology courses at the University of Notre Dame. They completed an online battery of self-report questionnaires in partial fulfillment of a course research exposure requirement or for extra course credit. The sample (mean age = 19.1 years) consisted of 199 men and 296 women; it was 81.8% White, 7.7% Asian, 2.4% African American, and 8.1% multiracial or other.

In addition, 392 of these participants (79.2%) completed the same set of questionnaires at a second session approximately 3 weeks later (mean interval = 19.7 days). This subsample will be used to assess the dependability of various trait measures, including the TAI.

Sample 6. The participants were 438 adults drawn from the greater South Bend metropolitan area. Individuals who had provided their contact information from previous studies conducted at CAMPP were recruited first; other community members who inquired about the study and met the participation criteria (age 18 years or older, comfortable

reading and writing in English) were also eligible to participate. The sample (age range = 18-77 years, mean age = 45.0 years) consisted of 138 men and 297 women (three participants did not specify their gender); it was 47.5% African American, 44.3% White, and 8.2% multiracial or other.

Participants were seen in two 3-hour sessions conducted at CAMPP; they were paid \$60 for each session. Session 1 consisted of an extensive battery of trait measures (including the TAI), plus a clinical interview. Session 2—which was completed approximately 3 weeks later (mean interval = 20.3 days)—consisted of a lengthy battery of psychopathology measures, plus an additional clinical interview. The large majority of the participants ($N = 410$, or 93.6%) completed this second session.

It should be noted that participants from previous CAMPP studies primarily were outpatients who were recruited from various sources, such as the Oaklawn Psychiatric Center.¹ Consequently, although not fully clinical in nature, this sample is characterized by a relatively high level of psychopathology. In fact, nearly half of the sample ($N = 207$, 47.3%) answered “yes” to one or more of these three questions: “Are you currently receiving psychological counseling/therapy for mental health issues?” “Have you received psychological counseling/therapy for mental health issues in the past?” “Are you currently taking medications to treat a mental illness?”

Trait Affectivity Measures

Temperament and Affectivity Inventory. All participants completed the final, 93-item version of the TAI.

Positive and Negative Affect Schedule—Expanded Form. Participants in Samples 1, 2, and 5 completed the trait version of the PANAS-X (Watson & Clark, 1999). The PANAS-X subsumes the original PANAS and therefore contains scales assessing General Negative Affect (10 items) and General Positive Affect (10 items). It also includes 11 factor-analytically derived scales that assess specific affects: Fear (six items; e.g., *scared, nervous*), Sadness (five items; e.g., *blue, lonely*), Guilt (six items; e.g., *ashamed, dissatisfied with self*), Hostility (six items; e.g., *angry, scornful*), Joviality (eight items; e.g., *cheerful, enthusiastic*), Self-Assurance (six items; e.g., *confident, bold*), Attentiveness (four items; e.g., *alert, concentrating*), Shyness (four items; e.g., *bashful, timid*), Fatigue (four items; e.g., *sleepy, sluggish*), Serenity (three items; e.g., *calm, relaxed*), and Surprise (three items; e.g., *amazed, astonished*). Respondents rated the extent to which they generally experienced each mood term on a 5-point scale (1 = *very slightly or not at all*, 5 = *extremely*).

Based on their item content, we hypothesized that nine TAI scales would converge strongly with counterparts on the PANAS-X. Specifically, we predicted strong convergence

between (a) TAI Anxiety and PANAS-X Fear, (b) TAI Depression and PANAS-X Sadness, (c) TAI Regret and PANAS-X Guilt, (d) TAI Anger and PANAS-X Hostility, (e) TAI Lassitude and PANAS-X Fatigue, (f) TAI Shyness and PANAS-X Shyness, (g) TAI Attentiveness and PANAS-X Attentiveness, (h) TAI Geniality and PANAS-X Joviality, and (i) TAI Experience Seeking and PANAS-X Self-Assurance.

Personality Measures

Big Five Inventory. Participants in Samples 1, 2, 4, and 5 completed the 44-item BFI (John & Srivastava, 1999), which contains 8-item scales assessing Neuroticism and Extraversion, a 10-item Openness scale, and 9-item measures of Agreeableness and Conscientiousness. Respondents rated the extent to which each statement characterized them on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The students in Sample 5 completed the BFI at both assessments.

Goldberg Markers. The Sample 5 participants completed a second set of Big Five scales at each assessment, using 45 adjectives selected from Goldberg's (1992) list of factor markers (see also, Chmielewski & Watson, 2009). Each trait was assessed using nine markers (e.g., *anxious* for Neuroticism, *talkative* for Extraversion). Respondents rated how well each term described them on a 5-point scale ranging from *very inaccurate* to *very accurate*.

NEO Personality Inventory–3. Participants in Sample 6 were assessed on the 240-item NEO Personality Inventory–3 (NEO-PI-3; McCrae, Costa, & Martin, 2005), which is an updated version of the widely used NEO PI-R (Costa & McCrae, 1992); the only change was that 38 NEO PI-R items were revised to lower the reading level and to make the instrument more appropriate for younger examinees and adults with lower educational levels. Items are rated on a 5-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

The NEO-PI-3 contains six 8-item facet scales to assess each higher order trait domain. For example, the Neuroticism facets are Anxiety, Angry Hostility, Depression, Self-Consciousness, Impulsiveness, and Vulnerability. We subsequently report correlations between the TAI and the Neuroticism, Extraversion, Agreeableness, and Conscientiousness facets; no Openness facet had a correlation $>|.35|$ with any TAI scale. We predicted that we would obtain particularly strong positive correlations between five scale pairs: (a) TAI Anxiety and NEO-PI-3 Anxiety, (b) TAI Anger and NEO-PI-3 Angry Hostility, (c) TAI Depression and NEO-PI-3 Depression, (d) TAI Vigor and NEO-PI-3 Activity, and (e) TAI Experience Seeking and NEO-PI-3 Excitement-Seeking. We also expected a strong negative association between (f) TAI Mistrust and NEO-PI-3 Trust.

Interview Ratings. The Sample 6 participants were assessed using the Mini-International Neuropsychiatric Interview (M.I.N.I.; Sheehan et al. (1998). The M.I.N.I. is a brief structured diagnostic interview that assesses symptoms of *DSM-IV-TR* (American Psychiatric Association, 2000) and *International Classification of Diseases–10th Revision (ICD-10; World Health Organization, 1993)* psychiatric disorders; we used an adapted version (with the authorization of the author) that incorporated diagnostic changes for *DSM-5* (American Psychiatric Association, 2013).² The following modules were administered in the first Sample 6 session: panic disorder, agoraphobia, PTSD, social anxiety disorder, OCD, alcohol use disorder, and (nonalcohol) substance use disorder. The modules for dysthymic disorder, MDD (which permitted the assessment of both the overall diagnosis and the nine individual MDD symptom criteria), GAD, mania, and psychotic disorders (which provides diagnoses of both psychotic disorder and mood disorder with psychotic features) were administered in Session 2.

Interviewers were graduate students and advanced undergraduate research assistants who underwent extensive training on the M.I.N.I. Graduate students had prior training in clinical interviewing and the use of the M.I.N.I., and served as trainers for the undergraduate research assistants. Training included in-depth review of *DSM* criteria for each disorder being assessed, didactics on clinical interviewing skills and administration of a semistructured interview, and a detailed overview of the administration of each item in the interview. Each research assistant was required to observe three administrations of the interview by a graduate student and subsequently be observed administering the interview on three separate occasions.

To assess interrater reliability, the interviews were audiotaped; a second rater independently scored 39 of the Session 1 interviews and 34 of the Session 2 interviews (due to audiotape problems, $N = 38$ and 33 , respectively, for some disorders). The kappa for psychotic disorder (.65) indicated good interrater reliability (see Cicchetti, 1994); values for all other ratings were in the excellent range (Cicchetti, 1994), with kappas ranging from .77 to 1.00.³

Results and Discussion

Internal Consistency. Table 1 presents internal consistency data (coefficient alphas) and average interitem correlations (AICs) for the TAI scales in the six samples. The alpha values generally were strong: 66 of 72 (91.7%) are .80 or greater, and all coefficients are $\geq .75$. Experience Seeking tended to have the lowest values, with coefficients ranging from .75 (Sample 5) to .81 (Sample 3).

Ideally, the AIC for a scale should be moderate, falling between .15 and .50 (Clark & Watson, 1995); given that these scales are designed to be specific in scope, one would expect their values to fall in the upper part of this range. The

Table 1. Internal Consistencies (Coefficient Alphas) and Average Interitem Correlations (AICs) for the TAI Scales.

