

**BEHAVIORAL COMPLEXITY IN LEADERSHIP:
THE PSYCHOMETRIC PROPERTIES OF A NEW INSTRUMENT**

Katherine A. Lawrence
University of Michigan Business School
701 Tappan Street
Ann Arbor, MI 48109-1234
Tel: 734-994-7904
Fax: 734-994-7935
E-mail: kathla@umich.edu

Robert E. Quinn
University of Michigan Business School
701 Tappan Street
Ann Arbor, MI 48109-1234
Tel: 734-615-4179
Fax: 734-615-4266
E-mail: requinn@umich.edu

Peter Lenk
University of Michigan Business School
701 Tappan Street
Ann Arbor, MI 48109-1234
Tel: 734-936-2619
Fax: 734-936-0274
E-mail: plenk@umich.edu

Submitted to *Organization Science* on 11/21/03

The authors wish to thank Rick Bagozzi, Gordon Cheung, Jeff Edwards, Kevin Kuan, Gretchen Spreitzer, Lynda St. Clair, and Tom Sy for their help and feedback on this paper.

BEHAVIORAL COMPLEXITY IN LEADERSHIP: THE PSYCHOMETRIC PROPERTIES OF A NEW INSTRUMENT

Abstract

Recent research in managerial and executive leadership has recognized the importance of cognitive complexity and behavioral complexity. In particular, behavioral complexity addresses the competing demands and roles expected of managers and leaders. Though some empirical research on behavioral complexity exists, progress has been hindered by the absence of a more rigorous measurement instrument. Here we design an elaborated, multi-dimensional instrument based on the Competing Values Framework (CVF). We test this second-order measurement model using structural equation modeling (SEM) to remove measurement error and to examine the underlying conceptual structure of the framework. In addition, we test the spatial relationship of the factors using a circumplex model. Our data largely supports the theoretical structure and stringent demands of the CVF model as applied to this measurement instrument. This new instrument should prove to be a valuable tool for exploring the behavioral complexity of leaders.

BEHAVIORAL COMPLEXITY IN LEADERSHIP: THE PSYCHOMETRIC PROPERTIES OF A NEW INSTRUMENT

The measurement of effectiveness in managerial and executive leadership has been a long-term interest in the organizational sciences (Yukl 1998). A recent, promising line of inquiry in this area focuses on the concept of behavioral complexity, but no clean, psychometrically sound instrument for measuring behavioral complexity exists. The purpose of this paper is to develop an instrument to better measure behavioral complexity.

Complexity and Leadership

In an extensive review of the literature on executive leadership, Zaccaro (2001) identifies four conceptual models that focus on the requisite qualities of executive leaders. The four models are conceptual complexity models, behavioral complexity models, strategic decision-making models, and visionary or inspirational models. The latter two are well recognized. The former two have been less visible. Conceptual complexity models focus on the level of information processing demands facing executives and their requisite need for high-level conceptual skills. Behavioral complexity focuses on the level of pressing social demands and the ability to play multiple roles that call for diverse or even competing behaviors.

Zaccaro (2001) identifies three prior approaches that advance our thinking on the behaviorally complex roles played by executive leaders. First, Mintzberg (1973) calls attention to the fact that managers are not careful planners but rather under continuous pressure to act. He suggests a typology of managerial roles that must be played by all managers. He does not, however, see the roles as competing. Second, Tsui (1984) focuses on leader success as a function of reputational effectiveness and the ability of executives to meet and balance role expectations. In developing her approach she expands our understanding to the complexity of the process but does not put emphasis on conflicting expectations. Third, Quinn and his colleagues also identify

diverse, multiple demands but argue that these demands inherently compete with each other and that effective leaders must master competing patterns of behavior (Denison et al. 1995, Hart and Quinn 1993, Hooijberg 1996, Hooijberg and Quinn 1992, Quinn 1984). Zaccaro (2001) notes that there is considerable value in this line of research because it establishes the linkage between cognitive complexity and social behavior and that the linkage can give rise to an integrated model of leadership. One purpose of this paper is to advance the research on behavioral complexity by increasing our capacity to measure it. Next we will briefly review the research specifically focused on behavioral complexity in response to competing demands.

Defining Behavioral Complexity within the Competing Values Framework

In its simplest form, behavioral complexity is the capacity of a given leader to engage in wide repertoire of behaviors (Hooijberg and Quinn 1992). A person with high behavioral complexity is able to engage in a wider array of behaviors than a person with low behavioral complexity is able to do. Specifically, behavioral complexity is “the ability to exhibit contrary or opposing behaviors (as appropriate or necessary) while still retaining some measure of integrity, credibility, and direction” (Denison et al. 1995: 526). These opposing behaviors may be organized using the dimensions of the Competing Values Framework (CVF) (Quinn 1984). Specifically, two pairs of contrasting values or capabilities define the behavioral breadth with which a manager might act. The first pair concerns organizational focus, either an internal focus or an external one. The second pair reflects an emphasis of stability or flexibility in the organizational structure. As we illustrate in Table 1, these dimensions produce a four quadrant model of managerial behaviors. One set of opposing quadrants contrasts behaviors that create continuity versus change. The other set contrasts priorities of results versus relationships. Because opposing quadrants are commonly thought to be mutually exclusive, their coexistence in an individual manager or leader presents a paradox. The CVF model provides a useful

foundation for the development of an instrument measuring behavioral complexity because this model emphasizes the inherent difficulty of addressing competing demands and gives a theoretical structure to the requisite behaviors for satisfying those demands.

Insert Table 1 about here

Behavioral complexity suggests that it is possible for a leader to transcend these paradoxes. A behaviorally complex leader both maintains continuity and leads change. For example, managers are expected to do such things as maintain control over projects, make sure employees understand policies and procedures, and insure the accuracy of the work, all of which promote and maintain organizational stability. A manager is also expected to do such things as anticipate the evolving needs of customers, initiate new programs, and inspire people to engage in new patterns of behavior, all of which support a need for change. Likewise, a behaviorally complex leader transcends the paradox of results versus relationships. Managers are expected to do such things as encourage hard work, emphasize speed, and focus on the competition in order to produce results. At the same time, there is an expectation that a manager should maintain relationships by showing concerns for the needs of others, encouraging participation, and assisting others in their career development. Thus, a behaviorally complex leader should be capable of meeting the multiple and competing needs of the organization (Tsui 1984).

Behavioral Complexity and Effectiveness

A basic theoretical proposition is that a person who has high behavioral complexity is a person with the requisite variety (Ashby 1952) to meet and integrate the paradoxical pressures of organizational life (Denison et al. 1995). While behavioral complexity does not guarantee that a manager will exercise perfect judgment in applying the right behaviors at the right time, it does enable a manager to draw on a wider repertoire of behaviors to effectively meet competing

demands. Two categories of evidence tend to support the proposed relationship between behavioral complexity and leadership effectiveness. First, there is direct evidence from studies designed to test the proposition. In a study of chief executive officers and firm performance, Hart and Quinn (1993) found that CEOs with high behavioral complexity were more effective than their counterparts. The firms run by more behaviorally complex CEOs outperformed other firms on three measures of financial performance. Also, studies of middle managers have demonstrated that people with high behavioral complexity are more effective leaders than are those with lower levels of behavioral complexity (Denison et al. 1995, Hooijberg 1996, Quinn et al. 1992).

In his recent review of the leadership literature, Zaccaro (2001) concludes that senior leaders are more effective when they are behaviorally complex. Although research indicates growing support for a leadership model that incorporates behavioral complexity, Zaccaro also indicates that the body of literature specifying and measuring such a model is still quite small. While some research has assessed behavioral complexity using different methodological approaches, little has been done to demonstrate the psychometric properties of such an assessment (e.g., convergent and discriminant validity, relationships between factors) (Zaccaro 2001, pp. 167-168). Thus, there is a need for a more reliable and valid measurement instrument.

