ISR

CENTER FOR POLITICAL STUDIES / INSTITUTE FOR SOCIAL RESEARCH / THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48106—1248

TO: Participants, NES Conference on Candidate Evaluation DATE: November 17, 1994

FROM: Vincent Price, Edward Czilli & David Tewksbury, University of Michigan

RE: Candidate Trait Assessments and Affective Responses to Candidates

The paper we are developing for the upcoming conference is still in progress, so we are circulating in its place this memorandum, as requested in the letter of invitation. As the following will make clear, we are still in the midst of our analysis and can offer only the most preliminary of conclusions at this point. Our goal here is to outline our principal lines of inquiry and explain the general drift of our findings to date.

Background and Significance

Our interest centers on the NES character trait assessments and measures of affective reactions to candidates. Prior studies drawing from NES data have identified the importance of character trait assessments in affecting candidate evaluations; these studies have also generated evidence that emotional responses to candidates contribute independently to voters' evaluations. The published evidence supporting these conclusions (e.g., Abelson et al., 1982; Kinder, 1986; Conover and Feldman, 1986; Ottati et al., 1992) is reasonably strong, although it is not without some serious drawbacks. In brief, there are three classes of issues yet to be adequately addressed, which may be listed in roughly increasing order of importance to theory and research in the field. These are: (1) questions about the dimensionality of NES candidate trait and emotional response batteries, especially questions concerning whether traits and emotions are each well described by separate positive and negative dimensions, and whether emotions are less evaluatively consistent along positive-negative lines than are trait assessments; (2) issues surrounding the differential predictive utility of emotions and traits, especially questions concerning whether emotional reactions, as measured by the NES surveys, are more predictive of global candidate evaluations than are trait assessments; and (3) questions concerning the possibility that both the trait and emotion measures may be products of, rather than contributors to, global candidate evaluations -- that is, the possibility that these measures are largely rationalized responses conforming to prior global assessments.

The first two sets of issues have been raised explicitly by the published literature, although neither have to date been conclusively resolved. The third is probably the broadest in its implications and to our knowledge has not yet received any careful attention. Our paper will provide a systematic evaluation of all three sets of questions, capitalizing on available NES resources and employing improved analytical strategies -- confirmatory factor analysis and latent variable structural equation modeling in place of the exploratory factor analysis and standard regression techniques presented in the literature to date.

Measurement issues.

Prior research has been marked by a tendency to evaluate two sets of separate measurement models, one for trait items and another for emotion items, with minimal effort to verify that the chosen models are indeed the best among alternatives -- especially models with less factorial complexity, which might posit that both trait and emotion measures tap a single, global, underlying evaluative dimension. Several of the trait assessments (e.g., "inspiring," "moral") would appear to have strong emotional components, and may thus tap evaluative dimensions also captured by specific items in the emotion battery (e.g, compare "inspiring," from the trait battery, with "proud" from the emotion battery). However, published work has separately analyzed and combined "belief" and "emotion" items, and only then evaluated issues of cognitive and emotional consistency and the like. Thus, even basic issues of discriminant validity have not been particularly well addressed.

Published studies have also compared separately computed factor intercorrelations (or correlations based upon simple counts of positive and negative traits and emotions), noting that positive and negative traits are more strongly inversely related than are positive and negative emotions. Unfortunately, these analyses provide somewhat weak evidence upon which to gauge comparative levels of evaluative consistency. The fact that differences in variances and/or reliabilities of trait and emotion indices cloud empirical comparisons has been clearly acknowledged (e.g, Abelson et al., 1982; Ottati et al., 1992) but not taken into account in the subsequent estimation of interfactor correlations or of parameters in regression models.