Scale	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5		Sample 6	
	α	AIC										
Anger (12)	.87	.37	.90	.42	.89	.39	.89	.40	.88	.38	.89	.40
Geniality (10)	.85	.36	.86	.38	.85	.36	.81	.30	.81	.30	.83	.33
Shyness (9)	.84	.37	.86	.40	.85	.38	.88	.45	.85	.39	.81	.32
Mistrust (9)	.85	.39	.89	.47	.87	.43	.85	.39	.83	.36	.86	.41
Experience Seeking (9)	.77	.27	.78	.29	.81	.32	.78	.28	.75	.25	.77	.27
Anxiety (8)	.86	.43	.89	.51	.88	.47	.86	.43	.86	.42	.89	.51
Depression (7)	.87	.49	.89	.55	.87	.49	.90	.57	.89	.53	.87	.48
Lassitude (7)	.83	.42	.90	.55	.85	.45	.84	.43	.84	.43	.85	.44
Self-Doubt (6)	.88	.55	.90	.61	.86	.51	.89	.58	.87	.53	.86	.50
Attentiveness (6)	.83	.46	.85	.49	.80	.40	.86	.51	.86	.50	.83	.44
Regret (5)	.82	.48	.85	.54	.82	.48	.85	.52	.80	.45	.79	.44
Vigor (5)	.82	.48	.84	.52	.85	.54	.81	.45	.82	.48	.82	.48

Note. $N = 562$ (Sample 1), 374 (Sample 2), 219 (Sample 3), 341 (Sample 4), 495 (Sample 5), 383 (Sample 6). TAI = Temperament and Affectivity Inventory. The number of items for each scale is shown in parentheses.

TAI scales generally met this expectation; in fact, 69 of the 72 AICs (95.8%) ranged from .25 to .55 across samples. The only exceptions were that Depression and Self-Doubt had AICs of .57 and .58, respectively, in Sample 4, and that Self-Doubt had an AIC of .61 in Sample 2. Overall, these results indicate that the TAI scales are internally consistent and relatively narrow in their scope, without being overly redundant in their item content.

Internal Structure of the Temperament and Affectivity Inventory

Scale correlations. Table 2 reports correlations among the TAI scales separately in students (collapsed across Samples 1, 4, and 5) and patients (Sample 3). Most correlations are low to moderate in magnitude. In the student data, only 9 of the 66 correlations (13.6%) are $\geq .50$, and only four (6.1%) are $\geq .60$. The values tend to be somewhat higher in the patients, however; here, 18 coefficients (27.3%) are $\geq .50$ and six (9.1%) are $\geq .60$.

The four core negative affectivity scales represent a notable exception to this general trend. The correlations among these scales ranged from .55 to .65 (mean $r = .61$ after r -to- z transformation) in the students, and from .64 to .76 (mean $r = .71$) in the patients. These values are high enough to raise potential discriminant validity concerns that we will address in subsequent analyses.

Factor structure. To examine the internal structure of the instrument, we subjected the 12 TAI scales to separate principal factor analyses in each sample. Two clear and interpretable factors emerged in all six samples; they were rotated to oblique simple structure using promax (see Table 3). Factor I had seven strong, consistent markers. Six scales—Regret (mean loading = .76), Depression (mean loading = .75), Anger (mean loading = .73), Anxiety

(mean loading = .67), Mistrust (mean loading = .61), and Lassitude (mean loading = .61)—mark the high end of this dimension; Attentiveness (mean loading = $-.54$) defines its low end. Self-Doubt also loaded primarily on the high end of this dimension in four solutions (loadings ranged from .62 to .84), but had stronger negative loadings on Factor II in Samples 4 ($-.53$) and 5 ($-.55$).

Factor II is considerably smaller, but was consistently defined by Experience Seeking (mean loading = .66) and Vigor (mean loading = .55). Geniality also marked this dimension in five samples (loadings ranged from .42 to .57), but had a loading of only .20 in Sample 5. Finally, Shyness loaded primarily on Factor I in Samples 3 (.43) and 6 (.53), but was a marker of Factor II in the four remaining solutions (loadings ranged from $-.51$ to $-.71$).

Structural stability. How replicable is this two-factor structure across samples? We examined this issue by computing comparability coefficients (Finn, 1986). Unlike congruence coefficients—which are based on the similarity of the factor loadings across different analyses—comparability coefficients involve deriving regression-based factor scores for each solution (Everett & Entekin, 1980; Finn, 1986; Gorsuch, 1983; Harman, 1976). The comparability coefficient provides a more direct and stringent test of factor similarity (Everett & Entekin, 1980; Finn, 1986); moreover, it avoids the problems that can make congruence coefficients difficult to interpret (Finn, 1986; Nesselroade & Baltes, 1970; Pinneau & Newhouse, 1964).

As noted, comparability coefficients involve comparing factor scores across solutions. The two-factor solutions each generated a set of regression-based factor scoring weights, which then can be applied to the participants' standardized responses in each sample. Consequently, we have a total of

Table 2. Correlations Among the TAI Scales in the Student and Patient Samples.

TAI scale	1	2	3	4	5	6	7	8	9	10	11	12
1. Anxiety	—	.74	.72	.64	.52	.49	.53	.47	-.37	-.12	-.29	-.36
2. Depression	.62	—	.76	.68	.50	.56	.47	.54	-.48	-.08	-.49	-.43
3. Self-Doubt	.65	.65	—	.69	.43	.51	.55	.41	-.48	-.05	-.37	-.35
4. Regret	.55	.61	.60	—	.50	.42	.43	.52	-.36	.06	-.32	-.33
5. Anger	.45	.47	.30	.44	—	.29	.29	.53	-.30	.10	-.22	-.42
6. Lassitude	.46	.52	.50	.45	.35	—	.25	.23	-.51	-.12	-.36	-.22
7. Shyness	.47	.47	.59	.37	.15	.31	—	.36	-.20	-.26	-.40	-.42
8. Mistrust	.36	.48	.30	.36	.44	.29	.25	—	-.35	.01	-.35	-.46
9. Attentiveness	-.31	-.35	-.36	-.35	-.26	-.48	-.18	-.22	—	.12	.42	.38
10. Experience Seeking	-.21	-.09	-.23	.02	.08	-.04	-.34	.07	-.07	—	.41	.37
11. Vigor	-.19	-.37	-.27	-.21	-.14	-.20	-.37	-.21	.27	.25	—	.55
12. Geniality	-.28	-.47	-.24	-.28	-.34	-.16	-.35	-.40	.15	.20	.46	—

Note. $N = 1,398$ (students), 219 (patients). Student correlations are presented below the diagonal; patient correlations are shown above the diagonal. Correlations $\geq |.50|$ are in bold. TAI = Temperament and Affectivity Inventory.

Table 3. Promax Factor Loadings of the Temperament and Affectivity Inventory (TAI) Scales.

Scale	Factor I						Factor II					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
Regret	.81	.82	.84	.62	.65	.84	.09	.06	.13	-.18	-.04	.15
Depression	.70	.77	.85	.61	.74	.85	-.22	-.16	-.07	-.35	-.16	-.10
Anger	.72	.77	.67	.74	.66	.83	.12	.20	.10	.21	.21	.18
Anxiety	.73	.74	.82	.43	.49	.81	-.07	-.06	.02	-.43	-.32	-.08
Self-Doubt	.67	.62	.84	.39	.36	.75	-.18	-.28	.00	-.53	-.55	-.07
Mistrust	.52	.63	.60	.68	.59	.61	-.04	.01	-.06	.10	.04	.12
Lassitude	.69	.61	.53	.62	.58	.61	.03	-.07	-.12	-.07	.00	-.21
Attentiveness	-.53	-.65	-.44	-.55	-.55	-.52	-.10	-.05	.24	-.01	-.06	.28
Experience Seeking	.25	.28	.24	.46	.37	.34	.66	.65	.70	.69	.64	.62
Vigor	-.04	-.20	-.19	-.01	-.13	-.14	.64	.55	.65	.51	.38	.59
Geniality	-.12	-.23	-.26	-.23	-.37	-.24	.55	.44	.57	.42	.20	.53
Shyness	.22	.30	.43	.07	.07	.53	-.54	-.51	-.30	-.68	-.71	-.21

Note. $N = 562$ (Sample 1), 374 (Sample 2), 219 (Sample 3), 341 (Sample 4), 495 (Sample 5), 383 (Sample 6). Loadings $\geq |.30|$ are in boldface. S1 = Sample 1; S2 = Sample 2; S3 = Sample 3; S4 = Sample 4; S5 = Sample 5; S6 = Sample 6.

12 factor scores (six representing each factor) that can be compared in all six samples. If different solutions are highly similar, then the corresponding weights for each factor will produce scores that are very highly correlated. Everett (1983) suggested that comparability coefficients $\geq .90$ indicate that the same factors emerged across solutions.

Across the six samples, we computed a total of 90 comparability coefficients for each factor (15 coefficients per sample \times 6 samples). The results revealed an impressive level of structural similarity. Comparability coefficients for Factor I ranged from .959 to .998, with an overall mean value of .990 (median = .993). Corresponding values for Factor II ranged from .750 to .994, with an overall mean of .923 (median = .943). These data establish that the same two factors emerged in each sample.

