

Huelsman, T. J., Furr, R. M., & Nemanick, R. C., Jr. (2003). Measurement of dispositional affect: Construct validity and convergence with a circumplex model of affect. *Educational and Psychological Measurement*, 63(4): 655-673. Published by Sage (ISSN: 1552-3888). doi:10.1177/0013164403251331.

Measurement of Dispositional Affect: Construct Validity and Convergence with a Circumplex Model of Affect

Timothy J. Huelsman, R. Michael Furr, and Richard C. Nemanick Jr.

ABSTRACT

A circumplex model of affect has recently gained tentative acceptance by prominent researchers investigating the structure of dispositional affect. The present study examines the issue of construct validity of several existing measures of affect and their fit with the circumplex model. A traditional multitrait-multimethod matrix approach is utilized first, followed by a confirmatory factor analytic approach. These analyses demonstrate that the data collected using these four scales are characterized by generally good concurrent and discriminant validity. Furthermore, the data are in partial agreement with the proposed circumplex model of affect. Issues for the further development of the circumplex model and the use of these measures are discussed.

Recent years, 1999 in particular, have witnessed a flurry of scholarly activity addressing the structure of affect. Long after Russell (1980) introduced his circumplex model of affect, the structure of affect still provokes lively discussion and disagreement. Unfortunately, a commensurate body of literature addressing the measurement of affect has not accompanied the debate over the structure of affect. The present study examines four measures of dispositional affect in the context of a circumplex model of mood that has been the recipient of support from several diverse sources.

STRUCTURE OF AFFECT

It is true that disagreements over the structure of affect have yet to be settled definitively, and that the resolution of these debates may have important implications for the measurement of affect. However, it is unlikely that the structure debate, which dates back to at least the 1950s (Russell & Carroll, 1999a, 1999b), will be resolved in the near term. One needs only to review the recent articles in the *Journal of Personality and Social Psychology* (Diener, 1999; Green, Salovey, & Traux, 1999; Russell & Feldman Barrett, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999), *Psychological Bulletin* (Russell & Carroll, 1999a, 1999b; Watson & Tellegen, 1999), and *Psychological Science* (Green & Salovey, 1999; Tellegen, Watson, & Clark, 1999a, 1999b) to see that the issue still presents many unanswered questions, along with several parties of willing and eminently qualified researchers to continue the discussion.

Given the current situation, it seems there are two paths to take regarding the measurement of affect. The first course requires patient waiting for the emergence of a truly consensual structure of affect, after which measurement efforts would commence; as noted, this may prove to be a long wait. The second course requires either the examination and improvement of the measures of affect presently available, or the development of new measures. Because affect will undoubtedly continue to be the focus of empirical inquiry across a wide variety of psychological specialties, we propose that the second course is the better choice.

The second course is, perhaps, even more reasonable, given the recent agreement on a common structure that has apparently been reached by several prominent researchers. Russell and Carroll (1999b) state that "it would be most unfortunate if quibbles on the sidelines were to obscure the agreement on center stage" (p. 611). They have adopted as a "working semantic model" (Russell & Carroll, 1999a, p. 7) the circumplex model of affect (see Russell, 1980; Russell & Feldman Barrett, 1999; Yik, Russell, & Feldman Barrett, 1999). In the most recent version of their circumplex model of affect, Russell and Carroll and their colleagues (Carroll, Yik, Russell, & Feldman Barrett, 1999; Russell & Carroll, 1999a) describe the structure of affect as consisting of bipolar "clusters," three positive affect clusters and three negative affect clusters, that vary in their levels of activation. Thus, they offer six clusters of affect arranged at opposite ends of three bipolar continua (see Figure 1): positive affect/high activation (e.g., excited) versus negative affect/low activation (e.g., lethargic), positive affect/medium activation (e.g.,

happy) versus negative affect/medium activation (e.g., unhappy), and positive affect/low activation (e.g., calm) versus negative affect/high activation (e.g., tense). Note that a potential fourth pair of clusters, pure activation versus pure deactivation, does not receive equal attention in their model.