The Measurement of Competing Values

While research on complexity and effectiveness is promising, it has been hindered by the lack of an efficient and rigorous measurement instrument to measure leadership behavior. The instrument used by Hart and Quinn (1993) has only one factor per quadrant. This provides a very limited measure of complexity. The instrument used by Denison, Hooijberg and Quinn (1995), has a small number of items representing each domain. This makes it difficult to separate random measurement error from unique variance specific to the measure (Bagozzi 1994). The instrument used in Cameron and Quinn (1999) displays considerable error variance, so it requires a

transformation to ipsative data in order to remove the error variance and reproduce the CVF structure. To date, no other instruments have been developed. Therefore, an instrument that will better measure behavioral complexity is needed.

However, creating such an instrument is difficult. The theoretical structure of the CVF imposes more rigorous demands on an instrument than does any other leadership theory. Consider any theory that identifies a set of twelve proposed behaviors. To establish discriminant validity in a confirmatory factor analysis, a researcher must distinguish just twelve first-order factors. Using the CVF, the same twelve factors would be partitioned into four quadrants; therefore, we must identify twelve first-order factors as well as demonstrate a second-order structure in which the factors represent the four predicted domains. Moreover, the behaviorally complex leader would be expected to have abilities in all four quadrants, so the four second-order factors should be correlated yet distinct.

The purpose of this paper is to develop an efficient (i.e., compact) instrument with appropriate reliability and multiple indicators in each quadrant. The factors must display both a first- and second-order structure that meets the requirements of the theory. In addition, the instrument should sufficiently represent the multi-dimensional space described by the theory. An instrument that meets these demands should enhance research on behavioral complexity. We will focus on four areas: the accuracy of psychometric tests, the specific detail of the instrument, the correlation of quadrants, and the spatial relationship between the instrument items.

Accuracy of Psychometric Tests

Only Kalliath, Bluedorn, and Gillespie (1999), working on the model at the cultural level, have used structural equations modeling (SEM) to test the CVF. They suggest that the advantage of SEM over other validation tools is that measurement error can be eliminated, allowing the parameters to be specified more accurately. Their results supported the underlying structure of

the CVF; however, it was measuring organizational culture, not managerial behavior. To date, no study has tested the framework as applied to individual managers by using SEM. This paper will develop and test a CVF instrument for leadership behavior using SEM, and we will also use standard tests of reliability and validity.

Instrument Specificity

Many papers that test the CVF use a limited number of indicators per quadrant. Most only test the four main constructs, using a few indicators per quadrant to measure each construct (Denison et al. 1995, Kalliath et al. 1999, Quinn and Spreitzer 1991). As we asserted earlier, a second-order model that uses multiple indicators to measure a set of first-order constructs is more useful for capturing latent constructs while remaining accurate enough to identify specific behaviors. This paper will describe and test a model that uses specific behavioral indicators to identify first-order constructs that have a specified relationship to the second-order constructs.

Correlation of Quadrants

Because the theoretical model is structured so that each quadrant shares a value dimension with each of its neighbors, we would expect that constructs representing a given quadrant would be somewhat correlated with neighboring constructs but not at all correlated with the constructs in the opposite quadrant of the framework. While the dimensions and roles of the CVF have been tested using multitrait-multimethod techniques, multidimensional scaling, exploratory factor analysis, and general psychometric analyses, the nature of such tests has rarely allowed for the likelihood that managers will demonstrate behaviors in all four models, focusing instead on discriminant validity between factors or items and convergent validity within factors. Instead, we would expect that empirical tests of behaviorally complex individuals would produce lower correlations between diagonally opposite quadrants.

Working on the model at the cultural level, Kalliath, Bluedorn, and Gillespie (1999) tested the hypotheses that neighboring quadrants sharing one of the dimensions of the CVF would have statistically significant, moderate correlations, while opposing quadrants of the framework would have correlations that were zero or close to zero. They found that the adjacent quadrants did, in fact, have significant correlations, however, they also found a strong correlation between the open systems and internal process models. They explain this finding as an artifact of the sample they used, which was undergoing change. This paradox is in keeping with our assertions in this paper. In contrast, we will test our measurement instrument to verify that neighboring quadrants are more highly correlated than opposing ones, but we will also examine whether the instrument fits more stringent, theoretical constraints of zero (or close to zero) correlations between opposing quadrants.

Spatial Relationships

We feel that it is important to establish the spatial relationship of the factors in each quadrant because the CVF hypothesizes an underlying spatial model that should take the form of a circle. To assess this, we use a circumplex model. Circumplex models are used in psychology to describe similarity and polarity among values (Anderson 1960, Browne 1992, Guttman 1954). After an appropriate sequencing of the items, the inter-item correlations progress from large positive, to zero, to large negative, back to zero, and back to large positive as one moves away from the diagonal of the correlation matrix. Thus, the circumplex model imposes the most demanding constraints on the CVF. The existence of circumplex correlations is consistent with the hypothesis of two, independent, bipolar dimensions, which measures the subject's relative position on orthogonal scales, when scale-usage effects are removed. In this case, the CVF could be considered to have two, bipolar dimensions, one representing the opposing quadrants of "People" and "Results" and the other representing "Processes" and "Change." Each first-order

factor should be associated with one of these dimensions. In a prior study of behavioral complexity, the data had a poor fit with a circumplex model (Denison et al. 1995). A strong fit would suggest a more powerful model for measuring behavioral complexity.

In summary, this paper attempts to fill the gap in the behavioral complexity literature by developing and testing an instrument that uses multiple indicators per behavior in a second-order model. Additionally, the use of SEM eliminates measurement error and allows us to examine the correlations between the quadrants as a reflection of the underlying conceptual structure of the framework. Finally, this paper tests this instrument against a spatial, circumplex model.

Methods

Sample and Data Collection Procedure

The sample used for construct validation included mid- and senior-level managers from an international information services organization. The survey was administered to 539 managers prior to their participation in an executive education course. Five-hundred twenty-eight of the respondents answered a 72-question survey (described below). Seventy-nine percent of the respondents were male, and eighty percent were between 31 and 45 years old (with a modal age of 36 to 40 years of age). These managers worked in locations around the globe (approximately half were based in Europe, the Middle East, or Africa, and slightly less than one-third were based in the Americas) but no racial demographics were collected. The sample size was further reduced from the original sample size because respondents were allowed to select “Don’t know” as an option (creating “missing” data) and listwise deletion was used for generating the matrices used in the factor analyses, resulting in a final sample size of 407 subjects.

These managers were also evaluated by subordinates using the same instrument. A total of 1610 subordinates were surveyed, producing an average of three subordinate evaluations per manager, but some respondents had as many as eleven subordinate evaluations. These

subordinates were 66 percent male, 25 percent were younger than 31 years old, and 64 percent were between 31 and 45 years old (with a modal age of 31 to 35 years of age).

Development of the Measures

The organization at which the respondents worked was interested in using an assessment that would be structured using the CVF. Interviews with 24 senior executives in the firm helped identify sets of behaviors that they thought were important. We matched these behaviors with the appropriate quadrants in the model, according to whether the behaviors represented continuity, change, relationships, or results. Ultimately, six sets of behaviors per quadrant were selected to represent a wide array of behaviors. (See Table 2 for a complete list of these behaviors.)

Insert Table 2 about here

A separate scale of three questions was used to measure each of the behaviors, resulting in a 72-item instrument. The wording of the questions was grounded in the interviews with the senior executives of the company. For the self-evaluation, each question was framed as “I would describe myself as being skilled in the following...,” and the surveys for the subordinates were phrased in a similar way. The questions were administered with a 5-point Likert-type scale (strongly disagree to strongly agree), plus an option for co-workers to indicate that they did not have sufficient information (coded as missing data). Questions were randomly ordered.

Analytic Procedures

The objective of the analyses was to select a set of questions that could be used to assess managerial behavior in each of the quadrants. To that end, we wanted the number of questions representing each quadrant to be manageable and practical. Because we expected the quadrants to correlate, we felt it would be cleaner to refine the sets of questions within each quadrant first and then verify that the quadrants were appropriately related. Consequently, we used a series of

exploratory analyses to determine the most effective set of constructs to represent each quadrant, and then we confirmed that the constructs appropriately fit within the higher-order framework. For the exploratory analyses, we used the mean of the subordinate evaluations for each respondent, and for the confirmatory analyses, we used the respondents' self evaluations as a way of cross-validating the suitability of the final model (Cudeck and Browne 1983).