Interpreting the factors. We also can use these factor scores to explicate the meaning and interpretation of these dimensions by correlating them with other higher order trait measures that were completed by the participants (see Table 4). First, we examined associations with the general affect scales of the PANAS-X; weighted mean correlations (averaged across Samples 1, 2, and 5) are presented in the top two rows of Table 4. These results establish a strong level of convergence between the higher order dimensions of the two instruments. Specifically, Factor I correlated .67 with the General Negative Affect scale, whereas Factor II correlated .62 with General Positive Affect. These strong correlations might be taken to suggest that the TAI factors can be interpreted as Negative Emotionality and Positive Emotionality, respectively.

Table 4. Correlations Between the TAI Factors and Higher Order Trait Scores on Other Instruments.

Instrument/Scale	Factor I	Factor II
PANAS-X		
General Negative Affect	.67	-.43
General Positive Affect	-.55	.62
Goldberg Markers		
Neuroticism	.70	-.47
Extraversion	-.35	.73
Openness	.06	.08
Agreeableness	-.44	.14
Conscientiousness	-.37	.14
BFI		
Neuroticism	.76	-.58
Extraversion	-.44	.75
Openness	-.16	.31
Agreeableness	-.57	.34
Conscientiousness	-.53	.32
NEO-PI-3		
Neuroticism	.88	-.47
Extraversion	-.51	.77
Openness	-.13	.32
Agreeableness	-.49	-.01
Conscientiousness	-.65	.45

Note. $N = 1,426$ (PANAS-X), $N = 492$ (Goldberg Markers), 1,759 (BFI), 380 (NEO-PI-3). Correlations $\geq .60$ are in boldface. TAI = Temperament and Affectivity Inventory; PANAS-X = Expanded Form of the Positive and Negative Affect Schedule; BFI = Big Five Inventory. NEO-PI-3 = NEO Personality Inventory-3.

Next, we examined relations with two sets of Big Five scales: the Goldberg Markers (assessed in Sample 5) and the BFI (these are weighted mean coefficients averaged across Samples 1, 2, 4, and 5). These data are displayed in the middle portion of Table 4. They reveal an even stronger degree of overlap with the general personality traits of Neuroticism and Extraversion: Factor I correlated .70 and .76 with Goldberg and BFI Neuroticism, respectively, whereas Factor II correlated .73 and .75 with Goldberg and BFI Extraversion, respectively. These very strong associations indicate that it would be more accurate to interpret the two TAI factors as reflecting Neuroticism and Extraversion, respectively.

Finally, we examined associations with the higher order NEO-PI-3 domain scales in Sample 6; these coefficients are shown at the bottom of Table 4. These relations are the strongest of all: Factor I correlated .88 with Neuroticism, whereas Factor II correlated .77 with Extraversion. These very substantial associations confirm that the TAI factors are best interpreted as reflecting Neuroticism and Extraversion, respectively. Moreover, they raise potential incremental validity concerns. That is, they suggest that the TAI is redundant with existing instruments such as the NEO

and, therefore, might not contribute any important new information beyond that already obtainable from current measures. We address these concerns in a subsequent series of analyses.

Relations With the PANAS-X. In establishing the construct validity of the TAI scales, it is critically important to explicate how they relate to traditional measures of trait affect. Table 5 presents correlations between the TAI and trait versions of the PANAS-X scales collapsed across two student samples (Samples 1 and 5; $N = 1,052$). Overall, these data establish a strong level of convergence between the instruments. Note that with the single exception of Mistrust, all the TAI scales correlate $\geq .50$ with at least one PANAS-X scale. Conversely, all but two PANAS-X scales—Serenity and Surprise—correlate $\geq .50$ with at least one TAI scale.

We hypothesized that nine TAI scales would correlate strongly with parallel measures on the PANAS-X. These predictions were strongly supported. Across the nine hypothesized scale pairs, the convergent correlations ranged from .53 (TAI Anger vs. PANAS-X Hostility) to .73 (TAI Depression vs. PANAS-X Sadness), with a mean value (after r -to- z transformation) of .60.

We can assess the specificity of these relations by comparing each of the convergent correlations with all the other values in its row or column of the matrix (the PANAS-X general affect scales were excluded from these analyses, given that they share items with several specific scales). We tested these differences formally using the Williams modification of the Hotelling test for two correlations involving a common variable (Kenny, 1987), comparing the convergent correlations with each of the discriminant coefficients in the same row or column. This yielded a total of 189 comparisons (21 comparisons \times 9 scales). Overall, 186 of these comparisons (98.4%) were significant ($p < .05$, one-tailed). The only exceptions were that (a) PANAS-X Guilt did not correlate more strongly with TAI Regret than with TAI Depression, (b) PANAS-X Joviality did not correlate more highly with TAI Geniality than with TAI Depression, and (c) PANAS-X Self-Assurance did not correlate more strongly with TAI Experience Seeking than with TAI Self-Doubt.

It is particularly encouraging that the TAI Anxiety, Depression, and Regret scales showed strong convergent and discriminant validity in these analyses. These data help establish that they can be meaningfully differentiated, despite the strong correlations reported in Table 2.

Finally, we correlated the TAI factor scores with the specific affect scales of the PANAS-X (not shown in Table 5). As would be expected—given its substantial association with General Negative Affect (see Table 4)—Factor I correlated most highly with Sadness ($r = .66$), Guilt ($r = .64$), Hostility ($r = .54$), Fear ($r = .52$), and Fatigue ($r = .52$); it also had a strong negative link to Joviality ($r = -.56$). Conversely, consistent with its strong association with

Table 5. Correlations Between the TAI and PANAS-X Scales (Combined Student Sample).

PANAS-X	Anxiety	Depress	Doubt	Regret	Anger	Lass	Shyness	Mistrust	Atten	Genial	Vigor	ES
General scales												
General NA	.58	.59	.44	.49	.45	.38	.29	.36	-.29	-.33	-.23	-.03
General PA	-.30	-.49	-.46	-.35	-.24	-.40	-.40	-.26	.41	.46	.51	.29
Specific scales												
Fear	.58	.47	.36	.35	.30	.29	.25	.27	-.25	-.26	-.19	-.03
Sadness	.46	.73	.50	.47	.36	.41	.34	.38	-.27	-.30	-.28	-.04
Guilt	.46	.61	.51	.60	.34	.34	.29	.32	-.26	-.33	-.23	-.01
Hostility	.42	.48	.27	.38	.53	.28	.21	.40	-.24	-.38	-.25	.06
Fatigue	.37	.39	.40	.34	.27	.70	.23	.23	-.39	-.08	-.14	-.02
Shyness	.41	.37	.40	.27	.11	.23	.61	.17	-.18	-.28	-.25	-.19
Attentiveness	-.27	-.39	-.37	-.30	-.21	-.38	-.26	-.19	.55	.34	.40	.15
Joviality	-.33	-.59	-.44	-.39	-.31	-.36	-.46	-.37	.25	.56	.44	.25
Self-Assurance	-.32	-.40	-.49	-.28	-.15	-.31	-.47	-.15	.23	.32	.38	.54
Serenity	-.43	-.44	-.38	-.32	-.41	-.33	-.19	-.25	.23	.33	.09	.21
Surprise	-.04	-.16	-.17	-.08	-.07	-.12	-.19	-.15	.10	.16	.17	.19

Note. $N = 1,052$. Correlations $\geq |.08|$ are significant at $p < .05$. Correlations $\geq |.50|$ are in boldface. TAI = Temperament and Affectivity Inventory; Depress = Depression; Doubt = Self-Doubt; Lass = Lassitude; Atten = Attentiveness; Genial = Geniality; ES = Experience Seeking; PANAS-X = Expanded Form of the Positive and Negative Affect Schedule; NA = Negative Affect; PA = Positive Affect.

General Positive Affect, Factor II was most highly related to Self-Assurance ($r = .64$) and Joviality ($r = .63$); in addition, it had a substantial negative link to Shyness ($r = -.55$).

Relations With the NEO-PI-3. Table 6 displays correlations between the TAI and the NEO-PI-3 facet scales in Sample 6 (as noted earlier, the Openness facets are omitted from these analyses). Overall, these data establish a strong level of convergence between the instruments at the lower order level. Note that with the single exception of Lassitude, all the TAI scales correlate $\geq |.60|$ with at least one NEO-PI-3 scale. The converse is not true, however; that is, many NEO-PI-3 facets do not correlate strongly with any TAI scale. This pattern is to be expected, given that the scope of the TAI (viz., emotionality) is narrower and more focused than that of the NEO-PI-3.

We hypothesized that six TAI scales would correlate strongly with parallel measures on the NEO-PI-3. These predictions were supported. Across the six hypothesized scale pairs, the convergent correlations ranged from $|.61|$ (TAI Vigor vs. NEO-PI-3 Activity; TAI Experience Seeking vs. NEO-PI-3 Excitement-Seeking) to $|.83|$ (TAI Mistrust vs. NEO-PI-3 Trust), with a mean value (ignoring sign) of $|.75|$. It is noteworthy that this mean value is considerably higher than that observed earlier with the PANAS-X (.60), which highlights the very strong links between the TAI scales and more traditional measures of personality.