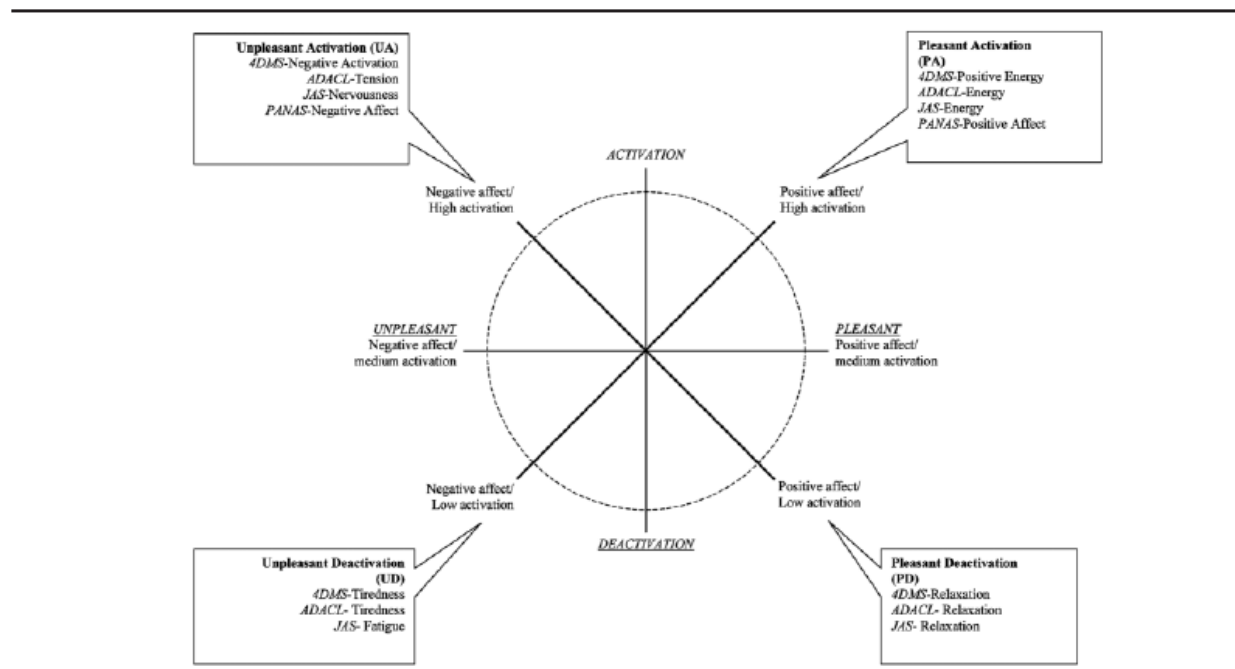


Figure 1. The circumplex model of affect.

Note. 4DMS = Four-Dimensional Mood scale; ADACL = Activation-Deactivation Check List; JAS = Job Affect scale; PANAS = Positive Affect Negative Affect Schedule.

Previously, Watson, Tellegen, and their colleagues (e.g., Watson & Tellegen, 1985) had introduced an influential model of affect built on two theoretically orthogonal dimensions, negative affect and positive affect. For years, it seems, the work of Watson, Tellegen, and their colleagues has been misunderstood (Russell & Carroll, 1999b). Much of the misunderstanding centers on their use of the terms *negative affect* and *positive affect*. In their terminology, high negative affect is associated with feelings of distress and irritability and not with sadness, as some may have assumed; high positive affect is associated with feelings of enthusiasm and excitement and not with happiness (see Figure 1). Thus, when Watson and Clark (1997) determined that negative affect and positive affect were largely independent (their most controversial finding), they did not mean that happiness and sadness were independent. To avoid any further confusion of this type, Tellegen et al. (1999a) have recently renamed NA and PA (their usual abbreviations for negative and positive affect) as *negative activation* and *positive activation*. Watson and Tellegen (1999) wrote that what we have is a specious controversy (and on this point Russell & Carroll, 1999b, p. 614, agree). Furthermore, recent analyses have demonstrated that there is considerable overlap between Watson and Tellegen's (1985) model and Russell's (1980) circumplex (Yik et al., 1999). We will take the words of Russell and Carroll (1999b) as a final statement on this issue:

One point of confusion is that (Watson and Tellegen's) definition of positive and negative affect was different from the colloquial use of the terms in that Watson and Tellegen were referring to positive and negative affect that was high in arousal. Now, 14 years later, they have modified their two-dimensional structure and their circular (circumplex) ordering of affect variables, acknowledged that the dimensions of their scales are not totally independent, and changed the names they use for various aspects of positive and negative affect. . . . When the terminological fog lifts, we can detect no substantive controversy. On the central issue, we are saying the same thing they are, just in different words. (p. 614)

To avoid some of the connotations and misunderstandings associated with the use of the terms *negative affect/activation* and *positive affect/activation*, we adopt a modified version of the terminology presented by Russell and Carroll (1999a) to describe two major dimensions of affect: the pleasant activation–unpleasant deactivation (PA-UD) dimension and the unpleasant activation–pleasant deactivation (UA-PD) dimension. Although these acronyms are inelegant, we advocate that others consider this terminology as an alternative to NA and PA.

MEASUREMENT OF AFFECT

One of the critical events in the empirical study of affect has been the development of a brief scale that has become the de facto standard for measuring affect in various domains of social science: the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988). Other measures of affect have been introduced and championed, but none has gained the frequency of use across specialties in psychology and allied disciplines as the PANAS. For example, a search of the Social Sciences Citation Index in January 2002 revealed that the Watson et al. (1988) article had been cited more than 1,400 times.

Despite its wide usage, Russell and Carroll (1999a) note that there are at least two related concerns with the PANAS. One issue is that the scales do not cover the entire range of what Watson and his colleagues propose negative affect and positive affect to be. In the present terminology, the PANAS measures PA and UA (the high activation ends of these dimensions), but not UD and PD (the low activation ends of these dimensions). Nemanick and Munz (1994) and Huelsman, Nemanick, and Munz (1998) have also expressed this concern. A second and related issue noted by Russell and Carroll is that the adjective sets of the PANAS do not include semantic opposites. Borrowing the examples provided by Russell and Carroll, *interested* is present, but *bored* is not; *active* is present, but *passive* is not. These two conceptual concerns with the PANAS will be examined in the present study to determine the extent to which they actually result in empirically demonstrated shortcomings.

In the present research, we examine the PANAS along with three other measures of the PA-UD and UA-PD dimensions of affect—measures that

apparently avoid the two potential problems with the PANAS: the Four Dimensional Mood scale (4DMS) (Huelsman et al., 1998), the Activation-Deactivation Adjective Checklist (ADACL) (Thayer, 1989), and the Job Affect scale (JAS) (Burke, Brief, George, Roberson, & Webster, 1989). Although these instruments were developed for different purposes, they may share a common structure tapping into a similar set of constructs. Each of these instruments includes scales to measure the dimensions of interest, explicitly assessing the low poles that are not included on the PANAS. The 4DMS comprises separate scales for positive energy (PA), tiredness (UD), negative arousal (UA), and relaxation (PD). The ADACL comprises separate scales for energy (PA), tiredness (UD), tension (UA), and relaxation (PD). The JAS comprises separate scales for enthusiasm (PA), fatigue (UD), nervousness (UA), and relaxation (PD). The PANAS comprises two scales: one for positive affect/activation (PA) and one for negative affect/activation (UA).