Sound empirical answers to questions about the dimensionality of these measures, or about whether assessments of positive and negative candidate traits are more highly correlated than positive and negative emotions, will require more complex confirmatory factor analyses that include *both* the emotion and trait items in the *same* statistical model. For the sake of argument, who is to say that a much simpler model -- for example, one stipulating that all of the emotion and trait items are together just measures of a single, general evaluative dimension -- wouldn't fit the data just as well? If both the emotion and trait measures were subjected together to confirmatory factor analysis, the necessary questions could be directly addressed: Do the emotion and trait items indeed have discriminant validity (or, after accounting for critical differences in measurement, do the trait items load substantially on emotion factors, and vice versa)? Are the emotion and trait batteries each indeed two-dimensional, or would a model with as single belief and a single emotion factor, or for that matter a single positive and single negative factor, fit the data just as well? If there are indeed two emotion factors (positive and negative) and two trait factors (positive and negative), would a model constraining the correlations between the respective positive and negative factors to be equal fit the data just as well as one allowing those correlations to vary? And, finally, if they are found to vary, is the correlation between positive and negative traits indeed larger? The 1980 NES data used by Abelson et al. (1982) could easily be subjected to the analyses described above.

Predictive power of trait appraisals and emotional reactions.

Dimensional issues are fundamental, and questions about the evaluative consistency of trait appraisals and emotional reactions are interesting; but the clearly more central and politically relevant questions concern whether or not emotions operate as partially unique indicators of general attitudes toward political figures. Again, the evidence is fairly strong but not without question. The problems cited above (differential variances and/or reliabilities) might well affect, not just interfactor correlations, but also correlations with the dependent variable (e.g., evaluations of Reagan or Bush) and all of the independent variables (e.g, party identification, emotions, and traits). If so, regression coefficients and incremental R-squares -- the statistics informing the conclusions advanced in prior work -- could also be affected. It is both possible and desirable to estimate regression models using latent variable structural equation methods that, by accounting for measurement error in at least some of the independent variables, would do a far better job of informing us on matters concerning comparative predictive power.

The conceptual nature of the dependent variable selected to represent global candidate evaluation is another important concern. It is worth noting that the dependent measure used in most studies in this area (as elsewhere) is the "feeling thermometer," which requests that respondents report how "warm" or "cold" they feel toward the target. Is this measure not tantamount to a global affective judgment? It is clearly more affective than, for example, the presidential approval questions that ask how "good a job" the president is doing. The emotional/cognitive character of the global evaluation used as the dependent variable should be absolutely central to interpretations of the "differential" or "partially unique" contributions of affective or cognitive components. It would thus be important to test, again using latent variable structural equations, the role of traits and emotions in predicting *both* global affective evaluations (as in the thermometer ratings) and also more "cognitive," less emotional judgments (e.g., presidential job performance). It could well be that emotions are more strongly predictive of the former than of the latter, while quite the opposite pattern might be obtained for the trait measures.

Issues of Causality -- Do Trait Assessments and Affective Reactions Inform Global Evaluations?

Perhaps more critical still is the potential for researchers to incorrectly interpret results of such regression models (whether estimated via latent variable structural equations or not) as necessarily indicative of the calculus used by voters in arriving at their evaluations of candidates. There is certainly the possibility that pre-existing global evaluations determine the more specific candidate-centered responses and not the other way around. In other words, in forming their responses to the emotion and trait questions, respondents may be largely rationalizing their already formed global evaluations of the target person (e.g., the President). Recently, for example, Rahn et al., (1993) have conducted multi-wave analyses of feeling thermometer ratings and versions of the NES candidate likes/dislikes inventories; their findings suggest that much of what is registered in response to candidate likes/dislikes batteries may be the product of rationalized, global candidate evaluations. Rahn et al. propose that more sophisticated and involved voters may be especially like to rationalize their candidate likes and dislikes.

Extrapolating to the candidate trait and emotion measures, we have reasonable cause for concern that our standard dependent variable is in fact a principal cause of our independent variables. Only experimental manipulation of trait appraisals and emotional responses (a daunting exercise) or some form of longitudinal analysis will offer much useful evidence in this regard. Fortunately, NES data resources include collections with feeling thermometers and presidential job performance evaluations, along with short-form candidate traits and emotions batteries available as repeated measures in a number of studies. Although only two- or three-wave repeated measures designs are available, these do allow for the estimation of simple cross lagged, latent variable models that can begin to inform us about the extent to which reports of candidate-centered emotions and trait evaluations indeed contribute to -- or are derived from -- global evaluations. Since it has been suggested that rationalization is especially common among more involved and sophisticated voters, grouped covariance matrices (e.g., for voters of high, average, and low sophistication) can be evaluated for differences, both in the fit of the basic measurement model and in critical structural parameters.