Refinement of the measurement model. First, we tested each set of measures for reliability using Cronbach's alpha (Cronbach 1951). A reliability estimate of 0.7 or better is considered to be good (Nunnally 1978). We also tested the distributions of the 72 items to check for range restriction and multivariate normality.

Second, we conducted four confirmatory factor analyses (CFAs) to explore the fit of and factor loadings in the first-order models of each quadrant. Because the CVF has a strong theoretical structure, the use of CFAs to examine and refine the measurement instrument is appropriate (Hurley et al. 1997). Each quadrant was examined individually to see which of the six first-order factors and sets of measures would best represent the second-order construct. The instrument was refined and reduced by examining the factor loadings and correlations, paying attention to the theoretical robustness and the content of the items that might cause problems (Anderson and Gerbing 1988). We wanted indicators that measured several distinct constructs and that did not tend to load on multiple factors.

Confirmation of the measurement model. The confirmatory analyses used the self evaluations. Once we reduced the model size, we conducted an exploratory factor analysis (EFA) of the items to verify that the items did not load on unintended factors (i.e., discriminant validity). Next, we conducted a second-order CFA to examine the entire reduced model using the scales as first-order factors in four second-order quadrants. We did this in two ways. First, we tested that the individual items loaded appropriately on each first-order scale factor, the first-

order factors appropriately loaded on the second-order factors, and the correlations between second-order factors fit the expected pattern of highly correlated neighboring quadrants and less highly correlated opposing quadrants. Second, we wanted to test the model in a way that would be comparable with the constraints imposed by the Bayesian circumplex model investigation (see discussion in next section). To do this, we added an additional first-order factor that would represent the subject-specific scale-usage effects that the circumplex model incorporates. All items loaded on this factor.

We used LISREL 8.30 with maximum likelihood estimation for all CFA tests. Several criteria were used to assess the overall model fit of each analysis. The overall chi-square statistic, which evaluates the deviation of the reproduced covariance matrix from the sample covariance matrix, is subject to influence by the sample size, leading models with good fit to be rejected when sample sizes are large (Gerbing and Anderson 1992). Consequently, we used several additional criteria. The root mean squared error of approximation (RMSEA) estimates the discrepancy per degree of freedom between the original and reproduced covariance matrices. Browne and Cudeck (1993) recommend a value of up to .05 as a close fit, and values up to .08 are reasonable. The 90% confidence interval should be within these values, such that the value should not be significantly greater than .05. We also used the Tucker-Lewis Index (TLI; also called the Non-Normed Fit Index or NNFI in the LISREL program) (Tucker and Lewis 1973), and the Comparative Fit Index (CFI) (Bentler 1990) as other goodness-of-fit measures that are independent of sample size. Both of these indexes represent the relative improvement in fit of the tested model as compared to the baseline or “null” model. Potential values range from 0 to 1, and a value of 1 indicates a perfect fit to the population. Despite ongoing debates on the subject, values of .90 or higher are considered to indicate adequate fit (Bentler and Bonett 1980). In addition, the NNFI rewards models that are less complex. Finally, we used the Standardized

Root Mean Square Residual (SRMR), which is the square root of the average squared residuals between the original and reproduced covariance matrices. A cutoff value of .08 for the SRMR is recommended by Hu and Bentler (1999).

Circumplex model. To examine the spatial relationships of the items and scales in the model, we tested the circumplex pattern using a Bayesian circumplex model. All circumplex model analyses used the self-evaluation data. Browne (1992) proposes using trigonometric functions to approximate the circumplex correlations. Lenk, Wedel, and Bockenholt (Lenk et al. 2002) implement a Bayesian version of Browne’s model. Their model accommodates subject-specific, scale-usage effects with ordinal data and incorporates theory-driven constraints on the relation among sets of items. In their paper, they apply the Bayesian circumplex model to the Schwartz and Bilsky instrument (1987, 1990) and confirm two bipolar dimensions for the values measured by the instrument: Self-Transcendence versus Self-Enhancement and Openness-to-Change versus Conservation. This Schwartz and Bilsky instrument has a similar structure to the CVF, suggesting that the application of the Bayesian circumplex model would be appropriate.

Rossi, Gilula, and Allenby (2001) provide a model for scale-usage effects (what could also be considered response bias) and document the bias when these effects are ignored. The cut-point model for ordinal data assumes that the ordinal response arises when an underlying continuous, latent variable falls between two cut points. In this application, each subject has his or her unique set of cut-points. The cut point model offers two benefits. First, it explicitly recognizes the ordinal nature of the data and transforms it to a continuous scale that is more appropriate for most statistical models. Second, the manifest ordinal variables can be highly non-normal, even multi-modal, but the latent variables are normal.

The latent variable for the circumplex model follows a constrained, three-factor model:

$$U_{i,j} = \varphi_i + \mu_j + \alpha_i \cos(\theta_j) + \beta_i \sin(\theta_j) + \varepsilon_{i,j}$$

$U_{i,j}$ is subject's i latent rating for item j where φ_i is a scale usage effect for subject i ; μ_j is the grand mean for item j ; a_i and b_i are factor scores for subject i ; $\cos(\theta_j)$ and $\sin(\theta_j)$ are factor loadings for item j ; and $\varepsilon_{i,j}$ are normally distributed, independent error terms. Because the factor loadings are constrained to the unit circle, the model uses a polar coordinate representation where the angles θ_j are between 0 and 2π . The orientation of the circle is determined by assuming without loss of generality that $\theta_1 = 0$ and $0 < \theta_2 < \pi$.

The scale-usage effect φ_i for subject i is a random effect from a normal distribution with mean 0 and standard deviation τ . If φ_i is positive, the subject tends to use the upper end of the scale, and if it is negative, he or she tends to use the lower end of the scale. We call this factor “scale-usage” even though it could also be a measure of basic managerial competency. For example, one manager might have a modal self-assessment score of 3 and another manager's score might be 4. The point difference could be attributable to subject-specific scale usage or to an accurate appraisal of leadership ability; consequently, scale-usage bias and general effectiveness are confounded within the data of this study.

The factor scores $\{\alpha_i, \beta_i\}$ are mutually independent and normally distributed with mean 0. If the two factor scores have the same standard deviation λ , then the correlations fall within a circumplex model. The factor loadings $\{\sin(\theta_j), \cos(\theta_j)\}$ are expressed in polar coordinates to restrict them to the unit circle. The error terms $\{\varepsilon_{i,j}\}$ are mutually independent and normally distributed with mean 0 and standard deviation σ_j for item j .

With these assumptions, the variances and covariances of the latent scores are:

$$\begin{aligned} \text{var}(U_{i,j}) &= \tau^2 + \lambda^2 + \sigma_j^2 \\ \text{cov}(U_{i,j}, U_{i,j'}) &= \tau^2 + \lambda^2 \cos(\theta_j - \theta_{j'}), j \neq j' \end{aligned}$$

The covariance between two items is determined by the difference in the factor loading angles. As this difference varies from 0 to π to 2π , the cosine ranges from +1 to -1 to +1. The scale-usage effect allows the correlations to be positive if $\tau^2 > \lambda^2$. Not including scale-usage effects can severely bias the resulting representation of the item angles.

Thus, within a circumplex model, we wanted to examine whether two, opposing quadrants could be considered to represent a bipolar dimension; therefore, one dimension represented the “Change” and “Processes” opposition and the other represented the “People” and “Results” opposition. We fit two circumplex models, treating the two bipolar dimensions as orthogonal. That is, knowing a respondent’s position on the “Change” and “Processes” dimension is not informative about his or her position on the “People” and “Results” dimension. The first model did not constrain the order of the item angles; they were allowed to vary freely around the circle. The second model imposed constraints on the angles so that the items would be expected to fall within their associated quadrant; all of the angles for items in “Change” preceded those in “People,” which preceded those for “Processes” and “Results.” In this way, we hoped to verify the theoretical validity of the CVF.