That the PANAS is used to measure the full circumplex of affect, including the lower activation ends of the dimensions, is one potential problem in the measurement of affect. But another potential problem is that the three other scales comprising the focus of this investigation may not conform to the emerging structure of affect presented by Russell and his colleagues and Watson and his colleagues. Both of these latter groups of researchers propose a two-factor model, with the factors being roughly orthogonal. However, the 4DMS, ADACL, and JAS were designed to include four separate scales (the 4DMS and JAS are most explicit in this regard, although the ADACL is also intended for such use). As such, we see a need to modify the existing instruments to reflect the model they are intended to measure.

Thus, the present study assesses a two-factor structure for the 4DMS, ADACL, and JAS. For each scale, the UD items are reverse-scored and added to the PA items to yield a single PA-UD dimension that is more reflective of the model described by Russell and his colleagues and by Watson and his colleagues (see Figure 1). Similarly, the PD items are reverse-scored and added to the UA items to yield a single UA-PD dimension.

One important goal of the present study is to examine the construct validity of the use of these four instruments to assess affect through an analysis of their convergent and discriminant validity utilizing, first, a traditional multitrait-multimethod (MTMM) matrix approach (Campbell & Fiske, 1959) and then a confirmatory factor analytic approach (CFA) (Kenny & Kashy, 1992). There is a paucity of published information concerning concurrent and discriminant validity for the various measures of mood and this study attempts to provide information toward that end. By including the PANAS in this set of measures, we can provide an initial assessment of its performance versus instruments designed to assess both ends of the PA-UD and UA-PD dimensions.

Another goal of the present study is to provide further evaluation of the circumplex model of affect. This study is consistent with the exploratory approach to personality research and test construction described by Tellegen and Waller (in press). Their inductive approach, as opposed to the more traditional deductive approach, moves “from ideas to data and from data to

ideas . . . [leading] to a better map with better constructs and scales that match these constructs” (p. 5). The present study is intended to stimulate further research on the structure and measurement of mood.

METHOD

Participants

For their participation in the study, 282 introductory psychology students at two universities received course credit. Approximately half of the students were from a medium-sized, rural, state university in the southeastern United States ($n = 143$), and the others were from a medium-sized, urban, private university in the Midwest ($n = 139$). Of the 282 participants, 250 provided usable responses. Participants were treated in accordance with APA ethical standards (American Psychological Association, 1992).

Measures

Four measures of dispositional mood were used in this study. Each was utilized with trait instructions. That is, participants were instructed to indicate the extent to which an adjective described how they generally feel, or how they feel on the average. Then, to create a continuum that might more fully reflect the full range of the PA-UD dimension for the 4DMS, ADACL, and JAS, the separate PA and UD scales were combined to form a single bipolar PA-UD dimension by reverse-scoring the UD items and adding them to the PA items. Similarly, UA and PD scales were combined to form a single bipolar UA-PD dimension by reverse-scoring the PD items and adding them to the UA items.

4DMS. The 4DMS (Huelsman et al., 1998) consists of 20 adjectives and is designed to measure positive energy (PA, 4 items), tiredness (UD, 5 items), negative activation (UA, 6 items), and relaxation (PD, 5 items). Participants rated each adjective on the extent to which it generally described their mood using a 5-point rating scale ranging from 1 (*slightly or not at all*) to 5 (*definitely*).

ADACL. The ADACL (Thayer, 1989) is a 20-item list of adjectives developed to measure energy (PA), tiredness (UD), tension (UA), and relaxation (PD). The scale includes 5 items to measure each construct. The instruction set was modified for the present study to measure trait mood. The response format was also changed so that it was more consistent with the other measures in the study. Participants rated each adjective on the extent to which it generally described their mood using a 4-point rating scale ranging from 0 (*slightly or not at all*) to 3 (*definitely*).

JAS. The JAS (Burke et al., 1989) is composed of 20 adjectives designed to measure energy (PA, 6 items), fatigue (UD, 4 items), nervousness (UA, 6 items), and relaxation (PD, 4 items). The JAS instruction set was modified for use in the present study. In the original version, participants were asked to

indicate how they felt at work during the past week. The present study used a trait instruction set, as described above. Participants rated each adjective on the extent to which it generally described their mood using a 5-point rating scale ranging from 1 (*slightly or not at all*) to 5 (*definitely*).

PANAS. The PANAS (Watson et al., 1988) comprises 10 adjectives to assess positive affect (PA) and 10 adjectives to assess negative affect (UA). Participants rated each adjective on the extent to which it generally described their mood using a 5-point rating scale ranging from 1 (*slightly or not at all*) to 5 (*definitely*).

Procedure

Werecruited participants using standard subject pool recruiting at the two universities (i.e., via sign-up boards posted in the Department of Psychology). Participants reported to a standard college classroom with other students who were participating in the study. Participants received a packet containing the study surveys and an optical scan sheet. Each participant received one of four versions of the survey packets; the versions differed only in that the sequence of the measures was counterbalanced to minimize order effects. Once they had completed the surveys and recorded their answers on the scan sheets, participants received a debriefing sheet and answers to any questions they had about the study.