Once again, we can assess the specificity of these relations by comparing each of the convergent correlations to all the other values in its row or column of the matrix; this yielded a total of 204 comparisons (34 comparisons \times 6 scales). Overall, 203 of these comparisons (99.5%) were

significant ($p < .05$, one-tailed); the only exception was that NEO-PI-3 Depression did not correlate significantly more strongly with TAI Depression (.76) than with TAI Self-Doubt (.74). Thus, these data further support the discriminant and construct validity of the TAI Anxiety, Anger, Depression, Vigor, Experience Seeking, and Mistrust scales.

Table 6 also reveals many strong associations beyond these hypothesized relations. For example, TAI Regret ($r = .68$) and Self-Doubt ($r = .74$) both correlated strongly with NEO-PI-3 Depression; the latter also was substantially related to Self-Consciousness ($r = .70$) and Vulnerability ($r = .61$). Geniality had strong links to both NEO-PI-3 Warmth ($r = .67$) and Positive Emotions ($r = .66$), whereas Shyness was associated with Self-Consciousness ($r = .63$) and low Assertiveness ($r = -.64$). Attentiveness was at least moderately correlated with all the Conscientiousness facets, but was most strongly related to Self-Discipline ($r = .68$).

Finally, we correlated the TAI factor scores with the NEO facet scales (not shown in Table 6). Factor I was strongly related to all six Neuroticism facets: Depression ($r = .81$), Anxiety ($r = .72$), Angry Hostility ($r = .72$), Vulnerability ($r = .72$), Self-Consciousness ($r = .67$), and Impulsiveness ($r = .59$); it also had substantial negative links to two Conscientiousness facets ($r_s = -.65$ and $-.60$ with Competence and Self-Discipline, respectively). Factor II had strong positive associations with four facets of Extraversion ($r_s = .63$ with Positive Emotions, $.62$ with Activity, $.55$ with Assertiveness, and $.51$ with Warmth) and with the Achievement Striving component of Conscientiousness ($r = .53$); in addition, it had substantial negative associations with two Neuroticism facets ($r_s = -.53$ and $-.50$ with Self-Consciousness and Vulnerability, respectively).

Table 6. Correlations Between the TAI and NEO-PI-3 Facet Scales (Sample 6).

NEO-PI-3 Scale	Anxiety	Anger	Depress	Doubt	Regret	Genial	Shyness	Vigor	ES	Mistrust	Atten	Lass
Anxiety	.78	.51	.66	.60	.52	-.34	.46	-.32	-.17	.34	-.52	.55
Angry Hostility	.54	.81	.61	.47	.56	-.27	.30	-.21	.17	.54	-.42	.45
Depression	.69	.54	.76	.74	.68	-.35	.50	-.38	.01	.42	-.55	.57
Self-Consciousness	.63	.33	.58	.70	.56	-.37	.63	-.33	-.15	.29	-.47	.48
Impulsiveness	.48	.52	.47	.49	.53	-.24	.22	-.23	.11	.29	-.47	.43
Vulnerability	.65	.51	.68	.61	.54	-.46	.43	-.36	-.09	.35	-.56	.49
Warmth	-.37	-.41	-.48	-.30	-.31	.67	-.42	.42	.09	-.44	.36	-.34
Gregariousness	-.25	-.22	-.29	-.26	-.18	.37	-.50	.35	.16	-.26	.21	-.29
Assertiveness	-.38	-.10	-.33	-.44	-.28	.34	-.64	.34	.34	-.10	.31	-.26
Activity	-.24	-.08	-.31	-.27	-.15	.33	-.37	.61	.35	-.05	.31	-.41
Excitement-Seeking	-.13	.10	-.11	-.09	.06	.26	-.22	.24	.61	.02	.07	-.15
Positive Emotions	-.37	-.28	-.55	-.37	-.28	.66	-.43	.46	.21	-.31	.39	-.42
Trust	-.35	-.52	-.47	-.28	-.39	.32	-.24	.24	-.13	-.83	.31	-.34
Straightforwardness	-.25	-.45	-.33	-.20	-.35	.16	.02	.02	-.46	-.28	.22	-.15
Altruism	-.26	-.38	-.37	-.21	-.28	.50	-.16	.22	-.18	-.33	.24	-.23
Compliance	-.24	-.67	-.35	-.15	-.38	.12	-.02	.04	-.37	-.45	.23	-.25
Modesty	.10	-.13	.03	.24	.08	-.10	.33	-.24	-.36	-.12	-.14	.19
Tender-Mindedness	-.03	-.12	-.11	.01	.05	.32	-.06	.10	-.01	-.18	.08	-.01
Competence	-.53	-.43	-.62	-.56	-.53	.48	-.40	.38	.02	-.32	.58	-.40
Order	-.22	-.24	-.32	-.30	-.29	.20	-.12	.21	-.02	-.15	.41	-.37
Dutifulness	-.38	-.46	-.48	-.37	-.46	.40	-.19	.39	-.13	-.34	.51	-.39
Achievement	-.33	-.30	-.47	-.44	-.36	.41	-.31	.50	.13	-.16	.50	-.44
Self-Discipline	-.46	-.42	-.55	-.51	-.48	.36	-.32	.39	.03	-.27	.68	-.51
Deliberation	-.31	-.46	-.38	-.32	-.46	.21	-.12	.17	-.32	-.25	.42	-.26

Note. $N = 432$. Correlations $\geq |.10|$ are significant at $p < .05$. Correlations $\geq |.60|$ are in boldface. TAI = Temperament and Affectivity Inventory; NEO-PI-3 = NEO Personality Inventory-3; Depress = Depression; Doubt = Self-Doubt; Genial = Geniality; ES = Experience Seeking; Atten = Attentiveness; Lass = Lassitude; Achievement = Achievement Striving.

Incremental Validity

Overview. The data we have presented clearly establish that the TAI scales and factors are strongly related to existing measures of trait affectivity and general personality—particularly the latter. As noted earlier, these strong associations raise significant concerns about the incremental validity of the instrument, which leads to the question: Why should one use the TAI, given that these other instruments already are available?

We examined the incremental validity of the TAI vis-à-vis the NEO-PI-3, which shows the strongest overall associations with the instrument (see Tables 4 and 6). We conducted two series of logistic regression analyses in which the TAI and NEO-PI-3 scales jointly were used as predictors of the interview ratings that were collected in Sample 6. In the first series of regressions (“Domain analyses”), each TAI scale and factor score was individually tested against the five NEO-PI-3 domain scales, so that there were six predictors in each analysis. For example, the five domain scales and TAI Depression were used to predict diagnoses of dysthymic disorder, PTSD, alcohol use disorder, and so on.

A significant limitation of these analyses is that they do not model the very strong facet-level relations displayed in

Table 6; note, for instance, that TAI Depression correlated .76 with NEO-PI-3 Depression in this sample. We therefore conducted a second series of logistic regressions (“Facet analyses”) in which each TAI scale and factor score was individually tested against the six facets within the most relevant NEO-PI-3 domain (i.e., the domain that correlated most highly with that TAI score); thus, there now were seven predictors in each analysis. For example, TAI Depression correlated most strongly with NEO-PI-3 Neuroticism at the domain level, so it and the six Neuroticism facets were used to predict diagnoses of dysthymic disorder, PTSD, and so on. Seven sets of analyses (those for Depression, Self-Doubt, Anxiety, Regret, Anger, Lassitude, and Factor I) involved facets of Neuroticism and five more (Shyness, Geniality, Vigor, Experience Seeking, and Factor II) used facets of Extraversion; finally, the regressions for Attentiveness and Mistrust included Conscientiousness and Agreeableness facets, respectively.

In the tables that follow, we report odds ratios (ORs) for the TAI scales/factor scores from these two series of analyses; to facilitate interpretation of these values, the scales were standardized to put them on the same metric. Note that the interview variables were scored as 0 = *absent*, 1 = *present*,

Table 7. Odds Ratios From Logistic Regression Analyses Predicting Indicators of Depression.