Our paper thus stands to address several commonly cited but unresolved questions about the role of emotions and trait appraisal in shaping candidate evaluations; it will also open new lines of critical inquiry in the area. The work should inform us both about existing NES instrumentation and about the dynamics of candidate appraisal.

Research Design

Although we expect in our final paper to draw from a variety of NES collections, here we will confine our analysis and results to data from the 1980 Major Panel Study. That study presents a particularly valuable resource, in that it included a fairly large number of trait and affect measures (nine character trait questions and seven affective response items) for a number of candidates (e.g., Carter, Reagan, Kennedy), repeated across multiple waves of interviews. It also provides an opportunity to compare results with those of Abelson et al. (1982), who analyzed data from the first wave of the panel.

Face-to-face interviews were conducted over a nine-month period. The first wave occurred in January and February of 1980; the second took place in June and July; and the third wave happened in September and early October of that year. Thus, the first and second waves were separated by three months, while the second and third waves were separated by approximately one and a half months (see Table 1). Although 1008 respondents completed interviews in the first wave, only about three-quarters of those respondents remained in the entire panel (n = 769).

Table 1 about here

Measures

Respondents on all three waves were asked to judge how well a set of character traits applied to several candidates (e.g., "How well does the word 'moral' describe Carter -- extremely well, quite well, not too well, or not well at all?"). The trait assessments included three negative traits ("dishonest," "weak," and "power hungry,") and seven positive traits ("inspiring," "moral," "knowledgeable," "solves our economic problems," "provides strong leadership," and "develops good relations with other countries"). Emotional responses to the candidates were assessed by asking them to report (yes or no) whether a particular figure had ever made them feel a certain emotion (e.g., "Think about Ronald Reagan. Now, has Reagan -- because of the kind of person he is, or because of something he as done -- ever made you feel angry?"). The affective responses included four negative feelings ("angry," "afraid," "disgusted," and "uneasy") and three positive feelings ("hopeful," "proud," and "sympathetic").

Dependent measures include the standard 100-point NES feeling thermometers, with which respondents indicate how warm or cold they feel toward candidates. In the case of Jimmy Carter (our focus here), the thermometer can be supplemented by respondents' ratings of his performance as president, recorded on a 4-point scale ("strongly approve," "approve," "disapprove," and "strongly disapprove").

Data Analysis

The data were analyzed using LISREL 7, a computer program that estimates parameters in latent variable structural equation models (Jöreskog and Sörbom, 1989). The LISREL program generates these estimates by minimizing the differences between a matrix of covariances obtained in a sample and

Table 1. NES 1980 Major Panel Study: Waves 1 to 3

	Wave 1 Jan. 22-Feb. 25	Wave 2 Jun. 4-Jul. 13	Wave 3 Sept 2-Oct. 1
Sample Size	1008	843	769
Completion Rate	75%	84%	90%

covariances predicted by a stipulated theoretical model. Goodness of fit between the hypothesized model and the sample data is assessed by evaluating discrepancies between the covariances predicted by the model and those actually observed in the sample.

Because most of the variables under study here are ordinal measures (save the feeling thermometers, which were treated as continuous), a matrix of tetrachoric, polychoric, and polyserial correlations rather than Pearson correlations or covariances was used in the analysis, and a weighted least-squares (WLS) method of estimation was employed. Computation of all matrices was carried out using a listwise deletion of missing data, resulting in 777 observations in the wave-one analyses and approximately 580 observations in each of the three-wave analyses.[1]

Results

Measurement Models

For the sake of brevity, we will focus on the results for character assessments of, and affective responses to, then-President Carter. Analyses to date suggest that results for Reagan and Kennedy follow essentially the same pattern (see also Abelson et al., 1982). We begin by fitting a measurement model to the full collection of affect and trait items in wave one. Following Abelson et al., (1982), the model we propose posits four latent factors: Positive affect, negative affect, positive traits, and negative traits. The model assumes that no variable loads on more than a single factor, and that all error variances are uncorrelated. Results are presented in Table 2.