Statistical Analyses

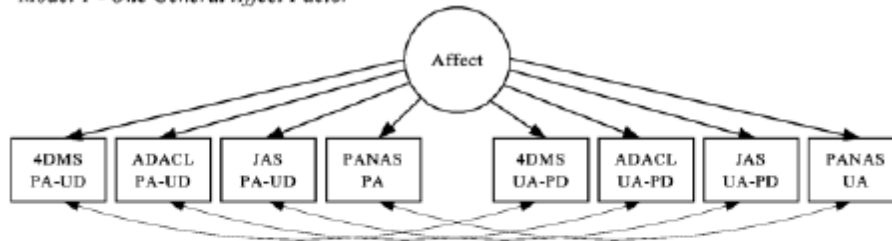
Weexamined the MTMM correlations among the eight subscales through two sets of analyses. First, we adopted Campbell and Fiske's (1959) recommendations for evaluating the zero-order correlations among the eight subscales. These analyses afford a useful overview of the measures' convergent and discriminant validities, and they provide a solid foundation for the second, more sophisticated set of analyses. Second, we evaluated a series of CFA models of trait and method effects. For the CFA, we used a correlated uniqueness model, which represents method effects through correlated error terms among the subscales of a measure. Although several CFA models are applicable to MTMM data, evaluations of such models have led researchers to recommended the correlated-uniqueness model over the alternatives (e.g., Marsh & Grayson, 1995), for both practical and interpretational reasons (Kenny & Kashy, 1992; Marsh & Bailey, 1991).

Weexamined a series of nested CFA models that differed in the structural representation of affect and of method effects (see Figure 2). Model 1 is a general affect (baseline) model that includes one trait factor on which all subscales load. Following the typical correlated-uniqueness method, the error terms between each inventory's two subscales were allowed to correlate. In Model 1, one factor underlies all eight measures of affect. Despite some suggestions that positive and negative affect are essentially two ends of a single bipolar dimension (Green, Goldman, & Salovey, 1993), we did not expect this model to fit the data well. It is included primarily as a baseline

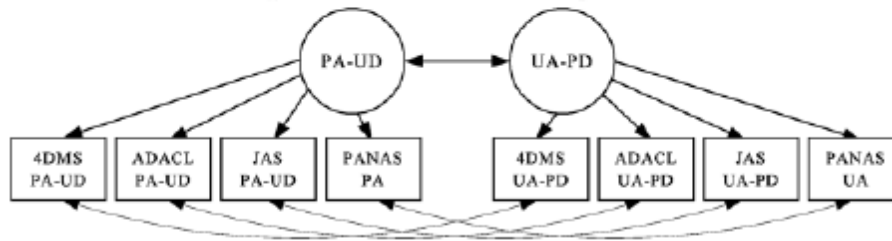
against which the next model may be gauged. Model 2 includes two correlated factors and correlated method variances; the four PA-UD subscales load on one trait factor and the four UA-PD subscales load on the other trait factor. Thus, Model 2 represents a two-factor model of affect that allows error variances to be correlated, modeling any shared method variance unique to an inventory. This model more closely resembles the circumplex, except that the two factors are not orthogonal. Model 3 includes two correlated factors and uncorrelated error variances to assess method effects. Model 3 is nested within Model 2 and differs only in terms of the presence or absence of the correlated error residuals. As articulated by Marsh and Grayson (1995), a significant decrement in fit between a model with correlated traits but no correlated error terms (e.g., Model 3) and a model with correlated traits and correlated error terms (e.g., Model 2) indicates the presence of method effects. Finally, Model 2 was compared to Model 4, which includes uncorrelated affect factors and is thus a purer representation of the circumplex model. This comparison allows us to assess the independence of the PA-UD and UA-PD dimensions.

A variety of statistics are available to evaluate the fit of a CFA model. We report several such statistics for each model: the chi-square test, the standardized root mean squared residual (SRMR) (Bentler, 1995), the Incremental Fit Index (IFI) (Bollen, 1989), the Comparative Fit Index (CFI) (Bentler, 1992), the Tucker-Lewis Index (TLI) (Tucker & Lewis, 1973), and the root mean squared error of approximation (RMSEA). The chi-square statistic has several problems, such as being strongly influenced by sample size, but it is particularly useful for determining the relative fit of nested models such as those presented here. SRMR values represent the difference between the observed correlation matrix and the reproduced correlation matrix, with small values, values approaching zero, indicating good fit. IFI, CFI, and TLI values range (roughly) between zero and 1.0, with values approaching 1.0 indicating good fit. RMSEA values represent the value of the fit function divided by the degrees of freedom, with lower values indicating better model fit. Hu and Bentler (1998, 1999) extensively evaluated a host of fit indexes and recommended a two-index strategy in which researchers present the SRMR along with at least one index from a set including the IFI, CFI, TLI, and RMSEA. In addition, they suggest that an SRMR of .09 is generally indicative of good fit in combination with values of .95 for the IFI, CFI, and TLI or a value of .06 for the RMSEA. They do stipulate, however, that with relatively small samples, such as the $N = 250$ in the current study, the TLI and the RMSEA are oversensitive, rejecting too many true models. So, for the sake of completeness, we present the TLI and the RMSEA but suggest that the SRMR, the IFI, and the CFI are the most appropriate fit indexes for the current sample.