Results

Initial Refinement of the 72-Item Model

Descriptive statistics. Using the subordinate evaluations we assessed the reliability, distribution, and the factor structure of the instrument. The Cronbach alpha reliability coefficients for the initial 24 scales averaged .82 and ranged from .71 to .93. All scales met the standard for reliability of .70, recommended by Nunnally (1978). The distribution of the items was not range restricted for any of the 72 items, with responses across the entire range. The questions did vary significantly from multivariate normality for skewness ($Z = -8.75$, $p < .05$) but not for kurtosis ($Z = 1.00$, $p = .32$); however, variables with statistically significant skewness do not

produce substantive differences in the analysis when the sample size is large, as it was in this case (Tabachnick and Fidell 1996). A visual examination of the shape of the distribution showed that almost all the variables had a slight bias towards higher scores (i.e., negative skewness) and the distribution was slightly peaked (i.e., positive kurtosis), but they did not look very different from the normal distribution curve.

CFA exploration. We used first-order CFAs to explore the fit of the preliminary model by quadrant and reduce the number of items. Table 3 reports the results from these explorations. Each model produced a statistically significant chi-square, which was expected based on the large sample size. The CFI and TLI values for each quadrant, however, were all above the .90 criterion, indicating excellent fit. All but one of the models had an RMSEA that met the recommended criterion, and all SRMR values were less than .08.

Insert Table 3 about here

Item coefficients for each quadrant model were good, ranging from .56 to .95, and correlations between first-order factors in each quadrant ranged from .14 to .82, though most were between .35 and .70. All item coefficients and correlations were statistically significant. We wanted the model to include items which were distinct so as to provide discriminant validity yet correlated so that they appropriately represented each second-order factor. Although some factors were highly correlated, tests of discriminant validity checking that the confidence interval around the correlation estimate between the two factors did not include 1.0 (Anderson and Gerbing 1988) indicated that the factors were distinct. However, we selected first-order factors from each quadrant that were moderately correlated (ranging from .42 to .66). We eliminated factors with indicators that seemed to be less distinct from other factors or that had higher variation in factor loadings (Anderson and Gerbing 1988). Each of the factors and indicators that

we eliminated had wording which, in retrospect, was less accurately representative of the associated CVF construct. Moreover, they were less evocative of the theoretical dimensions of the CVF framework (i.e. internal/external focus and flexible/stable structure). Two items were reassigned to replace an indicator in another factor because they seemed to have a better fit theoretically and statistically. Based on these analyses, we were able to reduce the number of behavioral constructs representing each quadrant from six to three (twelve scales in total), measured with 36 items.

The final 36-item model included three scales in each quadrant to provide a representative range of complex behaviors. The “People” quadrant measured “encouraging participation,” “developing people,” and “acknowledging personal needs.” The “Change” quadrant included “anticipating customer needs,” “initiating significant change,” and “inspiring people to exceed expectations.” The “Processes” quadrant focused on “clarifying policies,” “expecting accurate work,” and “controlling projects.” Finally, the “Results” quadrant assessed “focusing on competition,” “showing a hard work ethic,” and “emphasizing speed.” The exact wording of each of the items, grouped by construct, are listed in the Appendix.

Exploratory factor analysis (EFA). Before conducting further analyses using the self evaluations, we checked the 36 measures for multivariate normality. The distribution of the items was not range restricted for 30 of the 36 items, with responses across the entire range. The remaining 6 items were restricted on the low end of the scale (no respondents selected “strongly disagree”). On average, the questions did vary significantly from multivariate normality for skewness ($Z = -5.10, p < .05$) but not for kurtosis ($Z = 1.53, p = .13$); again, we did not expect this to produce substantive differences in the analysis because the sample size was large (Tabachnick and Fidell 1996). Like the subordinate evaluations, almost all the variables had a slight bias

towards higher scores (i.e., negative skewness) and the distribution was slightly peaked (i.e., positive kurtosis), but they did not look very different from the normal distribution curve.

With the 36 items, we conducted a maximum likelihood factor analysis, using varimax rotation and extracting 12 factors. Because varimax rotation assumes that the factors are not correlated, this analysis would also indicate whether a second-order model with correlated factors would be appropriate. In other words, if the varimax rotation analysis did not show distinct loadings, it would suggest that factors are correlated. The first ten factors had eigenvalues greater than one, accounting for 67 percent of the variance. All 12 factors explained 72 percent of the variance. Inspection of factor loadings showed that, although each factor corresponded to a particular set of items, other items also loaded on each factor. Next, we used oblimin rotation (with principal component analysis as the extraction method), allowing the factors to correlate. The results of this analysis showed very good simple structure. Looking at the pattern matrix, which indicates the partial correlation between the factors and items (removing the influence of the other factors), items loaded greater than .5 on their primary factor and less than .3 on the majority of all other factors. All but three of the factor loadings were above .6, the remainder of loadings in any particular factor was typically below .2 (only two items loaded higher than .3 on the wrong scale), and for most factors the difference between the highest set of loadings and the next highest loading was usually at least .45. A few loadings had a smaller difference. Generally, these overlaps can be explained by common words or sentiments in the scales (e.g., being concerned with people). Moreover, these overlaps are not surprising given that correlations are expected within quadrants and between neighboring second-order factors. Thus, the results of the factor analysis support the discriminant validity of most of the scales.

Testing the Reduced Model

We assessed the reduced, 36-item model with a second-order CFA. Table 3 and Figure 1 report these results.

Insert Figure 1 about here

Basic model fit, loadings, and correlations. First, we assessed the 36-item, second-order model as hypothesized in the CVF. The model produced a statistically significant chi-square, which was expected based on the large sample size. The CFI and TLI values for the entire model were .91 and .90, indicating a very good fit for a complex, second-order model. The model had an excellent RMSEA of .046 (not significantly different from the .05 criterion). The SRMR of .070 was also excellent. Standardized coefficients from items to factors were excellent, ranging from .52 to .87. Coefficients from first-order factors to second-order factors ranged from .55 to .85, though most were between .64 and .80. This variability was somewhat a problem, indicating that one 1st-order construct was perhaps less well suited than the others. All coefficients were statistically significant.

Correlations between the second-order factors were as expected in an empirical test of the CVF. Factors in adjacent quadrants were more highly correlated than the factors in opposite quadrants. The correlation between the “People” and “Processes” quadrants was somewhat lower (at .46) than the other correlations between neighboring quadrants (which were .65, .68, and .69). Correlations between factors in opposite quadrants were lower than the other correlations: the “People” and “Results” quadrants had a correlation of .42 and the “Change” and “Processes” quadrants had a correlation of .20. All correlations were statistically significant.

Model with scale-usage factor. Second, we assessed the 36-item, second-order model, adding the subject-specific, scale-usage factor. Table 3 reports these results. The model fit

improved slightly. The CFI value for the entire model was .94 and the TLI value was .92. The model had an RMSEA of .040 and a SRMR of .054, which also indicated an excellent fit. Standardized coefficients from items to factors dropped somewhat. Three items had factor loadings of .35, .35, and .45, but the remainder ranged from .54 to .88. Coefficients from first-order factors to second-order factors ranged from .53 to .84, with the exception of the loading of the first scale of the “Processes” quadrant (scale 7), which dropped to .13. The loadings of the items onto the scale-usage factor were quite low (all but six were below .45), and ten were statistically non-significant. All other loadings were statistically significant.

Correlations between the second-order factors were almost as hypothesized. Factors in adjacent quadrants were more highly correlated (.71 to .74) than in the other 36-item model, with the exception of the “People” and “Processes” quadrants, which were correlated by .09. Correlations between factors in opposite quadrants were low: the “People” and “Results” quadrants had a correlation of .38 and the “Change” and “Processes” quadrants had a correlation of .24. All correlations were statistically significant, except the exceptionally low correlation between “People” and “Processes.” These results were nearly in keeping with the expected relationships between the quadrants of the CVF model, with exception of the non-significant relationship between “People” and “Processes.”

This second analysis shows that the model with a scale-usage factor is robust. In comparison to the 36-item model without the scale-usage factor, this second model has weaker coefficients, suggesting that the model without the scale-usage factor would be preferred. Nevertheless, including the scale-usage factor in further analyses would not be inappropriate.