Table 2 about here

The obtained χ^2 (here used as a "badness of fit" rather than a goodness of fit measure) is 272.69 with 98 degrees of freedom and is clearly significant ($p < .001$). Previous results with structural equation modeling, however, have indicated that the χ^2 statistic can be large and significant even with good fits, particularly when the covariance matrix under analysis is based upon a very large number of observations (Carmines & McIver, 1981; Hayduk, 1987; Jöreskog and Sörbom, 1989). A ratio of χ^2 to degrees of freedom is sometimes used to judge model fit, and here the ratio is rather large (2.78). The goodness-of-fit index (GFI), which theoretically ranges from 0 to 1 and indicates the relative amount of variance and covariance explained by the model, is .98. A modified version of this index (AGFI = .97) adjusts the goodness-of-fit. Another rough indicator of model fit is the root mean square residual (RMSR = .06), which represents the typical deviation between the observed sample correlations and those predicted by the model. Taken together, these results suggest that the model fits the data reasonably well, although modifications might well improve the overall fit. Results from several sensible alternative models -- e.g., a model adding a fourth factor to capture covariation among negative items, cf. Kinder (1986) -- failed to produce significant improvements. Allowing certain error terms to correlate (in particular "moral," and "dishonest") or allowing some variables to load on more than one factor (e.g., "moral" on the positive affect factor) would improve the fit slightly; but the theoretical motivation for piecemeal modifications seems less than compelling and, at any rate, critical parameter estimates are little affected by these alternate models.

The parameter estimates produced by the model include sizable factor loadings (known as λ estimates in LISREL, averaging above .70), and correspondingly low error variances (θ). The differences in

Table 2. Fitted Factor Model for Carter -- Affects and Traits

Item	Affect Factors: Negative (λ)	Affect Factors: Positive (λ)	Trait Factors: Negative (λ)	Trait Factors: Positive (λ)	Error (θ)
Affects					
Angry	.76*				.42*
Afraid	.56*				.69*
Disgusted	.82*				.33*
Uneasy	.68*				.54*
Hopeful		.78*			.39*
Proud		.85*			.28*
Sympathetic		.53*			.72*
Traits					
Dishonest			.54*		.71*
Weak			.81*		.34*
Power Hungry			.49*		.76*
Moral				.51*	.74*
Knowledgable				.67*	.55*
Inspiring				.77*	.41*
Solve Econ. Prob.				.82*	.33*
Strong Leadershp.				.91*	.17*
Good For. Relat'ns				.73*	.47*

Correlations between Factors (ϕ)

	Pos Aff	Neg Aff	Pos Trt	Neg Trt
Positive Affect	1.00			
Negative Affect	-.46*	1.00		
Positive Traits	.68*	-.69*	1.00	
Negative Traits	-.73*	.77*	-.77*	1.00

Note: Estimates are standardized. N = 777. Model χ^2 = 272.69 with 98 df (p <.001). GFI = .98; AGFI = .97. Root mean square residual = .06.

* $p < .01$

loadings for positive and negative items noted by Kinder (1986) is not apparent, perhaps owing to the fact that the present analysis uses a matrix of polychoric and polyserial correlations and WLS as opposed to ML estimation.

Our major interests focus on the discriminant validity of the four dimensions presented in Table 2. There are several results bearing on the matter. First, a variety of alternative models of less factorial complexity -- positing a single, global evaluative dimension, or only one affect and one trait dimension, or single positive and negative factors that combine affect and trait measures -- were systematically tested, and all fit the data significantly worse (e.g., a model with one affective factor, and one trait factor, even when it allowed all negative items to have second loadings on a third "negativity" factor, produced a χ^2 of 722.05 with 96 degrees of freedom). These results suggest that a four-factor model of the form presented in Table 1 may well be the most appropriate conception of the data.