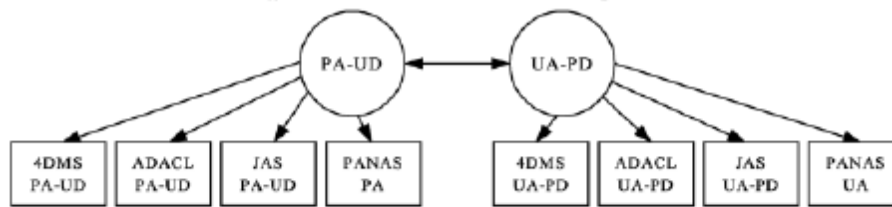
Model 1 - One General Affect Factor



Model 2 - Two Correlated Affect Factors, with Correlated Uniquenesses



Model 3 - Two Correlated Affect Factors, with Uncorrelated Uniquenesses



Model 4 - Two Uncorrelated Affect Factors, with Correlated Uniquenesses

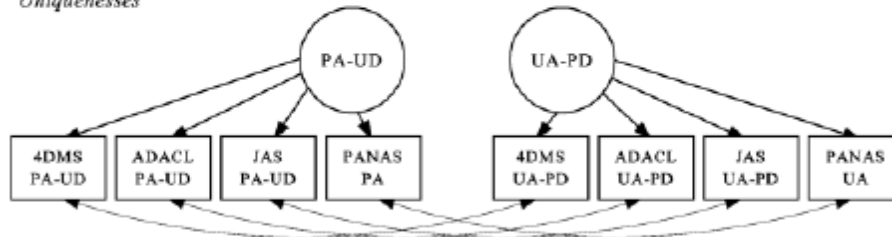


Figure 2. Structural models 1 to 4.

Note. 4DMS = Four-Dimensional Mood scale; ADACL = Activation-Deactivation Check List; JAS = Job Affect scale; PANAS = Positive Affect Negative Affect Schedule; PA = pleasant activation; UA = unpleasant activation; PD = pleasant deactivation; UD = unpleasant deactivation.

RESULTS

Descriptive Statistics

Descriptive statistics for all study scales are presented in Table 1. Preliminary analyses indicated that differences in the means and variances across samples were not statistically significant (we employed the Bonferroni technique dividing the conventional .05 alpha by the 14 tests, for a cutoff probability

level of .004), so the data were combined for subsequent analyses. Subscale scores from all four measures show acceptable reliability. The PANAS scores have the highest internal consistency reliabilities (mean $\alpha = .89$) and the ADACL has the lowest (mean $\alpha = .71$). Across the four measures, the UA-PD subscale scores have lower reliabilities (mean $\alpha = .77$) than the PA-UD subscale scores (mean $\alpha = .83$). In addition, all skew and kurtosis values are minimal, which suggests that all 8 subscales have relatively normal, symmetric distributions.

Table 1
Descriptive Statistics for the Affect Subscales (N = 250)

Scale/Subscale	Number of Items	Cronbach's α	<i>M</i>	<i>SD</i>	Skew	Kurtosis
4DMS						
PA-UD	9	.83	30.47	6.56	-.30	-.06
UA-PD	11	.85	27.44	6.54	.24	-.15
ADACL						
PA-UD	10	.80	26.34	5.77	-.12	-.62
UA-PD	10	.62	23.10	4.67	-.06	-.38
JAS						
PA-UD	10	.81	34.00	6.45	-.02	-.02
UA-PD	10	.72	24.56	5.00	.16	.20
PANAS						
PA	10	.89	34.79	7.38	-.21	-.09
UA	10	.88	19.86	6.88	.67	-.12

Note. 4DMS = Four-Dimensional Mood scale; ADACL = Activation-Deactivation Check List; JAS = Job Affect scale; PANAS = Positive Affect Negative Affect Schedule; PA = pleasant activation; UA = unpleasant activation; PD = pleasant deactivation; UD = unpleasant deactivation.

MTMM Correlations

Table 2 presents the MTMM correlations among the eight affect subscales and provides evidence for the convergent and discriminant validity of the four-factor structure of the scales. Campbell and Fiske (1959) offer three desiderata for MTMM correlations, all of which are generally satisfied by the correlations in Table 2. First, consider the convergent validity correlations—the correlations between two different measures of a single trait (monotrait heteromethod correlations in Campbell and Fiske's terminology). Most of these correlations (italicized) are relatively large (mean $r = .64$), and all are statistically significant at $p < .001$. Specifically, the JAS PA-UD scale has the strongest convergent correlations (mean convergent $r = .75$), and the ADACL UA-PD scale has the weakest convergent correlations (mean convergent $r = .45$).

Table 2
Correlations Among the Affect Subscales

Subscale	4DMS		ADACL		JAS		PANAS	
	PA-UD	UA-PD	PA-UD	UA-PD	PA-UD	UA-PD	PA	UA
4DMS								
PA-UD	1.0							
UA-PD	<i>-.57</i>	1.0						
ADACL								
PA-UD	<i>.65</i>	<i>-.42</i>	1.0					
UA-PD	<i>-.12</i>	<i>.45</i>	<i>-.10</i>	1.0				
JAS								
PA-UD	<i>.82</i>	<i>-.53</i>	<i>.65</i>	<i>-.09</i>	1.0			
UA-PD	<i>-.41</i>	<i>.75</i>	<i>-.26</i>	<i>.53</i>	<i>-.39</i>	1.0		
PANAS								
PA	<i>.62</i>	<i>-.39</i>	<i>.55</i>	<i>-.05</i>	<i>.79</i>	<i>-.25</i>	1.0	
UA	<i>-.48</i>	<i>.68</i>	<i>-.31</i>	<i>.38</i>	<i>-.42</i>	<i>.77</i>	<i>-.24</i>	1.0

Note. 4DMS = Four-Dimensional Mood scale; ADACL = Activation-Deactivation Check List; JAS = Job Affect scale; PANAS = Positive Affect Negative Affect Schedule; PA = pleasant activation; UA = unpleasant activation; PD = pleasant deactivation; UD = unpleasant deactivation. Convergent correlations (monotrait-heteromethod correlations) are in italics.