Reliability and univariate statistics. We tested the reliability of all the scales again, particularly to check the scales that were changed. Reliability coefficients are in Table 2. Cronbach’s alpha for the scales that had items replaced were not as good as we had hoped, but

were close to the .70 criterion, with reliabilities of .68 and .69. One other scale had a Cronbach's alpha of .69, and all other scales had excellent reliability coefficients. Given that the sample of subordinate evaluations produced much higher coefficients for almost every scale, we hope that these scores are an artifact of this particular sample of managers. The means and standard deviations of the final scales did not vary widely: scale means ranged from 3.58 to 4.24, and scale standard deviations ranged from .54 to .75. Scale correlations ranged from 0 to .46. Correlations between scales within the same quadrant averaged 0.37 while correlations between scales in different quadrants averaged 0.20.

Investigation of the Circumplex Model

Table 4 reports summary parameters for the posterior distributions of the angles. The posterior mean is estimated by using Markov chain Monte Carlo with the algorithm in Lenk, Wedel, and Bockenholt (2002). The posterior standard deviation measures the uncertainty for the angles (analogous to the standard error in non-Bayesian analyses). Given the posterior standard deviations, there is considerable overlap in the posterior distributions for the angles under the two models, if one keeps in mind that 0 and 2π (or 6.28) are the same value on the circle, so items 1c and items 2a through 2c overlap with items 10a through 10c.

Insert Table 4 about here

A Bayesian model selection criterion is the log marginal distribution (LMD) of the data under the two models. This criterion incorporates both the fit of the model to the data and the parameter uncertainty. The Bayesian Information Criterion (BIC) is the large sample approximation of the LMD, while the LMD is the exact computation. The LMD is -16761 for the unconstrained model and is -16732 for the model with constraints. Thus, the second model is slightly preferred to the first (larger values indicate a better fit). The fit of the two models are

similar, but the constraints reduce the posterior uncertainty about the angles. In Table 4, the posterior standard deviations for the constrained model tend to be smaller than the unconstrained model. The Brier scores, a squared error loss measure for ordinal data, are small, 1.429% and 1.427%, for the models.

Figure 2 plots the posterior means for these angles. The orientations for the two models are very similar, except that two of the “Results” items overlap with “Change” (as indicated in Table 4). Item angles within the same CVF quadrant differ by no more than $\pi/2$, which indicates that items within a quadrant are positively correlated. (The closer the item angles, the higher the correlation.) Item angles for “Change” oppose those for “Processes,” which indicates a negative correlation between these items. Similarly, “People” opposes “Results.” Conversely, an item angle from the “Change/Processes” dimension is nearly orthogonal to an item angle from the “People/Results” dimension. Except for three items in “Change” that are positioned just under 2π without constraints, the angles for the two models are highly correlated. This discrepancy between the two models is minor because the cosine is a periodic function on the circle: $\cos(\theta) = \cos(2\pi - \theta)$. That is, this apparent difference in the positioning of these angles does not substantially affect the correlations among the angles.

Insert Figures 2 and 3 about here

Figure 3 plots the averages of the angles within each quadrant for the model with constraints. Moving around the circle, the figure illustrates how neighboring quadrants are positively related to each other. For example, “Initiate significant change” and “Anticipate customer needs,” which both belong to the “Change” quadrant, are close to “Focus on competition,” which belongs to the “Results” quadrant. Similarly, in order to lead change, one

must “Inspire people,” which, in turn, is highly correlated with “Encourage participation” in the “People” quadrant. Conversely, “Inspire people” directly opposes “Expect accurate work,” which belongs to the “Processes” quadrant. Similarly, “Acknowledges personal needs” is in conflict with “Emphasize speed.”

Discussion

This study provides further support for the concept of behavioral complexity represented by the CVF. It adds to the literature that uses other methodological approaches and measures to examine the structure of the CVF. Specifically, this is the first study to use SEM as a way of testing an instrument for assessing managerial leadership, as opposed to culture or effectiveness. Moreover, our analysis allowed us to examine the factor structure of the framework while allowing for the expected correlations between neighboring quadrants in the model. Our instrument uses a second-order factor structure that allows each construct to be represented by multiple measures. Finally, this study locates the behavioral factors around a circumplex model.

Our results supported the four-quadrant structure of the CVF as applied to managerial behavior. Neighboring quadrants are significantly and moderately to highly correlated. Opposing quadrants were correlated significantly but to a lesser degree than neighboring quadrant correlations. The strong correlations between neighboring quadrants suggest that, in general, managers are reasonably competent in at least three areas while the moderate correlations across opposing quadrants additionally suggest that managers may actually be competent on all dimensions, albeit weaker in the quadrant opposite their strongest competency.

Also, our results suggest that the CVF is well represented as a second-order model. Even though it can be difficult to obtain a satisfactory fit with second-order models (Bagozzi 1994), our fit indices indicated that our data fit the theoretical model. The data mostly supported both

convergent and discriminant validity for the measurement model. With one exception, the data supported our hypothesized model.

Finally, our tests of the circumplex model provided important additional insights. When we imposed bipolar constraints on the model, it produced an even better fit than the unconstrained model. The scales are located in the appropriate quadrants where they would be expected. Furthermore, the scales within a quadrant are located closer to quadrants to which they are more closely related. For example, as shown in Figure 4, the “Controlling projects” scale in the “Processes” quadrant is closest to the “Results” quadrant, not the “People” quadrant on the other side. This ordering adds to the face validity and intuitive appeal of the model.

The only exception to our theoretical model was that the two internally-focused quadrants did not correlate as highly as we would have expected. In our first analysis of the 36-item model, these quadrants did have a slightly higher correlation than either of the diagonally-opposed quadrants, but the correlation was still low compared to the other adjacent quadrants. In our analysis that included a factor for scale-usage or response-bias, this correlation became non-significant. Given that each of these quadrants were significantly correlated with their other neighboring quadrants, this suggests that the dimension of “internal focus” might not have been effectively operationalized. It may be that the “Processes” quadrant was too focused on the stable structure dimension and “People” was too focused on flexibility, resulting in less emphasis on their common dimension. On the other hand, the low correlation could also be an artifact of our sample such that managers high in one or another of these profiles were not as high in the other.

Limitations

Although this study has developed a more robust instrument for measuring behavioral complexity, it has a few limitations. First, the sample represents managers in only one firm, which may limit generalizability. Similarly, approximately three-quarters of the respondents are

male. This disproportionate number of males may have influenced the distribution of ratings, offering another possible explanation for the unexpectedly low correlation between the quadrants representing internal focus. In the future, the generalizability of our findings should be tested with other samples, including more females and perhaps gathering data on racial demographics. Second, the fact that we used listwise deletion for the correlation matrices meant that we eliminated anyone who responded “don’t know” to one or more survey questions used in a particular analysis. These evaluators may be different from those retained in the sample in ways that we cannot know for certain.

Conclusion

In conclusion, we have deepened our understanding of the CVF as it applies to leadership and behavioral complexity. This instrument should help scholars capture the dynamic and contradictory nature of leadership behaviors. By examining a second-order factor structure with an elaborated measurement instrument, we have been able to unpack the behaviors and constructs that underlie the paradoxical nature of behavioral complexity and see how they are related to each other. Our results support the idea that many managers are moderately competent in three related areas of behavior, and more rarely, some are competent in all four. The versatility allowed by behavioral complexity should help leaders respond more effectively to varied challenges. To further our understanding of leader effectiveness, this instrument may be used in future research that combines behavioral complexity with models of conceptual complexity, strategic decision-making, and visionary leadership. Such studies might identify and compare sets of leaders with behavioral complexity and those without, allowing scholars to better understand how effective leaders maximize the benefits of a diverse set of behavioral strengths.