TAI scale	DD	MDD	C1	C2	C3	C4	C5	C6	C7	C8	C9
Domain analyses											
Depression	2.85	2.55	3.28	2.42	1.59	1.86	1.88	1.80	1.87	2.36	1.53
Self-Doubt	0.84	1.26	1.43	1.12	0.95	0.97	0.81	0.82	1.44	1.62	2.26
Anxiety	1.52	2.10	1.48	1.44	1.44	1.35	1.64	1.52	1.39	1.60	1.24
Regret	1.35	1.26	1.19	1.29	1.22	0.89	1.32	1.03	2.08	1.08	1.35
Anger	0.97	0.85	1.00	1.02	0.76	1.39	1.20	1.59	0.95	1.08	1.20
Lassitude	1.19	1.08	1.26	1.37	1.15	1.60	1.10	2.88	0.73	1.47	0.98
Shyness	0.69	1.06	1.05	0.94	1.03	0.97	0.97	1.18	0.85	1.04	1.04
Geniality	0.55	0.62	0.70	0.97	1.19	0.73	1.11	0.79	0.74	0.88	0.88
Vigor	0.95	0.66	0.78	0.82	0.84	0.83	0.76	0.77	0.82	0.70	1.07
Exper Seeking	0.94	1.02	1.04	1.00	1.18	1.24	1.28	0.89	1.23	1.20	1.64
Attentiveness	0.70	0.90	0.94	1.05	0.96	0.90	0.85	0.74	0.90	0.58	0.69
Mistrust	1.47	1.37	0.92	1.13	1.34	1.04	1.99	1.16	1.20	1.18	0.74
Factor I	2.72	2.66	2.82	2.31	1.56	1.98	2.42	2.54	2.36	3.04	2.30
Factor II	0.55	0.39	0.47	0.70	0.90	0.68	1.01	0.44	0.91	0.45	1.00
Facet analyses											
Depression	2.18	2.12	3.00	2.17	1.48	1.67	1.68	1.77	1.53	1.83	1.83
Self-Doubt	0.68	0.99	1.15	0.98	0.82	0.92	0.86	0.89	1.07	1.56	1.38
Anxiety	1.51	2.09	1.49	1.47	1.54	1.25	1.72	1.23	1.44	1.41	1.73
Regret	0.94	1.15	1.00	1.16	1.14	0.87	1.35	0.99	1.79	1.06	0.96
Anger	0.78	1.04	1.09	1.02	0.79	1.41	0.90	1.34	1.13	1.20	2.10
Lassitude	1.22	1.05	1.29	1.39	1.13	1.49	0.98	3.02	0.76	1.43	1.10
Shyness	1.24	1.83	1.87	1.37	1.49	1.29	1.56	2.15	1.59	1.70	2.02
Geniality	0.60	0.62	0.73	0.89	1.07	0.69	0.96	0.82	0.75	0.78	0.74
Vigor	0.76	0.49	0.65	0.69	0.82	0.70	0.66	0.78	0.64	0.60	0.95
Exper Seeking	0.90	1.32	1.30	1.25	1.35	1.29	1.44	0.98	1.56	1.46	2.50
Attentiveness	0.54	0.66	0.73	0.89	0.78	0.70	0.67	0.62	0.59	0.46	0.58
Mistrust	1.77	1.73	1.31	1.11	1.49	1.00	2.45	1.20	2.25	1.34	0.86
Factor I	1.54	2.37	2.43	2.00	1.47	1.73	1.99	2.19	1.92	2.43	2.98
Factor II	0.30	0.23	0.31	0.50	0.69	0.44	0.53	0.32	0.42	0.26	0.72

Note. $N = 402$. Significant odds ratios ($p < .05$) are in boldface. See text for details. TAI = Temperament and Affectivity Inventory; DD = dysthymic disorder; MDD = major depressive disorder; C1 = depressed mood; C2 = loss of interest; C3 = appetite disturbance; C4 = sleep disturbance; C5 = psychomotor disturbance; C6 = fatigue/anergia; C7 = worthlessness/guilt. C8 = cognitive problems; C9 = suicidal ideation; Exper Seeking = Experience Seeking.

so that an *OR* significantly less than 1.00 indicates that higher scores on a trait were associated with a reduced likelihood of receiving that rating (i.e., lower levels of psychopathology), whereas an *OR* significantly greater than 1.00 indicates that they were associated with an increased likelihood of receiving that rating (i.e., greater psychopathology).

In discussing these results, we focus primarily on those analyses in which a TAI scale or factor score displayed significant incremental validity in both series of regressions. These replicated findings offer the clearest, most compelling evidence of incremental validity.

Depression. We first consider analyses involving indicators of depression. We conducted 11 sets of regressions based on two *DSM-5* depressive disorder diagnoses (dysthymic

disorder and MDD) and the nine individual MDD symptom criteria. Table 7 presents the *ORs* from these analyses.

Table 7 results demonstrate strong incremental validity for the TAI. Overall, 106 of the 308 regressions (34.4%) yielded a significant incremental contribution for a TAI scale or factor score; moreover, 32 effects replicated across both sets of analyses. As would be expected, TAI Depression displayed particularly impressive incremental validity in these data. It made a significant incremental contribution in 19 of 22 regression analyses (86.4%) and produced nine replicable effects (including both diagnoses, as well as every MDD symptom criterion except for worthlessness/guilt and suicidal ideation); these substantial associations strongly support the construct validity of the scale. Factor I yielded seven replicable effects (MDD, plus every symptom criterion except for appetite disturbance, worthlessness/guilt, and

Table 8. Odds Ratios From Logistic Regression Analyses Predicting Indicators of Anxiety.

TAI scale	GAD	PTSD	Panic	Social	Agora	OCD
Domain analyses						
Depression	1.67	1.59	0.97	1.15	0.73	1.22
Self-Doubt	1.15	0.52	0.48	0.65	0.69	1.12
Anxiety	1.86	1.54	2.08	2.59	1.29	3.09
Regret	1.04	1.06	1.02	1.40	0.93	1.91
Anger	0.74	1.39	1.53	0.83	0.96	1.83
Lassitude	0.86	0.78	1.06	1.04	0.69	1.28
Shyness	1.31	0.81	0.78	1.80	0.74	0.70
Geniality	0.93	0.53	0.72	0.71	1.01	0.98
Vigor	0.79	0.79	0.63	0.98	0.89	1.10
Experience Seeking	0.93	0.64	0.60	1.18	0.86	0.75
Attentiveness	0.82	0.70	0.63	0.83	0.79	0.62
Mistrust	1.07	1.60	1.20	0.73	1.28	1.75
Factor I	1.57	1.52	1.26	1.52	0.71	3.52
Factor II	0.50	0.50	0.43	0.63	1.02	0.91
Facet analyses						
Depression	1.24	1.58	0.91	1.31	0.95	1.19
Self-Doubt	1.15	0.40	0.56	0.60	0.68	1.40
Anxiety	1.99	1.54	1.76	2.51	1.26	3.50
Regret	1.06	1.02	1.09	1.26	0.96	2.13
Anger	0.77	1.85	1.27	1.21	0.97	2.52
Lassitude	0.75	0.73	1.08	1.23	0.83	1.14
Shyness	2.37	1.79	1.47	2.55	1.17	1.53
Geniality	0.98	0.61	0.56	0.82	0.49	1.26
Vigor	0.64	0.52	0.50	0.66	0.82	1.12
Experience Seeking	1.15	1.05	0.89	1.09	1.04	0.89
Attentiveness	0.63	0.42	0.38	0.56	0.56	0.37
Mistrust	1.36	1.92	1.62	2.17	1.71	3.37
Factor I	1.24	1.53	1.09	1.79	0.95	3.66
Factor II	0.26	0.27	0.25	0.30	0.50	0.62

Note. $N = 402$. Significant odds ratios ($p < .05$) are in boldface. See text for details. TAI = Temperament and Affectivity Inventory; GAD = generalized anxiety disorder; PTSD = posttraumatic stress disorder; Panic = panic disorder; Social = social anxiety disorder; Agora = agoraphobia; OCD = obsessive-compulsive disorder.

suicidal ideation), and Factor II produced four consistent negative associations (in relation to MDD, depressed mood, fatigue/nergia, and cognitive problems). Other notable findings were that (a) Regret displayed a replicable association with MDD Criterion 7 (worthlessness/guilt) and (b) Lassitude was a consistent predictor of both Criterion 4 (sleep disturbance) and Criterion 6 (fatigue/nergia).

Anxiety. Next, we conducted logistic regressions using six *DSM* anxiety diagnoses as criteria: GAD, PTSD, panic disorder, social anxiety disorder, agoraphobia, and OCD; Table 8 reports the results from these analyses. Once again, the TAI displays substantial incremental validity in these data. Overall, 34 of the 168 regressions (20.2%) produced a significant incremental contribution for a TAI scale or factor score; moreover, six effects replicated across both sets of analyses. In terms of the individual scales, one would predict that TAI Anxiety should display the strongest incre-

mental/criterion validity in these data; this expectation was confirmed. TAI Anxiety made a significant incremental contribution in seven of 12 analyses (58.3%). Moreover, it showed replicable effects with GAD ($ORs = 1.86$ and 1.99 in the Domain and Facet analyses, respectively), social anxiety disorder ($ORs = 2.59$ and 2.51 , respectively), and OCD ($ORs = 3.09$ and 3.50 , respectively). These results further support the construct validity of the scale. In a related vein, it is noteworthy that TAI Shyness made a significant incremental contribution to the prediction of social anxiety disorder in both sets of analyses ($ORs = 1.80$ and 2.55 , respectively); again, these results provide important validity evidence for the scale.