Second, consider the evidence for the discriminant validity. Campbell and Fiske (1959) suggest that each convergent correlation should be higher than the heterotrait-heteromethod correlations in its row and column. Heterotrait-heteromethod correlations show the relationship between different traits measured by different scales. Again, the MTMM correlations in Table 2 generally satisfy the criterion, with two inconsistencies that emerge from the relatively weak convergent correlations of the ADACL UA-PD scale. First, the convergent correlation between the ADACL UA-PD subscale and the 4DMS UA-PD subscale ($r = .45$) is weaker than the heterotrait-heteromethod correlation between the 4DMS UA-PD subscale and the JAS PA-UD subscale ($r = -.53$), and second, the convergent correlation between the ADACL UA-PD subscale and the PANASUA subscale ($r = .38$) is weaker than the heterotrait-heteromethod correlations between the PANAS UA subscale and the 4DMS PA-UD subscale ($r = -.48$) and the JAS PA-UD subscale ($r = -.42$).

Third, and for further evaluation of discriminant validity, Campbell and Fiske (1959) suggest that the convergent correlations should be stronger than heterotrait-monomethod correlations—the correlations between different trait subscales within the same measure. For example, all of the convergent correlations for the two PANAS scales are stronger than the heterotrait-monomethod correlation between the two PANAS scales ($r = -.24$). Once again, results generally satisfy the criterion, with only one exception—the convergent correlation between the ADACL UA-PD subscale and the 4DMS UA-PD subscale ($r = .45$) is weaker than the heterotrait-monomethod correlation between the two 4DMS scales ($r = -.57$). The ADACL has the best discriminant correlation ($r = -.10$), and the 4DMS has the worst discriminant correlation ($r = -.57$).

Thus, the MTMM correlations provide general support for the convergent and discriminant validity of a two-trait structure for the eight subscales. Unfortunately, interpretation of raw MTMM correlations is somewhat limited by the different reliabilities of the eight subscales. To help clarify these issues, we turn next to CFA analyses that take error of measurement into account when modeling the relationships among subscales and traits.

CFA Analyses

We tested four CFA models to evaluate the factorial structure and the method variance of the affect scales. All models were fit using EQS 5.7b maximum likelihood estimation procedures. Initial analyses indicated that although each of the eight subscales had relatively normal distributions (see Table 1), the data were not multivariate normal in our sample of 251 respondents. Multivariate normality is assumed in maximum likelihood estimation procedures, which we used for our CFA analyses. Therefore, we dropped the data from the participant contributing most to multivariate kurtosis (resulting Mardia's coefficient = 7.40). This individual was an extreme outlier ($z = 4.1$) on the 4DMS UA scale. Thus, all results reported here are based on a final sample of 250, which matches the minimum sample size of 250 recommended by Marsh and Grayson (1995) for MTMM CFA analysis. As shown in Table 3, Model 1 (one trait: general affect) provides a poor fit with the data. The SRMR is well above .09, and the IFI and CFI are well below .95.

The more familiar two-factor model of affect, Model 2 (two traits: PA-UD and UA-PD) is indeed a statistically significant improvement over Model 1 (chi-square difference = 294.56, $df = 1$, $p < .001$) and provides a fairly good overall fit to the data. As Table 3 shows, the CFI and IFI approach .95 and the SRMR is below .09. LaGrange multiplier tests reveal several parameters that could be freed to improve the overall fit of the model. In particular, allowing the 4DMS UA-PD scale to load on the PA-UD factor could improve the model's fit. However, because Model 2 provides an arguably acceptable fit to the data and because we are not interested in reshaping the model based on post hoc modifications, we present the parameter estimates for Model 2 without any modifications (see Figure 3). These parameter estimates generally reiterate the initial evaluations of convergent validity—the JAS has the strongest factor pattern coefficients and the ADACL has the weakest. In addition, the JAS and the ADACL have no correlated uniquenesses, but the 4DMS and the PANAS do have method effects. Finally, the two traits are negatively correlated.

Table 3
Goodness-of-Fit Indices for Affect Models

	χ^2	df	SRMR	IFI	CFI	TLI	RMSEA
Model 1	386.22	16	.141	.723	.720	.511	.302
Model 2	91.66	15	.078	.943	.942	.892	.143
Model 3	105.21	19	.079	.935	.935	.904	.135
Model 4	147.89	16	.233	.901	.900	.826	.182

Note. SRMR = standardized root mean-square residual; IFI = Incremental Fit Index; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean squared error of approximation. All chi-square tests are statistically significant at $p < .001$.

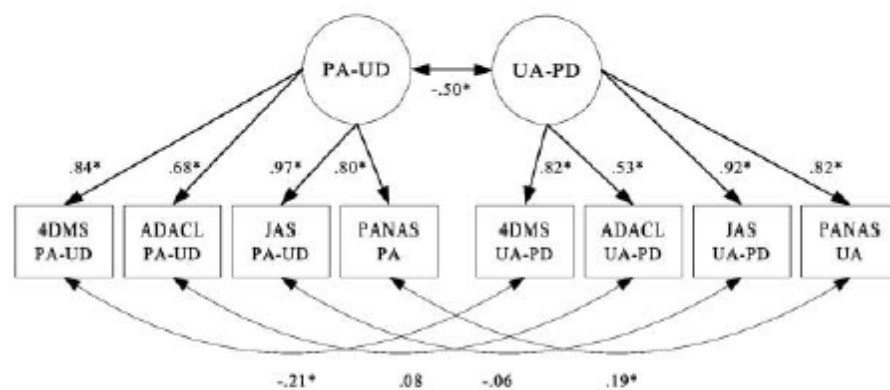


Figure 3. Parameter estimates for Model 2.