References

- Anderson, J. C., D. W. Gerbing. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psych. Bull.* **103** 411-423.
- Anderson, T. W. 1960. Some stochastic process models for intelligence test scores. K. J. Arrow, S. Karlin, P. Suppes, eds., *Mathematical methods in the social sciences*. Stanford University Press, Stanford, CA, 205-220.
- Ashby, W. R. 1952. *Design for a brain*. Wiley, New York.
- Bagozzi, R. P. 1994. Structural equation models in marketing research: Basic principles. R. P. Bagozzi, ed., *Principles of marketing research*. Blackwell, Oxford, UK, 317-385.
- Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psych. Bull.* **107** 238-246.
- Bentler, P. M., D. G. Bonett. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psych. Bull.* **88** 588-606.
- Browne, M. 1992. Circumplex models for correlation matrices. *Psychometrika.* **57** 469-497.
- Browne, M. W., R. Cudeck. 1993. Alternative ways of assessing model fit. K. A. Bollen, J. S. Long, eds., *Testing structural equation models*. Sage, Newbury Park, CA, 136-162.
- Cameron, K. S., R. E. Quinn. 1999. *Diagnosing and changing organizational culture*. Addison-Wesley, Reading, MA.
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika.* **16** 297-334.
- Cudeck, R., M. W. Browne. 1983. Cross-validation of covariance structures. *Multivariate Behavioral Res.* **18** 147-167.
- Denison, D. R., R. Hooijberg, R. E. Quinn. 1995. Paradox and performance: Toward a theory of behavioral complexity in managerial leadership. *Organ. Sci.* **6** (5) 524-540.
- Gerbing, D. W., J. C. Anderson. 1992. Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Res.* **21** 132-160.

- Guttman, L. 1954. A new approach to factor analysis: The radex. P. F. Lazarsfeld, ed., *Mathematical thinking in the social sciences*. Columbia University Press, New York, 258-348.
- Hart, S. L., R. E. Quinn. 1993. Roles executives play: CEOs, behavioral complexity, and firm performance. *Human Relations*. **46** (5) 543-574.
- Hooijberg, R. 1996. A multidirectional approach toward leadership: An extension of the concept of behavioral complexity. *Human Relations*. **49** (7) 917-947.
- Hooijberg, R., R. E. Quinn. 1992. Behavioral complexity and the development of effective managers. R. L. Phillips, J. G. Hunt, eds., *Strategic management: A multiorganizational-level perspective*. Quorum, New York.
- Hu, L., P. M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. **6** 1-55.
- Hurley, A. E., T. A. Scandura, C. A. Schriesheim, M. T. Brannick, A. Seers, R. J. Vandenberg, L. J. Williams. 1997. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *J. Organ. Behavior*. **18** (6) 667-683.
- Kalliath, T. J., A. C. Bluedorn, D. F. Gillespie. 1999. A confirmatory factor analysis of the competing values instrument. *Educ. Psych. Measurement*. **59** (1) 143-158.
- Lenk, P., M. Wedel, U. Bockenholt. 2002. Bayesian estimation of random effects circumplex models subject to prior theory constraints and person-specific usage bias. Unpublished Working Paper, University of Michigan Business School, Ann Arbor, MI.
- Mintzberg, H. 1973. *The nature of managerial work*. HarperCollins, New York.
- Nunnally, J. C. 1978. *Psychometric theory*. McGraw-Hill, New York.
- Quinn, R. E. 1984. Applying the competing values approach to leadership: Toward an integrative framework. J. G. Hunt, D.-M. Hosking, C. A. Schriesheim, R. Stewart, eds., *Leaders and*

- managers: International perspectives on managerial behavior and leadership*. Pergamon Press, New York, 10-27.
- Quinn, R. E., G. M. Spreitzer. 1991. The psychometrics of the competing values culture instrument and an analysis of the impact of organizational culture on quality of life. *Research in Organizational Change and Development*. **5** 115-142.
- Quinn, R. E., G. M. Spreitzer, S. L. Hart. 1992. Challenging the assumptions of bipolarity: Interpenetration and managerial effectiveness. S. Srivastva, R. E. Fry, eds., *Executive and organizational continuity: Managing the paradoxes of stability and change*. Jossey-Bass, San Francisco.
- Rossi, P. E., Z. Gilula, G. M. Allenby. 2001. Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statistical Assoc.* **96** 20-31.
- Schwartz, S. H., W. Bilsky. 1987. Toward a psychological structure of human values. *J. Personality and Soc. Psych.* **53** 550-562.
- 1990. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *J. Personality and Soc. Psych.* **58** 878-891.
- Tabachnick, B. G., L. S. Fidell. 1996. *Using multivariate statistics*. 3rd ed. HarperCollins, New York.
- Tsui, A. S. 1984. A role set analysis of managerial reputation. *Organ. Behavior and Human Performance*. **34** 64-96.
- Tucker, L. R., C. Lewis. 1973. The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. **38** 1-10.
- Yukl, G. 1998. *Leadership in organizations*. 4th ed. Prentice-Hall, Saddle River, NJ.
- Zaccaro, S. J. 2001. *The nature of executive leadership: A conceptual and empirical analysis of success*. American Psychological Association, Washington, DC.

Table 1:
Behavioral Complexity in the Competing Values Framework

| | Internal Focus | External Focus |
|---------------------------|---|--|
| Flexible Structure | Relationships <i>“Relating to People”</i> | Change <i>“Leading Change”</i> |
| Stable Structure | Continuity <i>“Managing Processes”</i> | Results <i>“Producing Results”</i> |

Table 2: Reliability Coefficients of Behavioral Scales in the Original 72-Item Model

| Original | Final | Descriptive Scale Name | Reliability with Subordinate Ratings | Reliability with Self Ratings |
|---------------------------|-------|--|--------------------------------------|-------------------------------|
| Relating to People | | | | |
| 1 | 1 | Encouraging participation | .83 | .69 |
| 2 | | Communicating concern (item moved to final scale #3) | .88 | |
| 3 | 2 | Developing people | .84 | .72 |
| 4 | 3 | Acknowledging personal needs (item added from original scale #2) | .80 | .68 |
| 5 | | Building a diverse unit | .79 | |
| 6 | | Representing employee concerns | .79 | |
| Leading Change | | | | |
| 7 | 4 | Anticipating customer needs | .76 | .75 |
| 8 | | Developing a strategic vision | .87 | |
| 9 | 5 | Initiating significant change | .86 | .83 |
| 10 | | Persisting in the face of adversity | .71 | |
| 11 | | Convincing senior people to support new ideas | .81 | |
| 12 | 6 | Inspiring people to exceed expectations | .84 | .78 |
| Managing Processes | | | | |
| 13 | 7 | Clarifying policies | .87 | .86 |
| 14 | 8 | Expecting accurate work | .86 | .80 |
| 15 | | Providing consistent discipline | .82 | |
| 16 | | Stabilizing internal processes | .79 | |
| 17 | 9 | Controlling projects | .84 | .86 |
| 18 | | Making decisions carefully | .75 | |
| Producing Results | | | | |
| 19 | 10 | Focusing on competition | .82 | .81 |
| 20 | | Focusing on immediate financial results | .74 | |
| 21 | 11 | Showing a hard work ethic | .86 | .81 |
| 22 | 12 | Emphasizing speed (item added from original scale #23) | .79 | .69 |
| 23 | | Resolving problems quickly (item moved to final scale #12) | .84 | |
| 24 | | Running efficient meetings | .93 | |

Table 3: Results of Confirmatory Factor Analyses

| Model | # Items | # 1st-order Factors | # 2nd-order Factors | χ^2 | <i>df</i> | RMSEA | CFI | TLI | SRMR |
|--|----------------|----------------------------|----------------------------|-----------|-----------|--------------|------------|------------|-------------|
| <i>First-order: 72-item preliminary analysis (subordinate evaluations)</i> | | | | | | | | | |
| Relating to people | 18 | 6 | 0 | 331.31** | 120 | .062** | .954 | .941 | .048 |
| Leading change | 18 | 6 | 0 | 229.30** | 120 | .044 | .976 | .969 | .032 |
| Managing processes | 18 | 6 | 0 | 232.93** | 120 | .046 | .973 | .966 | .035 |
| Producing results | 18 | 6 | 0 | 274.61** | 120 | .057 | .964 | .954 | .057 |
| <i>Second-order: 36-item reduced (self evaluations)</i> | | | | | | | | | |
| Full model | 36 | 12 | 4 | 1105.56** | 576 | .046 | .909 | .900 | .070 |
| Full model with scale usage factor | 36 | 13 | 4 | 917.61** | 540 | .040 | .935 | .924 | .054 |

Note: RMSEA = root mean squared error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis Index, SRMR = standardized root mean squared residual. Test for RMSEA is the deviation from .05, representing close fit; models were not statistically different from this criterion, unless noted.