Second, we should consider the interfactor correlations estimated by the model (ϕ parameter estimates, also presented in Table 2). These are all substantial in size, and markedly larger than those reported by Abelson et al., (1982). The main difference is that the present estimates have been disattenuated for error. Where Abelson et al. estimated the correlation between additive positive and negative affect scales to be -.23, our estimate is -.46; the Pearson correlation between additive positive and negative trait scales was reported to be -.53, compared with our estimate of -.77. We thus replicate -- and underscore -- the Abelson et al. finding that positive and negative affect are less evaluatively constrained than positive and negative trait assessments. An explicit statistical test of the difference is possible, by fitting a model constraining these two interfactor correlations to be equal and comparing the two model χ^2 statistics. In this case, the difference χ^2 is 45.16 with one degree of freedom, clearly significant. So we reject the hypothesis that the correlation between positive and negative affect and between positive and negative trait assessments are equal.

To this point, then, our results confirm the basic measurement model that has generally been applied to the NES affect and trait batteries, supporting prior claims about the discriminant validity of these four dimensions, and in particular the weaker relationship between positive and negative affect.[2]

Predictive Power of the Trait and Affect Measures

As noted above, questions about the evaluative consistency of trait appraisals and emotional reactions are interesting; but the more critical questions concern whether or not emotions and trait assessments operate as partially unique indicators of general attitudes toward political figures. To date, analyses bearing on these questions have relied upon regression analyses using additive scales built from the NES affect and trait batteries (Abelson et al., 1982; Kinder, 1986; Ottati et al., 1992); however, the above results suggest that such scales may result in considerable understatement of the true magnitude of factor intercorrelations and, we might suspect by extension, correlations with dependent variables in regressions as well (e.g., feeling thermometers). Any effort to evaluate the comparative predictive power of traits or affects must contend with the possibility that the differential variance and/or reliability of affect and trait scales might cloud statistical results.

As a first step, respondents' feeling-thermometer ratings of Carter were added (as a directly observed variable) to the basic factor model presented above. For comparison, evaluations of Carter's performance as president were also added (again as a directly observed variable). With all four factors free to correlate separately with the feeling thermometer and job performance ratings, it did appear -- as expected -- that the affective responses correlated more highly with the more "affective" of the two measures (e.g., estimated r between positive affect and the thermometer = .75, compared with .68 for the job performance measure). But difference χ^2 tests comparing models constraining the correlations to be equal to models allowing them to be freely estimated indicated these differences were not significant. As it turns out, the feeling thermometer and job performance measures are so highly correlated (r = .77) that they are essentially redundant measures. For this reason, they were combined as indicators of a fifth latent variable, namely overall evaluations of Carter. Results of the confirmatory factor analysis are presented in Table 3.

Table 3 about here

Goodness-of-fit measures show almost no change, and the factor loadings for all of the trait and affect measures remain unchanged as well. The most relevant pieces of information in Table 3 are the correlations in the last row of the interfactor correlation matrix (Φ). As can be seen, these correlations are uniformly high. In fact, just the positive trait factor alone would account for 72% of the variance in global evaluations of Carter, and the factor most weakly correlated with global evaluations (negative traits) accounts for well over half of the variance.

The fact that all of the trait and affect measures are so highly intercorrelated, and so comparably and highly correlated with global evaluations, renders a regression of global evaluations on all four measures somewhat uninformative -- at least as a way of determining comparative predictive power. A path model in which global evaluations are regressed on all four affect and trait factors is statistically equivalent to the model presented in Table 3 and fits the data identically. The four factors account together for 84% of the variance in global evaluations. The path coefficients vary considerably from one factor to the next (for negative affect, β=-.43; for positive affect, β=.43; for negative traits, β=.01, *n.s.*; and for positive traits, β=.22). But the predictor variables are somewhat interchangeable. For example, a regression model using only negative affect and positive traits fits the data about as well (χ^2 = 300.51 with 127 degrees of freedom (p <.001). GFI = .98; AGFI = .97; RMSR = .06. It accounts for 81% of the variance in global evaluations but produces very different path coefficients (for negative affect, β=-.24; for positive traits, β=.70). The failure of positive traits to produce a large path coefficient in the full regression model is largely due to the fact that it shares so much common variance with the other independent variables.

