Measuring Mood in the 1995 NES Pilot Study

by

George E. Marcus
Department of Political Science
Williams College
Williamstown, MA 01267

and

Michael MacKuen
Department of Political Science
University of Missouri-St. Louis
St. Louis, MO 63121

I. Mood and politics

The candidate affect items, added to the NES series in 1980, have proved to be a continuing source of interest and research. A part of the 1995 NES Pilot was dedicated to revisiting the measures of candidate generated affect. As detailed in the Marcus et al (1995) proposal, the items originally were derived from the "discrete" approach though they have also been successfully used by scholars who have applied them in the "arousal" approach.

II. Mood as discrete and Mood as arousal

The discrete approach posits a number of distinct emotional states. While the number of discrete, or "basic", emotions is not a settled issue, four (fear, anger, happiness and disgust) were included in the 1995 NES Pilot (form A). In a departure from prior practice, in addition to ascertaining whether subjects had "every felt ____", a simple dichotomous measure (yes/no), subjects were also asked the frequency of occurrence (*most of the time, often, occasionally* (or "*sometimes*"), *rarely*).

The arousal approach posits just two distinct arousal states, an anxiety dimension that ranges from calm to anxious, and an enthusiasm dimension, that ranges from depressed to enthusiastic. Previous studies have consistently replicated these two dimensions from standard NES affect items in each NES study that contained the standard NES affect items (Bruce, 1991; Marcus, 1988; Marcus, & MacKuen, 1993). These two dimensions are most commonly named Negative Affect and Positive Affect. Though the standard NES affect items have proved suitable for use by scholars applying arousal theoretical approaches, they do have a serious restriction. Past practice, as previously noted, has used a dichotomous response format. By treating the group of positive affect terms (such as proud, hopeful or sympathetic) as multiple measures, it has been possible to create simple summated scales that would approximate varying degrees of arousal from low, or non-existent, to high. Analysis of the individual items themselves has required applying polychoric and polyserial correlations (since the theoretical presumption is that underlying the dichotomous items is a continuous distribution). The diminishing number of affect items included in the NES series (now down to four items) poses a serious problem: it is difficult precisely to calibrate degrees of affect. In order to address that problem, we proposed the adoption of affect items that have evolved from the arousal approach.

III. Mood as Arousal Measurement

David Watson (1988; 1988) has done the principal work on the measurement of affect from an arousal approach. He has demonstrated that affect terms consistently reproduce two dimensions of mood whether ascertained by measures of intensity or by measures of frequency and that the results are invariant across the response format.

Reproduced below is Table 5 from Watson, et al., (1988) which displays the median varimax-rotated factor loadings of the mood terms Watson has used to define the Positive and Negative Affect dimensions. We present these results to represent the typical solution obtained when mood is ascertained.

Table 1: Median Varimax-Rotated Factor Loadings of the Positive and Negative Affect Schedule (PANAS) Descriptors Across the Six Solutions. Source: (Watson, et al., 1988)

<u>Loadings On</u>

PANAS Descriptor	Positive Affect	Negative Affect
Enthusiastic	.75	-.12
Interested	.73	-.07
Determined	.70	-.01
Excited	.68	.00
Inspired	.67	-.02
Alert	.63	-.10
Active	.61	-.07
Strong	.60	-.15
Proud	.57	-.10
Attentive	.52	-.05
Scared	.01	.74
Afraid	.01	.70
Upset	-.12	.67
Distressed	-.16	.67
Jittery	.00	.60
Nervous	-.04	.60
Ashamed	-.12	.59
Guilty	-.06	.55
Irritable	-.14	.55
Hostile	-.07	.52

We have previously experimented with variants of this approach, successfully replicating this structure with adapted thermometer scales (rating from 0 to 100) and seven and five point Likert scales (Marcus, MacKuen, & Glassberg, 1989). It is important to note, as we shall see below, that a common feature of the arousal measurement, whether using frequency or intensity, or 100 point, five point or seven point scales, is that they focus the respondents attention to a single mood term (e.g., enthusiasm).

Watson (1988) also reports on the eight week test-retest reliability of the Positive and Negative Affect scales. As these data provide the basis for comparisons to the test-retest of NES measures in the 1995

Pilot and 1994 NES studies (the 1995 Pilot being a re-interview of subjects also used in the 1994 NES study), we present these below.