Note. 4DMS = Four-Dimensional Mood scale; ADACL = Activation-Deactivation Check List; JAS = Job Affect scale; PANAS = Positive Affect Negative Affect Schedule; PA = pleasant activation; UA = unpleasant activation; PD = pleasant deactivation; UD = unpleasant deactivation.

* $p < .05$.

Model 3 (two traits, uncorrelated error variances) demonstrated a worse fit than Model 2 (chi-square difference = 13.55, $df=4$, $p < .01$). Even though the fit indices presented in Table 3 show a relatively acceptable fit for Model 3, these analyses underscore the unique method variance found for the 4DMS and the PANAS (see Figure 3, correlated error terms). Finally, estimation of Model 4 (two uncorrelated traits) allows us to formally evaluate the correlation between PA-UD and UA-PD. As Table 3 shows, Model 4 results in worse fit than Model 2 (chi-square difference = 56.23, $df = 1$, $p < .001$). Again, this suggests that, given the inventories used in this study, PA-UD and UA-PD are negatively correlated.

DISCUSSION

One of the purposes of the present study was to assess the convergent and discriminant validity of the data collected by using the 4DMS, ADACL, JAS, and PANAS. The traditional MTMM approach provides general support for each of these four scales. The strongest support was garnered for the JAS

because of its superior convergent validity and acceptable discriminant validity, and for its inclusion of the deactivation scales. The JAS also had the highest factor pattern coefficients and low correlated method variances in the CFA analyses. The 4DMS and PANAS were somewhat weaker in this regard, with good trait factor pattern coefficients but somewhat more unique method effects than the JAS. Note that the correlation between the unique variances was negative for the 4DMS but positive for the PANAS. This reflects the difference in the zero-order correlations between the two scales from each measure (see the heterotrait-monomethod correlations in Table 2). The latent trait correlation is $-.50$ (see Figure 3), but the two 4DMS scales were even more negatively correlated ($-.57$) with each other and the two PANAS scales were less negatively correlated ($-.24$) with each other. Thus, the correlations between the unique variances in the confirmatory model adjust for the difference between the latent trait correlation and the measures' zero-order correlations. The ADACL appears to be the weakest of the scales, principally because of its low factor pattern coefficients in Model 2 of the CFA analyses. This should not be viewed as an indictment of the ADACL. In fairness, it was originally proposed as a measure of state affect—a purpose for which it may be well suited. Additionally, we modified its original response format to be more consistent with the other measures in the present study. These changes may have resulted in its poor performance.

These analyses also indicate that the failure of the PANAS to explicitly measure deactivation (PA-UD and UA-PD) may not represent a shortcoming for the PANAS. The factor pattern coefficients of the PANAS PA scale on the latent PA-UD dimension and of the PANAS UA on the latent UA-PD dimension were lower than those of the JAS but were comparable to those of the 4DMS. Although the PANAS may actually measure deactivation, this is unlikely given an item response theory (IRT) analysis reported by Baker, Zevon, and Rounds (1994) that indicated that low levels of negative affect were not measured by a set of adjectives similar to those assessing negative affect on the PANAS. A more plausible explanation is that the 4DMS, ADACL, and JAS may not truly represent deactivation. This reasoning assumes that measuring deactivation is possible, but that the adjectives chosen on these scales are not the optimal markers of deactivation. An alternative notion is that deactivation is unmeasurable because it is the absence of emotion, and the only way to measure emotion is to measure its presence. Analogously, when measuring temperature, one measures the presence or absence of heat but not the presence of cold. As Zevon and Tellegen (1982) originally noted, “If we define emotions as aroused-engaged states, then Positive and Negative Affect are best characterized as *descriptively bipolar* but *affectively unipolar* dimensions” (p. 112). It was on this reasoning that Watson et al. (1988) originally decided not to include measures of the lower ends of their constructs.

However, the fact that our results do not show that measuring deactivation adds value to the measurement model does not preclude it from being of value in examining substantive relationships involving mood. Therefore, another test of the merit of the inclusion of deactivation on scales of dispositional affect would include their use in predicting behavior, cognition, attitudes, state mood and emotion, or other personality variables. We suggest

that future research involving affect should include the PANAS and the JAS or 4DMS to see if prediction is improved by the inclusion of scales that attempt to measure PA-UD and UA-PD rather than just PA and UA (as in the PANAS).

In this study, we introduced an alternative method of scoring the 4DMS, JAS, and ADACL that renders these scales more conceptually consistent with the circumplex model of affect model offered by Russell, Carroll, and their colleagues (Carroll et al., 1999; Russell & Carroll, 1999a). This alternative scoring method works well, as illustrated by the correlations between the PA scale on the PANAS and the PA-UD scales of the JAS and 4DMS (mean $r = .71$), and the correlations between the UA scale on the PANAS and the UAPD scales on the JAS and 4DMS are strongly positive (mean $r = .73$). Furthermore, the loadings of the PA-UD and UA-PD scales for the JAS and 4DMS in Figure 3 are comparable to those for PA and UA scales for the PANAS, as previously noted. These results, along with the reasonably good fit of the data to Model 2 in the present study, suggest that this scoring protocol is viable.