These findings arouse some suspicion that the trait and affect measures may be, in effect, differently measured global candidate evaluations rather than especially useful independent variables in explanatory models. The measures do seem to tap four distinct dimensions, as our preceding confirmatory factor analyses indicated. But as a practical matter, all four measures relate so similarly to general evaluations of Carter that it is difficult to make a case for one or another of them as independent contributors to candidate evaluation. Another illustration of this can be found in the pattern of near identical relationships between all five latent factors -- including general evaluation -- and party identification (for negative affect the correlation with party ID is -.32; for positive affect, r = .27; negative trait, r = -.32; positive trait, r = .35; global evaluation, r = .38).

Issues of Causality -- Overtime Analyses

It may well be, then, that the NES measures of character trait assessments and of affective responses to candidates are as much a product as a contributor to general candidate evaluations. In an effort to gain

Table 3. Fitted Factor Model for Carter -- Affects, Traits, and Global Evaluation

Item	Affect Factors: Negative (λ)	Affect Factors: Positive (λ)	Trait Factors: Negative (λ)	Trait Factors: Positive (λ)	Global Evaluation (λ)	Error (θ)
Affects						
Angry	.76*					.42*
Afraid	.55*					.69*
Disgusted	.83*					.33*
Uneasy	.67*					.54*
Hopeful		.77*				.39*
Proud		.85*				.28*
Sympathetic		.54*				.72*
Traits						
Dishonest			.54*			.71*
Weak			.80*			.34*
Power Hungry			.50*			.76*
Moral				.51*		.74*
Knowledgable				.67*		.55*
Inspiring				.78*		.41*
Solve Econ. Prob.				.82*		.33*
Strong Leadershp.				.90*		.17*
Good For. Relat'ns				.72*		.47*
Evaluations						
Job Performance					.90*	.19*
Feeling Thermometer					.90*	.19*

Correlations between Factors (ϕ)

	Pos Aff	Neg Aff	Pos Trt	Neg Trt	Eval
Positive Affect	1.00				
Negative Affect	-.46*	1.00			
Positive Traits	.68*	-.69*	1.00		
Negative Traits	-.72*	.77*	-.77*	1.00	
Evaluation	.78*	-.78*	.85*	-.75*	1.00

Note: Estimates are standardized. $N = 777$. Model $\chi^2 = 285.38$ with 125 df ($p < .001$). GFI = .98; AGFI = .98. Root mean square residual = .06.

* $p < .01$

some additional leverage on this issue, we turn to longitudinal data analysis. Our three waves of data should allow us, in principle, to estimate simple cross-lagged, latent variable models or models with synchronous reciprocal paths between trait and affect factors and general candidate evaluations.

Because of panel attrition and the loss of cases due to listwise deletion of missing data, our three-wave models provide us with a much smaller sample (less than 600). Because asymptotic variances and covariances of estimated correlations require a minimum number of cases -- at least $1.5k(k + 1)$, where k is the number of variables -- it was decided that only three indicators of each latent construct would be selected for use in the longitudinal analyses. This provision would allow at least the modeling of each latent trait and affect factor, separately coupled with latent global evaluations, over three waves. Model estimates (e.g., factor loadings, errors, and factor intercorrelations) generated by the smaller three-wave subset of variables were compared with those generated by the wave-one analyses of the full complement of variables. In general, estimates were quite similar.

Before attempting to model overtime relationships between affect and trait factors and general candidate evaluations, simple autoregressive models were estimated separately for each latent variable. Basic simplex models without correlated errors over time were first evaluated. Only in the case of global evaluations (indicated by job performance ratings and feeling thermometer scores) did such a model fit well ($\chi^2 = 1.31$ with 7 degrees of freedom, $p = .99$). For each of the trait and affect models, allowing errors in observed variables to correlate (e.g., allowing the error for "angry" at wave one to correlate with the errors for "angry" at waves two and three, and so on) produced very significant improvements in model fit. Results are presented in Table 4. As is clear from the model χ^2 statistics, each of the models fits the data quite well. The most striking feature of these results is the very high levels of stability for all of the measures, even over the three-month period separating waves one and two. For the one-and-a-half-month period separating waves 2 and 3, stability coefficients average over .90, meaning that all but about 19% of the variance in these factors is accounted for by measures taken at the previous wave. These estimates exceed even the high levels of stability that Kinder (1986) found for character trait assessments over a one-month period (which he reported as ranging from .79 to .84). As might be expected, the present WLS estimates are somewhat higher, having been disattenuated for error.