Table 2: Test-Retest Reliabilities of the Positive and Negative Affect Schedule (PANAS) Scales (8-Week Retest Interval) Source: (Watson, et al., 1988)

Time instruction	PANAS PA scale	PANAS NA scale
Moment	.54	.45
Today	.47	.39
Past Few Days	.48	.42
Past Week	.47	.47
Past few Weeks	.58	.48
Year	.63	.60
General	.68	.71

The test-retest reliabilities are, as is expected, modestly higher for the positive affect scale than for the negative affect. And, further, the mood measurements posited within a more immediate time frame are less stable across time than are those set in a more enduring time frame.

On the basis of the substantial measurement literature on mood measurement and on our prior field experience, cited above, we proposed four mood terms to measure positive and negative mood states. For each (enthusiastic-unenthusiastic; hopeful-discouraged; anxious-calm; and upset-relaxed) we proposed a five point Likert like response format (e.g., Does ___ make you feel *very enthusiastic, somewhat enthusiastic, neither enthusiastic nor unenthusiastic, somewhat unenthusiastic,* or *very unenthusiastic*). Note that these terms juxtaposed two distinct mood terms rather than asking subjects to rate how intensely or frequently they experience a specific mood.

Discussions in the Pilot Planning Committee lead to a further change in these formats. The response format for enthusiastic vs. unenthusiastic was changed to enthusiastic vs. indifferent, further enhancing the apparent bipolar character of the dimensions. Though we did not expect that this or the above changes would have a material impact, this alteration had serious consequences . . . as we shall see below.

IV. Results

The analysis of the 1995 Pilot data for Form B shows unambiguously that a single dimension underlies the four items meant to serve as indicators of positive and negative affect. This is the case both for the items used to measure emotional response evoked by Clinton and for the items used to measure emotional response evoked by Dole.

Table 3: Form B Affect Items in the 1995 NES Pilot

Correlation Matrix - Clinton - 1995 NES Pilot

	V2059	V2060	V2061	V2062
V2059 (enthusiastic)	1.000	.705	-.530	-.545

V2060 (hopeful)	.705	1.000	-.615	-.578
V2061 (anxious)	-.530	-.615	1.000	.538
V2062 (upset)	-.545	-.578	.538	1.000

222 observations were used in this computation.
17 cases were omitted due to missing values.

.

Correlation Matrix - Dole - 1995 NES Pilot

	V2064	V2065	V2066	V2067
V2064 (enthusiastic)	1.000	.653	-.387	-.536
V2065 (hopeful)	.653	1.000	-.503	-.680
V2066 (anxious)	-.387	-.503	1.000	.519
V2067 (upset)	-.536	-.680	.519	1.000

214 observations were used in this computation.
25 cases were omitted due to missing values.

The correlation matrixes clearly demonstrate that the correlations, item-to-item, across the PA and NA dimensions are far too high to yield two dimensions and far higher than reported other measurement approaches or using standard NES items. (This inference is confirmed in dimensional analytics.)

Is this a failure of measurement or a failure of theory? It seems to us that three possibilities exist.

First, it may be the case that emotional responses are different when evoked outside of an election campaign. During the summer of 1995 both Clinton and Dole were major figures for the politically attentive, but it many citizens had had only minimal emotional involvement with either one. A lack of distinctive emotional experience might produce an "artificial" single dimensionality when people "manufacture" emotions to fit into a survey response.

Second, the theory may be all wrong. The verbal anchors "enthusiasm" and "anxiety" may reflect only poorly the "basic" discrete emotions that people experience. Of course, it is not obvious why such a disjuncture would produce such high inter-item correlations.

And finally, the measurement scales may at fault. The Form B items are explicitly bipolar in design. That is, we gave respondents contrasting response options that they might interpret as "like-dislike" prompts. (We conjecture here that respondents try to make sense of the interview situation--that they guess that we are trying to figure out whether they approve or disapprove of the candidates. Their doing so would only mean that they consciously adopt the unidimensional view of emotions that has been prevalent--even in Political Science. Our theoretical proposition is that people, unconsciously, use multiple emotions to make sense of the political world.)

To evaluate these possibilities, we may turn to other data. First, consider the Form A emotion items imbedded in the other half sample of the Pilot. By way of contrast with the Form B items, the Form A items evince the dimensions of positive and negative affect. Analyzing the NES traditional discrete items, using the *frequency* format, shows the multidimensionality of mood quite clearly. A principal components analysis of the measures for Clinton, with varimax rotation, reveals three dimensions (eigenvalues of 4.61, 1.82, and 1.24).