**p<.01

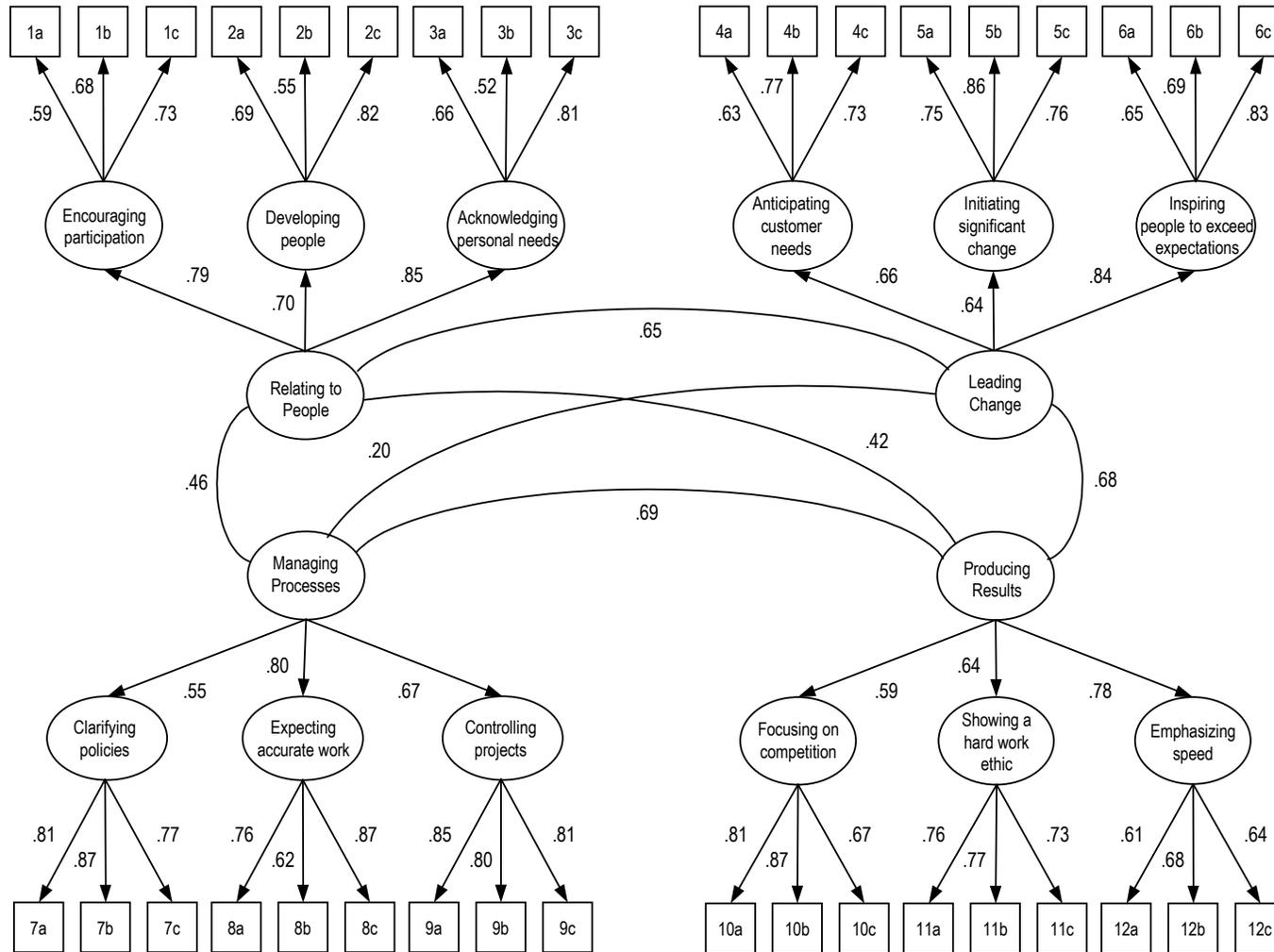
Table 4: Estimated Angles without Constraints and with Constraints

| Item | Posterior Mean ^a | | Posterior Standard Deviation ^b | |
|------|-----------------------------|-------------------------|---|-------------------------|
| | <i>No Constraints</i> | <i>With Constraints</i> | <i>No Constraints</i> | <i>With Constraints</i> |
| 1a | 0.000 | 0.000 | 0.000 | 0.000 |
| 1b | 0.101 | 0.192 | 0.083 | 0.119 |
| 1c | 5.902 | 0.018 | 0.211 | 0.101 |
| 2a | 5.878 | 0.061 | 0.229 | 0.115 |
| 2b | 5.785 | 0.010 | 0.235 | 0.101 |
| 2c | 5.984 | 0.136 | 0.240 | 0.130 |
| 3a | 0.521 | 0.953 | 0.229 | 0.154 |
| 3b | 0.760 | 1.197 | 0.230 | 0.155 |
| 3c | 0.604 | 1.043 | 0.222 | 0.132 |
| 4a | 1.107 | 1.570 | 0.238 | 0.157 |
| 4b | 1.354 | 1.809 | 0.239 | 0.153 |
| 4c | 1.111 | 1.572 | 0.238 | 0.155 |
| 5a | 1.556 | 2.017 | 0.249 | 0.157 |
| 5b | 1.974 | 2.442 | 0.264 | 0.184 |
| 5c | 1.580 | 2.041 | 0.245 | 0.156 |
| 6a | 1.654 | 2.122 | 0.251 | 0.171 |
| 6b | 1.634 | 2.081 | 0.239 | 0.159 |
| 6c | 1.436 | 1.884 | 0.231 | 0.148 |
| 7a | 2.925 | 3.400 | 0.260 | 0.175 |
| 7b | 2.789 | 3.264 | 0.254 | 0.167 |
| 7c | 2.785 | 3.252 | 0.242 | 0.149 |
| 8a | 3.590 | 4.036 | 0.239 | 0.160 |
| 8b | 3.670 | 4.121 | 0.256 | 0.181 |
| 8c | 3.575 | 4.019 | 0.231 | 0.145 |
| 9a | 3.847 | 4.297 | 0.243 | 0.150 |
| 9b | 3.758 | 4.219 | 0.249 | 0.157 |
| 9c | 3.845 | 4.291 | 0.234 | 0.147 |
| 10a | 5.623 | 5.980 | 0.248 | 0.145 |
| 10b | 5.727 | 6.093 | 0.230 | 0.114 |
| 10c | 5.911 | 6.171 | 0.233 | 0.083 |
| 11a | 4.390 | 4.829 | 0.240 | 0.151 |
| 11b | 4.500 | 4.939 | 0.239 | 0.146 |
| 11c | 4.430 | 4.874 | 0.260 | 0.166 |
| 12a | 4.596 | 5.052 | 0.258 | 0.157 |
| 12b | 4.832 | 5.281 | 0.243 | 0.140 |
| 12c | 4.593 | 5.054 | 0.259 | 0.154 |

^aPosterior means are the estimates of the angles.

^bPosterior standard deviations are a measure of spread for the posterior distribution and are analogous to standard errors.

Figure 1: Results of Second-Order Confirmatory Factor Analysis on 36-Item Model



Note: Parameters shown are for standardized model. All paths and correlations are statistically significant.

Figure 2: Posterior Means of the Item Angles for 36-item Model

A is without constraints on the angles, and B is with dimension constraints.