Mania, substance use, and psychosis. Table 9 reports the results of logistic regressions using indicators of mania, substance use (alcohol use disorder; nonalcohol substance use disorder), and psychosis (psychotic disorder; mood

Table 9. Odds Ratios From Logistic Regression Analyses Predicting Mania, Substance Use, and Psychosis.

TAI scale	Mania	Alcohol	Substance	Psychotic	Mood-Psy
Domain analyses					
Depression	1.51	1.33	1.06	1.32	1.77
Self-Doubt	0.84	1.21	1.22	1.22	1.99
Anxiety	1.09	1.22	2.18	2.91	5.57
Regret	1.10	1.66	2.09	0.95	1.97
Anger	1.28	1.57	1.24	1.16	0.55
Lassitude	0.73	0.92	0.90	0.61	1.26
Shyness	0.98	1.05	0.95	1.00	3.14
Geniality	0.77	1.08	1.89	1.51	1.07
Vigor	0.70	0.83	1.40	1.32	0.89
Experience Seeking	1.50	1.96	1.78	1.14	1.51
Attentiveness	0.60	1.87	1.31	0.93	1.14
Mistrust	1.35	0.84	1.55	0.73	0.49
Factor I	1.67	1.86	2.16	1.52	3.16
Factor II	0.88	1.98	3.47	1.45	0.47
Facet analyses					
Depression	1.10	1.27	0.90	1.27	3.48
Self-Doubt	0.96	1.25	1.05	1.33	1.28
Anxiety	1.03	1.50	1.93	2.19	8.75
Regret	1.27	1.83	2.35	1.09	1.28
Anger	0.76	1.56	1.46	1.15	1.89
Lassitude	0.59	0.84	0.80	0.67	2.16
Shyness	1.59	0.99	1.02	1.33	5.48
Geniality	0.75	1.05	1.92	1.22	1.01
Vigor	0.49	0.68	1.15	1.35	0.51
Experience Seeking	1.67	1.91	2.12	1.82	2.18
Attentiveness	0.53	1.52	1.02	0.79	0.95
Mistrust	1.35	1.01	1.63	0.99	1.04
Factor I	1.03	1.91	2.06	1.52	6.44
Factor II	0.28	1.34	1.95	1.38	0.19

Note. $N = 402$. Significant odds ratios ($p < .05$) are in boldface. See text for details. TAI = Temperament and Affectivity Inventory; Alcohol = alcohol use disorder; Substance = nonalcohol substance use disorder; Psychotic = psychotic disorder; Mood-Psy = mood disorder with psychotic features.

disorder with psychotic features) as the criteria. The TAI again displays substantial incremental validity in these data. Overall, 34 of the 140 regressions (24.3%) produced a significant incremental contribution for a TAI scale or factor score; moreover, 11 effects replicated across both sets of analyses. TAI Anxiety produced three replicable effects: It was associated with the increased likelihood of nonalcohol substance use disorder ($ORs = 2.18$ and 1.93 in the Domain and Facet analyses, respectively), psychotic disorder ($ORs = 2.91$ and 2.19 , respectively), and mood disorder with psychotic features ($ORs = 5.57$ and 8.75 , respectively). Experience Seeking and Regret were the strongest predictors of substance use, as they each made significant incremental contributions in all four regression analyses (ORs ranged from 1.66 to 2.35 in these analyses). Finally, Attentiveness was associated with an increased likelihood of alcohol use disorder ($ORs = 1.87$ and 1.52 in the Domain and Facet analyses, respectively) but with a reduced probability of mania ($ORs = 0.60$ and 0.53 , respectively).

Dependability

Attrition analyses. Before turning to our primary analyses, it is important to examine whether our results are influenced by biased attrition. To examine this issue, we compared the Time 1 scores of the Sample 5 participants who did ($N = 392$) versus those who did not ($N = 103$) complete the Time 2 assessment. Only 1 of the 35 analyses (13 PANAS-X scales, 12 TAI scales, 5 BFI scales, 5 Goldberg scales) produced a significant difference: The participants who completed the Time 2 assessment were lower in TAI Experience Seeking ($M = 27.8$) than those who did not ($M = 29.2$; $t = 2.57$, $p < .05$). Thus, we see very little evidence of biased attrition in these data.

Dependability analyses. Table 10 reports short-term dependability correlations for the BFI, Goldberg, PANAS-X, and TAI scales in Sample 5. The BFI (mean $r = .84$, range = $.80$ to $.89$) and Goldberg scales (mean $r = .84$, range = $.78$ to $.89$) provide a useful benchmark to evaluate

Table 10. Two-Week Dependability Correlations in Sample 5.

Scale	Dependability correlation
Goldberg Extraversion	.89
BFI Extraversion	.89
TAI Anger	.86
Goldberg Conscientiousness	.85
BFI Neuroticism	.85
TAI Shyness	.85
Goldberg Neuroticism	.84
BFI Conscientiousness	.84
PANAS-X Joviality	.83
TAI Depression	.82
TAI Attentiveness	.82
PANAS-X General Positive Affect	.81
BFI Openness	.81
TAI Self-Doubt	.81
PANAS-X Self-Assurance	.81
BFI Agreeableness	.80
Goldberg Agreeableness	.80
TAI Vigor	.80
TAI Experience Seeking	.80
TAI Anxiety	.79
Goldberg Openness	.78
TAI Lassitude	.77
TAI Mistrust	.77
PANAS-X Sadness	.77
PANAS-X Guilt	.77
PANAS-X General Negative Affect	.75
PANAS-X Shyness	.75
TAI Regret	.74
TAI Geniality	.74
PANAS-X Serenity	.74
PANAS-X Fatigue	.73
PANAS-X Fear	.70
PANAS-X Attentiveness	.67
PANAS-X Hostility	.66
PANAS-X Surprise	.59

Note. $N = 392$. TAI values are shown in boldface. BFI = Big Five Inventory; TAI = Temperament and Affectivity Inventory; PANAS-X = Expanded Form of the Positive and Negative Affect Schedule.

the dependability of the trait affect scales. The PANAS-X scales had a mean dependability correlation of .75 (range = .59 to .83) in this sample, consistent with previous findings that they are more susceptible to transient error than standard personality measures (Chmielewski & Watson, 2009; Watson, 2004). The TAI scales had a mean dependability coefficient of .80 (range = .74 to .86) in these data, suggesting that they had narrowed this gap but not eliminated it entirely.

To quantify these effects more precisely, we conducted two sets of analyses using the Pearson-Filon Test, which evaluates the difference between two correlations consisting

of four nonoverlapping variables from the same sample (Kenny, 1987). In the first series, we compared the BFI and Goldberg Neuroticism scales with parallel sets of negative affect scales from the PANAS-X (Fear, Sadness, Guilt, Hostility) and the TAI (Anxiety, Depression, Regret, Anger). The PANAS-X scales showed significantly poorer dependability than the Neuroticism scales in all eight comparisons (z s ranged from 2.73 to 6.30). In contrast, the TAI comparisons yielded mixed results. The TAI Anxiety and Regret scales were less dependable than Neuroticism in all four comparisons (z s ranged from 1.95 to 4.07; $p < .05$, one-tailed); however, none of the comparisons involving Depression or Anger approached significance (z s ranged from -0.53 to 1.35).

Second, we directly compared the dependability coefficients for the nine parallel pairs of PANAS-X and TAI scales identified earlier. Five TAI scales were significantly more dependable ($p < .05$, one-tailed) than their PANAS-X counterparts: TAI Anxiety ($r = .79$) > PANAS-X Fear ($r = .70$), TAI Depression ($r = .82$) > PANAS-X Sadness ($r = .77$), TAI Anger ($r = .86$) > PANAS-X Hostility ($r = .66$), TAI Shyness ($r = .85$) > PANAS-X Shyness ($r = .75$), and TAI Attentiveness ($r = .82$) > PANAS-X Attentiveness ($r = .67$). Three other comparisons were nonsignificant: TAI Regret ($r = .74$) versus PANAS-X Guilt ($r = .77$), TAI Lassitude ($r = .77$) versus PANAS-X Fatigue ($r = .73$), and TAI Experience Seeking ($r = .80$) versus PANAS-X Self-Assurance ($r = .81$). Finally, PANAS-X Joviality ($r = .83$) was significantly more dependable than TAI Geniality ($r = .74$).

Overall, these results are quite similar to those obtained earlier with the TEQ (Chmielewski & Watson, 2009; Watson, 2004). That is, they suggest that although the TAI scales are somewhat more dependable than their PANAS-X counterparts, they still tend to be less stable than standard measures of the Big Five.