Table 4: Form A- Principal Components with Varimax Rotation - 1995 NES Pilot - Clinton

	Factor 1	Factor 2	Factor 3
V2023 (afraid)	.073	-.028	.797
V2024 (anxious)	.108	-.011	.758
V2025 (worried)	.247	-.218	.749
V2026 (angry)	.582	-.273	.501
V2027 (bitter)	.678	-.106	.357
V2028 (resentful)	.639	-.210	.350
V2029 (proud)	-.104	.864	-.044
V2030 (hope)	-.164	.839	-.180
V2031 (enthusiastic)	-.005	.849	-.056
V2032 (disgusted)	.425	-.397	.466
V2033 (hatred)	.782	.032	-.053
V2034 (contempt)	.712	-.060	.086

Afraid, anxious, and worried, clearly define negative affect (Factor 3). Similarly, proud, hopeful and enthusiastic define positive affect (Factor 2). In the third dimension (Factor 1), a general feeling of moral disapprobation is clearly defined by angry, bitter, resentful, hatred and contempt (perhaps also disgusted). We are utterly intrigued by the strong emergence of the third dimension, though it must be said that the discrete model anticipates a differentiation between anger (one of the basic emotions) and disgust (another distinct emotion). These data provide little support for such a distinction. Yet, it does appear some Americans translated negative affect into distinctive "anxiety" and "anger-disgust" components when they considered Clinton in 1995.

The same analysis for the Form A items for Dole is clearer still. Here just two dimensions emerge (eigenvalues of 5.36 and 2.19).

Table 5: Form A - Principal Components with Varimax Rotation - 1995 NES Pilot - Dole

	Factor 1	Factor 2
V2047 (afraid)	.820	-.154
V2048 (anxious)	.704	.013
V2049 (worried)	.855	-.058
V2050 (angry)	.777	-.179
V2051 (bitter)	.726	-.051
V2052 (resentful)	.833	-.071
V2053 (proud)	-.098	.857
V2054 (hope)	-.052	.895
V2055 (enthusiastic)	-.079	.862
V2056 (disgusted)	.798	-.246
V2057 (hatred)	.634	-.033
V2058 (contempt)	.607	.033

For the three discrete negative distinctions to have consequence then items meant to discriminate between the three negative discrete "basic" emotions should define three factors. However, there is no evidence for more than one negative factor in the nine negative items used to assess Dole.

(Note that a journalistic view of politics suggests that Clinton elicited anger or disgust from his opponents when Dole, a more "inside-the-Beltway" figure in 1995, did not. Understanding how negative affect gets turned into anger or disgust is, of course, an intriguing question. At a minimum, these data argue for a strong battery of "emotional response" items in any survey, items that can illuminate subtleties in real-world politics.)

Beyond the matter of dimensionality, we may use over-time correlations to see whether the positive affect and negative affect scales behave as psychological theory predicts. The 1995 Pilot contains an implict panel back to 1994, including the four standard NES affect items for Clinton (proud, hopeful, angry and afraid). Thus, we can compare the test-retest reliabilities obtained by using these four items from 1994 to define two scales for PA and NA, and six items from the 1995 Form A Pilot (proud, hope, enthusiastic for PA and afraid, anxious, and worried for NA).

Based on the results obtained by Watson (described above) we would expect that reliabilities would be in the range of .5 to .7 and that the test-retest reliability of NA would be lower than that of PA. For NA_{94} and NA_{95} r = .49. For PA_{94} and PA_{95} r = .67. These results are quite similar to those obtained by Watson (see Table 2 above). Thus, test-retest conforms with our theoretical expectations.

Demonstrating that we can use the frequency measures of Form A to construct measures of positive and negative affect helps a bit. We now know that the public mood toward the candidates in 1995 was (at least) two-dimensional--there was nothing strange about the period that simplified emotional reactions. And we now know that the Form A items do produce positive and negative mood (and perhaps other affective responses). However, we do not yet know while the Form B items failed. We turn to evidence that will help us make a determination.

In the spring of 1995, we were asked to suggest some measures of candidate affect for an ABC/Washington Post survey. Like the 1995 Pilot, the survey would be a split half design (in the ABC/Washington Post survey, a single favorability question, their standard measure, was used in half the sample, in the other half single measures of enthusiasm and anxiety were used). Like the 1995 NES Pilot, the design produces a random national sample (with an N = 1548). The survey was in the field July 14 - 17, 1995. The ABC/Washington Post survey and the 1995 NES Pilot were administered in relatively close proximity, so much so that we believe we can treat them as a quasi-experiment that tests contrasting question formats.

The affect items used the following format:

"We're interested in how you feel about some of the possible candidates for president next year. How enthusiastic are you about (NAME) - *very enthusiastic*, *fairly enthusiastic*, n*ot too enthusiastic* or *not enthusiastic at all*?"