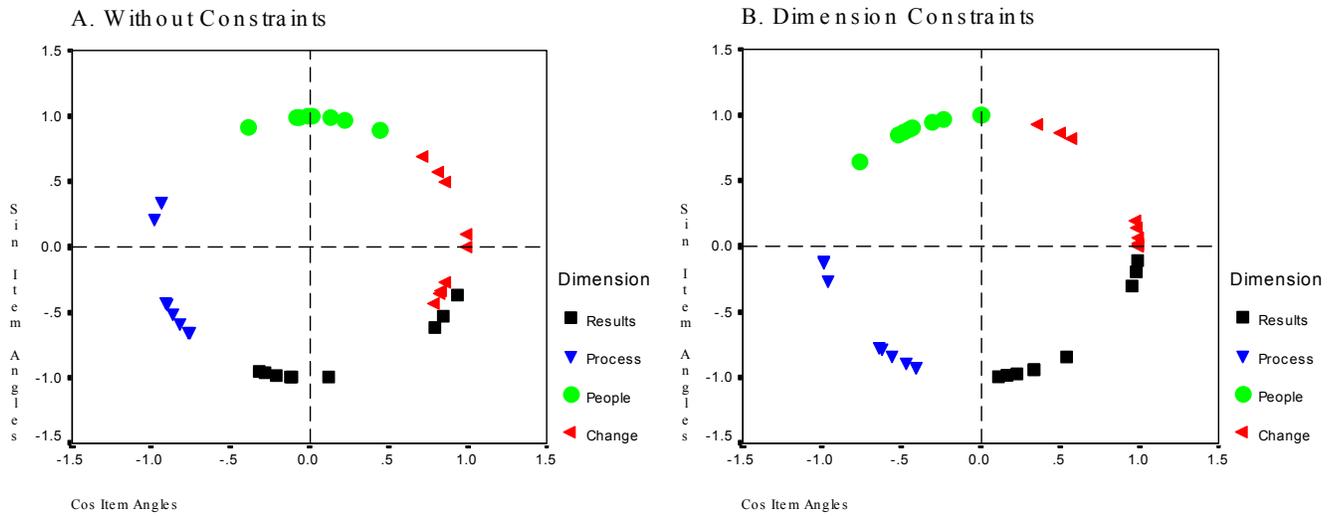
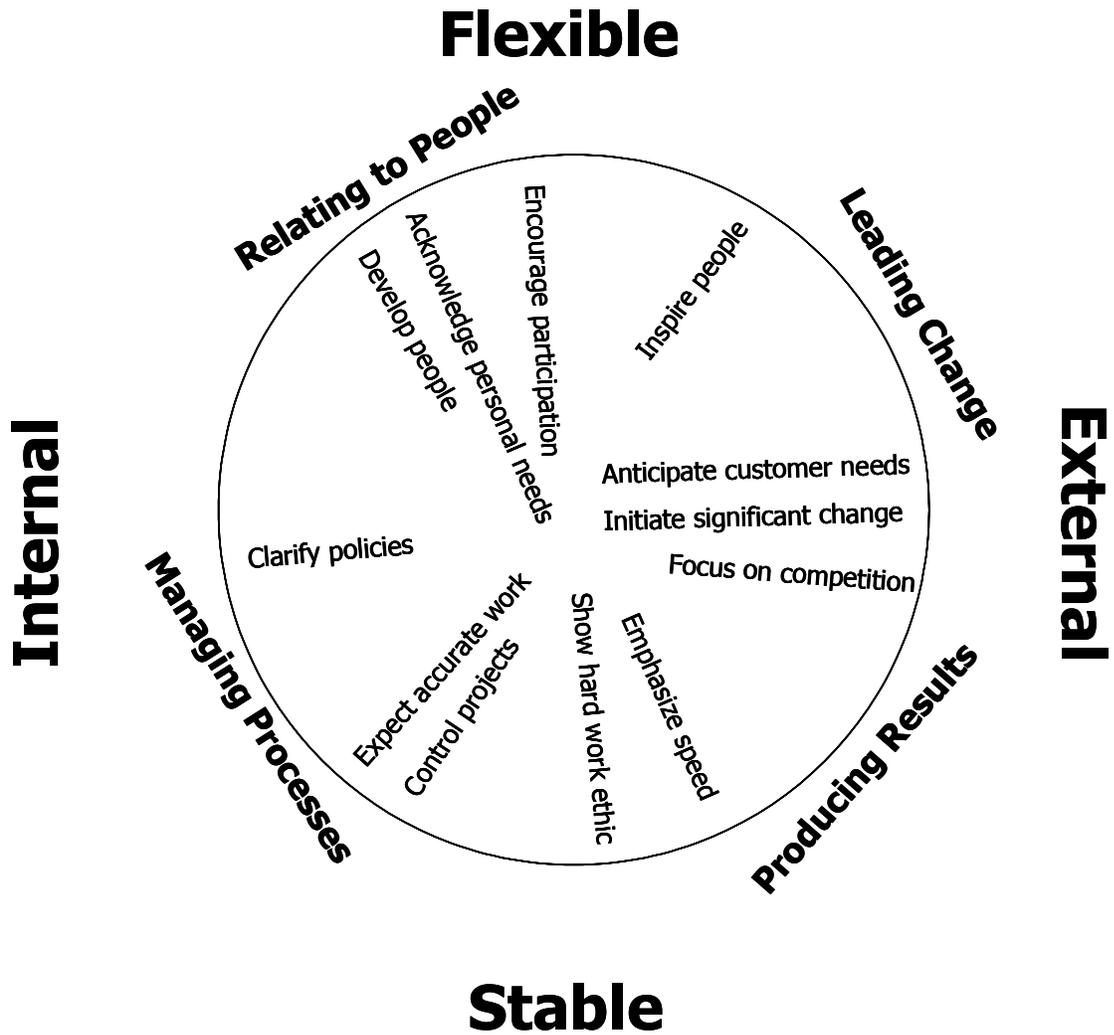


Figure 3: Average of Item Angles as Positioned in the CVF Quadrants



Appendix: Items in the 36-Item CVF Managerial Behavior Instrument

For the self-evaluation, the phrase “I would describe myself as being skilled in the following...” appears at the top of the page, and the surveys for external raters are phrased in a similar way. The questions are administered with a 5-point Likert-type scale (strongly disagree, disagree, neither agree/disagree, agree, strongly agree), plus an option “don’t know,” which is treated as missing data. Items were randomized so that constructs were not grouped together.

Relating to People

1. *Encouraging participation*
 - 1a. Making it legitimate to contribute opinions.
 - 1b. Employing participative decision making.
 - 1c. Maintaining an open climate for discussion.
2. *Developing people*
 - 2a. Encouraging career development.
 - 2b. Seeing that everyone has a development plan.
 - 2c. Coaching people on career issues.
3. *Acknowledging personal needs*
 - 3a. Being aware of when people are burning out.
 - 3b. Encouraging people to have work/life balance.
 - 3c. Recognizing feelings.

Managing Processes

7. *Clarifying policies*
 - 7a. Seeing that corporate procedures are understood.
 - 7b. Insuring that company policies are known.
 - 7c. Making sure formal guidelines are clear to people.
8. *Expecting accurate work*
 - 8a. Emphasizing the need for accuracy in work efforts.
 - 8b. Expecting people to get the details of their work right.
 - 8c. Emphasizing accuracy in work efforts.
9. *Controlling projects*
 - 9a. Providing tight project management.
 - 9b. Keeping projects under control.
 - 9c. Closely managing projects.

Leading Change

4. *Anticipating customer needs*
 - 4a. Meeting with customers to discuss their needs.
 - 4b. Identifying the changing needs of the customer.
 - 4c. Anticipating what the customer will want next.
5. *Initiating significant change*
 - 5a. Initiating bold projects.
 - 5b. Starting ambitious programs.
 - 5c. Launching important new efforts.
6. *Inspiring people to exceed expectations*
 - 6a. Inspiring direct reports to be creative.
 - 6b. Encouraging direct reports to try new things.
 - 6c. Getting unit members to exceed traditional performance patterns.

Producing Results

10. *Focusing on competition*
 - 10a. Emphasizing the need to compete.
 - 10b. Developing a competitive focus.
 - 10c. Insisting on beating outside competitors.
11. *Showing a hard work ethic*
 - 11a. Showing an appetite for hard work.
 - 11b. Modeling an intense work effort.
 - 11c. Demonstrating full exertion on the job.
12. *Emphasizing speed*
 - 12a. Getting work done quicker in the unit.
 - 12b. Producing faster unit outcomes.
 - 12c. Providing fast responses to emerging issues.