Another important purpose of the present study was to further examine Russell and Carroll's (Carroll et al., 1999; Russell & Carroll, 1999a) circumplex model of affect. The CFA analyses demonstrate that a one-factor model of affect (Model 1) did not fit the data particularly well. The other models tested variations of the two-factor model of affect. All of these latter models show relatively acceptable fit to the data, but Model 2, comprising two correlated factors with correlated method variances, fit the data best, demonstrating statistically significant improvements over Model 3 (two correlated factors with uncorrelated error variances) and Model 4 (two uncorrelated factors). These results provide equivocal support for the structural model of affect proposed by Russell, Carroll and their colleagues. The data clearly fit Model 2 better, although Model 4 would provide evidence for orthogonality of the factors, a key component of the circumplex model.

The critical weakness in the fit between the current scales and the circumplex model is the correlation between PA-UD and UA-PD (see Table 2 and Figure 3). These correlations are not consistent with the orthogonality assumption of the circumplex model. However, the magnitude of these correlations is consistent with much of the extant literature (e.g., Huelsman et al., 1998; Watson & Walker, 1996; for a discussion of the issue, see Watson et al., 1988). In fact, the issue of orthogonality prompted extensive discourse by Russell and Carroll (1999a, 1999b) and Watson and Tellegen (1999). Going beyond the issues that they discuss, we suggest that it is possible that a lack of pure measures of the Deactivation poles contributes to the lack of orthogonality between the PA-UD and UA-PD dimensions. A measure of "pure" PD should be uncorrelated with PA, and a measure of "pure" UD should be uncorrelated with UA. However, this may not be the case. Rather, these correlations suggest that the "measured" UD and PD are located away from "pure" UD and PD—likely located near to the core Pleasantness-Unpleasantness dimension (this assumes that PA and UA are measured correctly and UD and PD are not—as suggested by Watson et al., 1988; Zevon & Tellegen, 1982). Such a relocation of the measured Deactivation poles may subtly shift the entire measured UA-PD and PA-UD dimensions, resulting in

nonorthogonality. Certainly, the issue of measuring Deactivation and its relationship with orthogonality deserves further attention.

CONCLUSION

The present study has utilized Tellegen and Waller's (in press) inductive approach to construct development to simultaneously examine several measures of dispositional affect and a circumplex model of affect. This study supports three main points. First, the JAS, PANAS, and 4DMS do converge to measure the PA-UD and UA-PD dimensions of dispositional affect. Second, the concerns about potential limitations of the PANAS appear unfounded in the present analyses. Third, the present data are in partial agreement with the circumplex model of affect proposed by Russell and Carroll (1999a).

REFERENCES

- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611.
- Baker, J. G., Zevon, M. A., & Rounds, J. B. (1994). Differences in positive and negative affect dimensions: Latent trait analysis. *Personality and Individual Differences*, 17, 161-167.
- Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, 112, 400-404.
- Bentler, P. M. (1995). *EQS structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Bollen, (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, 17, 303-316.
- Burke, M. J., Brief, A. P., George, J. M., Roberson, B. S., & Webster, J. (1989). Measuring affect at work: Confirmatory analyses of competing structures with conceptual linkage to cortical regulatory systems. *Journal of Personality and Social Psychology*, 57, 1091-1102.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carroll, J. M., Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). On the psychometric principles of affect. *Review of General Psychology*, 3, 14-22.
- Diener, E. (1999). Introduction to the special section on the structure of emotion. *Journal of Personality and Social Psychology*, 76, 803-804.
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, 64, 1029-1041.
- Green, D. P., & Salovey, P. (1999). In what sense are positive and negative affect independent? A reply to Tellegen, Watson, and Clark. *Psychological Science*, 10, 304-306.

- Green, D. P., Salovey, P., & Traux, K. M. (1999). Static, dynamic, and causative bipolarity of affect. *Journal of Personality and Social Psychology*, 76, 856-867.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Huelsman, T. J., Nemanick, R. C., Jr., & Munz, D. C. (1998). Scales to measure four dimensions of dispositional mood: Positive energy, tiredness, negative activation, and relaxation. *Educational and Psychological Measurement*, 58, 804-819.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15, 47-70.
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
- Nemanick, R. C., Jr., & Munz, D. C. (1994). Measuring the poles of negative and positive mood using the Positive Affect Negative Affect Schedule and Activation Deactivation Adjective Check List. *Psychological Reports*, 74, 195-199.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- Russell, J. A., & Carroll, J. M. (1999a). On the bipolarity of positive and negative affect. *Psychological Bulletin*, 125, 3-30.
- Russell, J. A., & Carroll, J. M. (1999b). The phoenix of bipolarity: Reply to Watson & Tellegen (1999). *Psychological Bulletin*, 125, 611-617.
- Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76, 805-819.
- Tellegen, A., & Waller, N. G. (in press). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In S. R. Briggs & J. M. Cheek (Eds.), *Personality measures: Vol. 1—Development and evaluation*. Greenwich, CT: JAI Press.
- Tellegen, A., Watson, D., & Clark, L. A. (1999a). On the dimensional and hierarchical structure of affect. *Psychological Science*, 10, 297-303.
- Tellegen, A., Watson, D., & Clark, L. A. (1999b). Further support for a hierarchical model of affect: Reply to Green and Salovey. *Psychological Science*, 10, 307-309.

Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.

Tucker, L. R., & Lewis, C. (1973). Reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment*, 68, 267-296.

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

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

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

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.

Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77, 600-619.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Applied Psychology*, 43, 111-122.