"And does this person make you feel *very anxious, fairly anxious, not too anxious,* or *not anxious at all*"?

Note that these items are directly derived from the mood arousal theory, but that they are essential unipolar in format. (People are enthusiastic or not, anxious or not.) We can now see whether the theory

travels well.

For Clinton, in the ABC/Washington Post Survey, the enthusiasm and anxiety items are very weakly correlated, r = .14 and for Dole the items are only modestly correlated, r = .26 (compare with the 1995 NES Pilot item-item correlations, Table 3).

Happily, the ABC/ Washington Post survey included five other candidates as prompts, so we can see emotional responses to other types of political figures. The overall pattern of enthusiasm-anxiety item-to-item correlations is consistent with the theory: the two emotional responses are only weakly correlated--as we would expect based on prior research and on the theoretical claim that enthusiasm and anxiety are distinctive arousals.

This contrast of item-to-item correlations between the 1995 Pilot Form B and the ABC/Washington Post suggests that the culprit is not theory but instead measurement. When we juxtapose explicitly distinct affect terms, we lead subjects to "harmonize" their verbalized feelings to a single "like-dislike" dimension. As we have previously shown, reversing the order of presentation of the anxiety terms, asking subjects to rate political candidates as to how safe to anxious they feel, presenting either the term "safe" or "comfortable" first, has a similar effect of reducing the orthogonality of the relationship between the negative affect and positive affect dimensions (Marcus, et al., 1989). In both cases, we suspect that respondents are "helping us out" by putting a unidimensional "like-dislike" order to their expressed emotions. Thus, we conclude that tapping emotional responses requires careful attention to instrumentation.

V. Conclusion and Recommendations

Based on the analysis of data from Form A and Form B from the 1995 NES Pilot and the ABC/Washington Post survey, we can conclude that reliable measures of positive and negative affect can be obtained by using either the standard NES items or the items included in the ABC/Washington Post Survey. It is also the case that these items work with multiple or dichotomous response formats, either as frequency or as intensity measures. Clearly, however, items that make explicit a bipolar frame of reference, such as those used in Form B, fail to provide suitable measures of positive and negative affect. We believe our analyses show that this is a failure of measurement resulting from using items that had not previously been field tested. However, we believe that the ABC/Washington Post Survey items demonstrate that it is possible to design measures that work as expected.

Whether we retain the standard NES items or use items such as those used in the ABC/Washington Post survey, it is of considerable importance that we gain multiple response categories because dichotomous response items force reliance on polychoric and polyserial correlations (which provide less reliable measures of the true item-item covariation than do items that have four or five response options).

We hope these analyses are sufficient to persuade the board of the wisdom of ensuring that the NES series will retain multiple measures of at least PA and NA (at least two each--which can, as we have seen overlap the discrete items). While we have not conducted analyses to assess whether the discrete items produce the distinctive effects that they are purported to have, we have corroborated that measures of positive and negative affect do have the distinctive effects that have previously been shown. And, given the central importance these dimensions of mood have for voter judgment, voter learning and voter attention, we believe that their importance warrants continued inclusion of measures of mood.

References

Bruce, John M. 1991. "Emotion and Evaluation in Nomination Politics." *Washington, DC*.

Marcus, George E. 1988. "The Structure of Emotional Response: 1984 Presidential Candidates." *American Political Science Review*. 82:735-761.

Marcus, George E., Don Kinder, Michael MacKuen, Wendy Rahn and Lisa D'Ambrosio. 1995. "Measuring Emotional Responsese in the 1995 Pilot: A Proposal to the 1995 NES Pilot Planning Committee."

Marcus, George E. and Michael MacKuen. 1993. "Anxiety, Enthusiasm and the Vote: The Emotional Underpinnings of Learning and Involvement during Presidential Campaigns." *American Political Science Review*. 87:688-701.

Marcus, George E., Michael MacKuen and Andrew D. Glassberg. 1989. The Role of Emotional Response in Presidential Campaign Dynamics: Excitement and Threat. Paper presented at the American Political Science Association Meetings.

Watson, David. 1988. "The Vicissitudes of Mood Measurement: Effects of Varying Descriptors, Time Frames, and Response Formats on Measures of Positive and Negative Affect." *Journal of Personality and Social Psychology*. 55:128-141.

Watson, David, Lee Anna Clark and Auke Tellegen. 1988. "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales." *Journal of Personality and Social Psychology*. 54:1063-1070.