Table 4 about here

The fact that the data series are so stable introduces some problems for the estimation of cross-lagged or synchronous reciprocal paths between global evaluations and each of the trait and affect factors. Such models generally resulted in estimated paths with coefficients reversed in direction, produced negative error variances, or estimated stability coefficients over 1.0. In fact in each case, when one of the trait or affect variables was paired in an analysis with global evaluations over time, the model that fit the data most successfully was one in which the two variables a) are correlated at the outset, b) are completely autoregressive (that is, dependent only upon measures of the same latent construct at the preceding wave), and c) have modestly correlated residual variances at waves two and three. In other words, these models posit that traits, affects, and global evaluations began the nine-month study period highly correlated and remained that way without mutual effects for the full course. Overall measures of fit for these models were as follows: For negative affect and global Carter evaluations, ($\chi^2 = 46.27$ with 74 degrees of freedom, $p = .99$; RMSR= .03); for positive affect and Carter evaluations, ($\chi^2 = 59.19$ with 74 degrees of freedom, $p = .89$; RMSR= .04); for negative traits and evaluations, ($\chi^2 = 181.51$

Table 4. Estimated Overtime Stabilites -- Affects, Traits, Global Evaluation

Latent Factor	Wave 1 - 2	Wave 2 - 3	χ^2	*df*	*p*	*N*
Negative Affect	.85*	.89*	15.95	16	.46	591[a]
Positive Affect	.85*	.95*	15.23	16	.51	591
Negative Traits	.74*	.89*	17.15	16	.38	553
Positive Traits	.82*	.85*	20.21	16	.21	550
Global Evaluation	.82*	.93*	1.31	7	.99	591

Note: Estimates are standardized. For each of the trait and affect models, errors for each variable (θ) were allowed to correlate across waves. In the global evaluation model, errors for each variable were assumed uncorrelated across waves.

a. In order retain enough cases for the overtime analyses, only three indicators of each latent affect or trait factor were retained. They were as follows: negative affect ("angry," "disgusted," "uneasy"); positive affect ("hopeful," "proud," "sympathetic"); negative traits ("dishonest," "weak," "power-hungry"); and postive traits ("knowledgeable," "solve economic problems," "provide strong leadership").

* $p < .01$

with 74 degrees of freedom, $p < .01$; RMSR= .06); for positive traits and evaluations, ($\chi^2 = 65.93$ with 74 degrees of freedom, $p = .74$; RMSR= .03).

Preliminary Conclusions and Ongoing Work

We are continuing to evaluate overtime models, and to estimate second-order factor analyses (see note 2), with an eye toward sorting out the most accurate way to understand the NES trait and affect batteries. But our work to date suggests strongly that these measures, when modelled carefully as indicators of latent variables, tend to mirror quite closely those measures traditionally treated as indicators of global candidate evaluation (e.g., feeling thermometers and presidential job performance ratings). They shar similarly large correlations with general candidate evaluations, are just as stable as those global evaluations, and correlate with third variables (e.g., party ID) just as do more general evaluations.

As noted at the outset, our conclusions at this point can be little more than tentative. Still, we suspect that our results, should they be follow the same general pattern as we turn to other candidates and NES data collections, will raise some serious questions about the meaning and potential uses of these measures. To put things in the briefest possible way, it appears that these measures may not properly belong on the right-hand side of an equation seeking to explain general candidate evaluations.

The main thrust of our findings is that we may do well to conceive of these measures less as independent variables that *contribute* to general candidate evaluations, and more as *alternative indicators* of those general evaluations. The idea that they serve as "partially unique predictors" of candidate evaluation should be reconsidered in light of the high interfactor correlations we have observed, and especially in view of the fact that they seem to relate quite uniformly to global candidate evaluations (and party identification). This is not to say that all four factors provide strictly identical information, but that the differences appear subtle at best.