General Discussion

Summary of Results

Basic Properties of the Temperament and Affectivity Inventory. We created the TAI to address the limitations of existing trait affect inventories. Our basic goal was to create a hybrid instrument that (a) contained the types of content included within existing affectivity measures but (b) used a standard personality questionnaire format. We believe that we were successful in achieving that goal. The results presented in Table 5 establish a strong level of convergence between the TAI and the trait version of the PANAS-X. It is particularly noteworthy that with the single exception of Mistrust (which was designed to tap content similar to the MACL Skepticism scale), all the TAI scales correlated $\geq .50$ with at least one PANAS-X scale. Conversely, all but two PANAS-X scales—Serenity and Surprise—correlated $\geq .50$

with at least one TAI scale. Moreover, nine scale pairs demonstrated a clear convergent/discriminant pattern across the two instruments. Thus, our data indicate that the TAI assesses the same basic content as the PANAS-X.

Nevertheless, the TAI more closely resembles a standard personality questionnaire than does the PANAS-X. For instance, Watson et al. (1999) reported that the PANAS-X General Negative Affect and General Positive Affect scales correlated .58 and .51 with Neuroticism and Extraversion, respectively (see also Beer et al., 2013). Although these represent substantial associations, our data indicate that the TAI is even more strongly related to measures of the five-factor model. Most notably, the two TAI factors correlated .88 and .77 with NEO-PI-3 Neuroticism and Extraversion, respectively, in Sample 6. In a related vein, several TAI scales had correlations $>|.70|$ with specific NEO-PI-3 facet scales (see Table 6). Based on this evidence, it appears that our efforts to create a hybrid measure were largely successful.

Reliability Data. The data presented in Table 1 indicate that the TAI scales are internally consistent and are relatively narrow in their scope. Across our six samples, 66 of 72 coefficient alphas were $\geq .80$, and all values were at least .75. Furthermore, 69 of the 72 AICs ranged from .25 to .55 across samples, indicating that the scales generally consist of moderately correlated items.

The TAI scales also produced strong evidence of dependability (i.e., test-retest reliability). The scales had a mean dependability coefficient of .80 (range = .74 to .86) in the Sample 5 data (see Table 10). It is noteworthy, moreover, that five TAI scales—Anxiety, Sadness, Anger, Shyness, and Attentiveness—were significantly more stable than their counterparts on the PANAS-X (although it also should be noted that PANAS-X Joviality was significantly more stable than TAI Geniality). Although these results are quite encouraging, the dependability values for the TAI still were not as consistently high as those for the BFI (mean $r = .84$) and Goldberg scales (mean $r = .84$). Furthermore, targeted comparisons revealed that the TAI Anxiety and Regret scales were significantly less stable than Neuroticism in all four comparisons.

It therefore appears that although the TAI scales have narrowed the dependability gap between traditional trait affect measures and standard personality scales, they have not eliminated it entirely. In this regard, the specific factors that influence the dependability of trait measures remain poorly understood (for discussions of this issue, see Chmielewski & Watson, 2009; McCrae, 2015; McCrae, Kurtz, Yamagata, & Terracciano, 2011; Wood & Wortman, 2012). This is an important issue for future research.

Validity Data. We presented two series of analyses to establish the convergent and discriminant validity of the TAI

scales. The latter issue is particularly important in light of the strong associations we observed among the four core negative affect scales (Anxiety, Depression, Self-Doubt, Regret; see Table 2), which raise significant concerns regarding their distinctiveness. First, we compared nine TAI scales with hypothesized counterparts on the PANAS-X (see Table 5). The convergent correlations ranged from .53 to .73, with a mean coefficient of .60. Furthermore, these convergent correlations were significantly higher than all other values in their row or column of the matrix in 186 of 189 comparisons (98.4%). Second, we predicted that six TAI scales would be strongly related to NEO-PI-3 facets. These analyses (see Table 6) yielded a mean convergent correlation of $|.75|$; moreover, these coefficients were significantly greater than all other comparisons in 203 of 204 analyses (99.5%) of discriminant validity.

Taken together, these data establish the convergent and discriminant validity of 11 TAI scales. For example, the TAI Anxiety scale displays strong and specific associations with PANAS-X Fear and NEO-PI-3 Anxiety, TAI Depression converges well with PANAS-X Sadness and NEO-PI-3 Depression, TAI Regret is strongly linked to PANAS-X Guilt, and TAI Anger is related to PANAS-X Hostility and NEO-PI-3 Angry Hostility. One significant limitation of these data, however, is that we did not establish similar links for the TAI Self-Doubt Scale: Although it is substantially related to other measures (e.g., $r = .74$ and $.70$ with NEO-PI-3 Depression and Self-Consciousness, respectively), it does not show a clear convergent/discriminant pattern with any one of them. This is an important limitation that needs to be addressed in future research.

Finally, we examined the incremental and criterion validity of the TAI by conducting joint logistic regression analyses with the NEO-PI-3 domain and facet scales (see Tables 7-9). The TAI scales demonstrated substantial incremental validity in these regressions. Overall, the TAI scales and factor scores made a significant incremental contribution in 174 analyses; this included 49 effects that replicated across the Domain and Facet analyses. Furthermore, one or more TAI scales was a replicable predictor of 18 of the 22 interview criteria (81.8%) that were included in these analyses; the only exceptions were suicidal ideation, PTSD, panic disorder, and agoraphobia. Clearly, the scales tap a range of clinically important variance.

Consistent with expectation, TAI Depression displayed particularly impressive incremental/criterion validity in predicting the depression criteria. It made a significant incremental contribution in 19 of 22 regressions (86.4%) and produced nine replicable effects. Similarly, TAI Anxiety was the strongest predictor of the anxiety diagnoses: It made a significant incremental contribution in 7 of 12 analyses (58.3%) and showed three replicable effects. Other noteworthy findings were that (a) Regret displayed a replicable association with interview ratings of worthlessness/

guilt, (b) Lassitude was a consistent predictor of both sleep disturbance and fatigue/anergia, (c) Shyness made a significant incremental contribution to the prediction of social anxiety disorder diagnoses, and (d) Experience Seeking and Regret both contributed significantly in all the substance use analyses.

Limitations and Future Directions

Further Validation of the Temperament and Affectivity Inventory. The results we have presented are very encouraging and indicate that the TAI provides reliable and valid measures of specific types of trait affectivity. Validity is a complex and ongoing process, however, and additional research is needed to explicate the construct validity of the instrument more fully. We already have noted, for example, that further work is needed to establish the nature and correlates of the Self-Doubt scale. More generally, the TAI scales need to be related to a broader range of variables. For example, as was discussed earlier, traditional trait affect ratings display substantial links to indicators of happiness, subjective well-being, and satisfaction (e.g., life satisfaction, job satisfaction, marital satisfaction; Chang & Sanna, 2001; Connolly & Viswesvaran, 2000; Heller et al., 2002; Watson et al., 2004). It will be important to establish that the TAI scales display similar associations with these types of variables.

In a related vein, further data are needed to explicate the incremental validity of the TAI. We were able to show that the instrument shows strong incremental predictive power in relation to the NEO-PI-3 (see Tables 7-9), but it will be important to conduct additional analyses along these lines. In particular, it will be informative to evaluate the TAI against the PANAS-X. As was discussed earlier, the trait form of the PANAS-X has demonstrated strong criterion validity in relation to measures of affect, personality, satisfaction, and psychopathology. The data we have presented in this article demonstrate that the TAI offers certain advantages over the PANAS-X, including (a) stronger relations with personality and (b) higher levels of dependability. Nevertheless, additional data clearly are needed to establish that the TAI provides important information that is not already available from more traditional measures of trait affectivity.

Establishing incremental validity is particularly important because the TAI (which consists of 93 full sentences) takes somewhat longer to complete than older trait affect measures such as the PANAS-X (which consists of 60 single words or short phrases). Although this additional time expenditure may be a concern to some researchers, it should be kept in mind that the TAI still is much more focused than broader personality instruments such as the NEO-PI-3 (which consists of 240 full sentences) and typically takes respondents only 10 to 15 minutes to complete. Moreover,

similar to the PANAS-X (see, Watson & Clark, 1999), the TAI is modular in form, such that investigators facing more severe time constraints can select and assess only those scales that are most relevant to their research.

Further Examination of Stability. It also will be important for future research to examine the temporal stability of the instrument more broadly. The data we presented here are limited to comparing the short-term dependability of the TAI with that of conceptually and empirically related trait measures. As we have discussed, although the TAI scales compare favorably to the PANAS-X, they still do not fully achieve the level of dependability seen with standard Big Five measures. Will these differences replicate in a new sample? Will they be maintained over longer time spans in which true change may occur (e.g., intervals of 1 or 2 years)? These are important questions that need to be addressed in further research.

Need for Informant Data. As discussed previously, traditional trait affect scores show substantially poorer self-other agreement than standard personality scales, even strongly related measures with very similar content (see, Beer et al., 2013; Watson et al., 2000b). In this regard, one particularly salient limitation of our data is that we were not able to examine the extent to which the TAI scales improve on traditional instruments such as the PANAS-X by displaying superior levels of self-other agreement. This is a crucial issue for future research.