How then, should we use these measures? As dependent variables, perhaps coupled with more conventional measures of candidate evaluation (e.g., thermometers) in analyses such as those illustrated here, they can provide a slightly more multidimensional view of those evaluations. Taken as a whole, they might well comprise a set of "subtests" of general candidate evaluation, and in so doing serve to increase construct validity. We might conjecture (with Kinder, 1980) that separate affective and character-trait components are more or less responsive to shifts and changes in the political context. On the other hand, if they are as stable as they appear from these analyses, what opportunities can be found to observe reasonable change in these measures? Perhaps for some segments of the population (e.g., the less knowledgeable), we might find far less (or more) unidimensionality, and greater responsiveness to changing context, than we have so far observed. These are topics we are now pursuing and would like to pose for discussion next month.

[I write this, somewhat anxiously, as the moment of "no reimbursement for airfare" draws near. With apologies for being somewhat brief (and perhaps too cryptic), I look forward to meeting and working with you in Berkeley. -- VP]

Notes

1. Listwise deletion of missing values was required for computation of asymptotic covariance matrices necessary for calculating WLS weighting factors. In general, listwise deletion of missing values produces a correlation matrix with a number of desirable properties (Bollen, 1989), but it does in this instance eliminate a substantial number of cases. Furthermore, in calculating matrices across three waves, some variables were eliminated in order to maintain a sample size sufficiently large in relation to the number of variables examined (Jöreskog and Sörbom, 1989; see details below).

2. There remains, however, the possibility that each of the four factors are yet products of a single, more global evaluative dimension (i.e., general like or dislike of Carter). One way to test this possibility is to conduct a second-order factor analysis, in which each of the factors depicted in Table 1 are hypothesized to load on another latent, second-order factor. In this model, the four latent factors -- positive affect, negative affect, positive traits, and negative traits -- are posited to be unique products of another, overarching latent factor we might conceptualize as a global evaluation of Carter. In specifying the model, all of the same first-order factor loadings are stipulated for the observed variables; but the correlations between those first-order factors are hypothesized to stem from their condition of being unique products of a joint underlying cause. Such a model fits the data about as well as the basic four-factor model. The obtained χ^2 is 287.99 with 99 degrees of freedom (as before, clearly significant ($p < .001$). The ratio is just slightly larger (2.9 as compared with 2.78). The goodness-of-fit indices (GFI and AGFI are identical, as is the root mean square residual (.06). Loadings on the second-order global evaluation factor were quite substantial, as follows: negative affect, -.71; positive affect, .77; negative traits, -.79; positive traits, .99.

References

Abelson, R. P., Kinder, D. R., Peters, M. D., Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology, 42,* 619-630.

Carmines, E. G., and J. P. McIver (1981). Analyzing models with observed variables: Analysis of covariance structures. In G. W. Borhnstedt & Borgatta (Eds.), *Social measurement*, pp. 65-115. Beverly Hills, CA: Sage.

Conover, P. J., & Feldman, S. (1986). Emotional responses to the economy: I'm mad as hell and I'm not going to take it anymore. *American Journal of Political Science, 30,* 50-78.

Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and advances.* Baltimore: Johns Hopkins University Press.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications*. 2nd ed. Chicago: SPSS Inc.

Kinder, D. R. (1986). Presidential character revisited. In R. R. Lau and D. O. Sears (Eds.), *Political cognition: The 19th Annual Carnegie Symposium on Cognition* (pp. 233-256). Hillsdale, NJ: Lawrence Erlbaum.

Ottatti, V. C., Steenbergen, M. R., & Riggle, E. (1992). The cognitive and affective components of political attitudes: Measuring the determinants of candidate evaluations. *Political Behavior, 14,* 423-442.

Rahn, W. M., Krosnick, J. A., & Breuning, M. (1993). Rationalization and derivation processes in political candidate evaluation. Paper presented to the annual meetings of the American Association for Public Opinion Research, St. Charles, Illinois.