Clarifying the Nature and Structure of Affect. In developing the TAI, we felt that it was important that its content be directly linked to more traditional measures of affect. Consequently, in creating the original TAI item pool, we relied heavily on the content included in prominent multi-affect mood questionnaires such as the DES-IV, MACL, MAACL-R, PANAS-X, and POMS 2. In this way, we were able to ensure that the final instrument modeled content similar to that included in these older mood measures. For instance, 9 of the 12 TAI scales show a clear convergent and discriminant pattern with corresponding scales on the PANAS-X (see Table 5).

Nevertheless, this assessment strategy raises two potential concerns. First, the domain of mood/affectivity is somewhat ill defined, and one might legitimately question whether some of the content in these older instruments—and, by extension, in the TAI itself—truly represents affect per se. For example, the TAI Mistrust scale is modeled after content subsumed within the MACL Skepticism scale, which includes descriptors such as *suspicious*, *dubious*, and *skeptical*. Similarly, TAI Experience Seeking was modeled after the content contained in MAACL-R Sensation Seeking (which contains such terms as *wild*, *adventurous*, and *daring*) and PANAS-X Self-Assurance (which includes *bold*,

daring, and *fearless*). Do these scales actually assess affect? Given that we currently lack consensus regarding the nature and scope of this domain, it is impossible to provide a definitive answer to that question. However, we again will point out that the TAI is modular in form, such that investigators do not necessarily need to include these scales in their studies.

Second, we currently lack a consensual structure of affect at the specific, lower order level (see, Watson & Vaidya, 2013). Accordingly, it is unclear what types of affect need to be modeled in a comprehensive measure of affectivity. Although the TAI represents a good start, it seems likely that other important types of affectivity can be identified and assessed. We therefore encourage other researchers to build on our work with the TAI to create a more complete and comprehensive approach to the assessment of trait affectivity.

Acknowledgments

We thank Lee Anna Clark, Stephanie Ellickson-Larew, John Humrichouse, Kasey Stanton, and Nadia Suzuki for their help in the preparation of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Institute of Mental Health Grant R01-MH068472 to David Watson.

Notes

1. Fifty-five of these participants also are included in Sample 3. Because of this overlap, they are dropped from all analyses reporting results from both samples (reducing the N to 383).
2. This study was ongoing when *DSM-5* was finalized. The version we used included the proposed changes for GAD that later were rejected by the American Psychiatric Association. Thus, our version of the GAD diagnosis differs slightly from that in *DSM-5*.
3. Kappas could not be computed for two low base-rate diagnoses—agoraphobia and mood disorder with psychotic features—because none of the rescored cases met criteria for these disorders.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Beer, A., Watson, D., & McDade-Montez, E. (2013). Self-other agreement and assumed similarity in neuroticism, extraversion, and trait affect: Distinguishing the effects of form and content. *Assessment, 20*, 723-737.
- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.
- Chang, E. C., & Sanna, L. J. (2001). Optimism, pessimism, and positive and negative affectivity in middle-aged adults: A test of a cognitive-affective model of psychological adjustment. *Psychology and Aging, 16*, 524-531.
- Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology, 97*, 186-202.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.
- Connolly, J. J., & Viswesvaran, C. (2000). The role of affectivity in job satisfaction: A meta-analysis. *Personality and Individual Differences, 29*, 265-281.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18*, 197-218.
- Everett, J. E., & Entekin, L. V. (1980). Factor comparability and the advantages of multiple group factor analysis. *Multivariate Behavioral Research, 2*, 165-180.
- Finn, S. E. (1986). Structural stability of the MMPI in adult males. *Journal of Consulting and Clinical Psychology, 54*, 703-707.
- Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*, 45-55.
- Gnambs, T. (2014). A meta-analysis of dependability coefficients (test-retest reliabilities) for measures of the Big Five. *Journal of Research in Personality, 52*, 20-28.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.
- Heller, D., Judge, T. A., & Watson, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal of Organizational Behavior, 23*, 815-835.
- Heuchert, J. P., & McNair, D. M. (2012). *POMS 2*. North Tonawanda, NY: Multi-Health Systems.
- Hogan, R. T. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *1982 Nebraska Symposium on Motivation* (pp. 55-89). Lincoln: University of Nebraska Press.

- Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, *64*, 847-860.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102-138). New York, NY: Guilford.
- Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3-25). New York, NY: Russell Sage Foundation.
- Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston, MA: Little, Brown.
- McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*(2), 97-112. doi:10.1177/1088868314541857
- McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*, 261-270.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*, 28-50.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- Naragon-Gainey, K., Watson, D., & Markon, K. E. (2009). Differential relations of depression and social anxiety symptoms to the facets of extraversion/positive emotionality. *Journal of Abnormal Psychology*, *118*, 299-310.
- Nesselroade, J. R., & Baltes, P. B. (1970). On a dilemma of comparative factor analysis: A study of factor matching based on random data. *Educational and Psychological Measurement*, *30*, 935-948.
- Nowlis, V. (1965). Research with the Mood Adjective Check List. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality* (pp. 352-389). New York, NY: Springer.
- Nowlis, V. (1970). Mood: Behavior and experience. In M. B. Arnold (Ed.), *Feelings and emotions: The Loyola Symposium* (pp. 261-277). New York, NY: Academic Press.
- Pinneau, S. R., & Newhouse, A. (1964). Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika*, *29*, 271-281.
- Pytlík Zillig, L. M., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral and cognitive processes represented in the Big 5 personality inventories. *Personality and Social Psychology Bulletin*, *28*, 847-858.
- Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3-30.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513-523.
- Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61-84). New York, NY: Russell Sage Foundation.
- Schwarz, N., & Sudman, S. (1994). *Autobiographical memory and the validity of retrospective reports*. New York, NY: Springer-Verlag.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavas, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*, 34-57.
- Stone, A. A. (1987). Event content in a daily survey is differentially associated with concurrent mood. *Journal of Personality and Social Psychology*, *52*, 56-58.
- Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26-39). New York, NY: Russell Sage Foundation.
- Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, *20*, 663-678.
- Thayer, R. E. (1978). Factor analytic and reliability studies on the Activation-Deactivation Adjective Check List. *Psychological Reports*, *42*, 747-756.
- Vaidya, J. G., Gray, E. K., Haig, J., Mroczek, D. K., & Watson, D. (2008). Differential stability and individual growth trajectories of Big Five and affective traits during young adulthood. *Journal of Personality*, *76*, 267-304.
- Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology*, *83*, 1469-1484.
- Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*, 319-350.
- Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. *Journal of Personality and Social Psychology*, *62*, 489-505.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. Unpublished manuscript, University of Iowa, Iowa City.
- Watson, D., Clark, L. A., & Stasik, S. M. (2011). Emotions and the emotional disorders: A quantitative hierarchical perspective. *International Journal of Clinical and Health Psychology*, *11*, 429-442.
- 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., Hubbard, B., & Wiese, D. (2000a). General traits of personality and affectivity as predictors of satisfaction in intimate relationships: Evidence from self- and partner ratings. *Journal of Personality*, *68*, 413-449.
- Watson, D., Hubbard, B., & Wiese, D. (2000b). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546-558.

- Watson, D., & Humrichouse, J. (2006). Personality development in emerging adulthood: Integrating evidence from self- and spouse-ratings. *Journal of Personality and Social Psychology, 91*, 959-974.
- Watson, D., Klohnen, E. C., Casillas, A., Nus Simms, E., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality, 72*, 1029-1068.
- Watson, D., O'Hara, M. W., Simms, L. J., Kotov, R., Chmielewski, M., McDade-Montez, E., . . . Stuart, S. (2007). Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological Assessment, 19*, 253-268.
- Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin, 125*, 601-610.
- Watson, D., & Vaidya, J. (2013). Mood measurement: Current status and future directions. In J. A. Schinka & W. Velicer (Eds.), *Handbook of psychology. Volume 2: Research methods in psychology* (2nd ed., pp. 369-394). New York, NY: Wiley.
- Watson, D., & Walker, L. M. (1996). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology, 70*, 567-577.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820-838.
- Wood, D., & Wortman, J. (2012). Trait means and desirabilities as artifactual and real sources of differential stability of personality traits. *Journal of Personality, 80*, 665-701.
- World Health Organization. (1993). *International Classification of Diseases* (10th rev.). Geneva, Switzerland: Author.
- Zuckerman, M. (1960). The development of an Affect Adjective Checklist for the measurement of anxiety. *Journal of Consulting Psychology, 24*, 457-462.
- Zuckerman, M., & Lubin, B. (1965). *Manual for the Multiple Affect Adjective Check List*. San Diego, CA: Educational and Industrial Testing Service.
- Zuckerman, M., & Lubin, B. (1985). *Manual for the MAACL-R: The Multiple Affect Adjective Check List—Revised*. San Diego, CA: Educational and Industrial Testing Service.