

Running Head: ASSESSING FLEXIBILITY

**Assessing the Behavioral Flexibility of Managers:
A Comparison of Methods**

Jennifer T. Lindberg
jlindberg@kaplandevries.com

KAPLAN DEVRIES INC. & **NC STATE UNIVERSITY**

Robert B. Kaiser
rkaiser@kaplandevries.com

KAPLAN DEVRIES INC.

Abstract

It is almost a cliché to say today's managers must be flexible—television commercials sell organizations on becoming more agile, business magazines warn that the only constant is change, and so on. Despite keen interest in both practitioner and scholarly circles, several questions remain about managerial flexibility: How best to define the construct? How to measure it? How to help managers develop it? We analyzed three different methods for assessing the flexibility of managers in the context of developmental feedback. Our results indicate that the way organizations typically measure flexibility—as a global trait-like characteristic—is severely deficient. However, more complex models that represent the paradoxes confronting managers show promise. Specifically, measures that represent flexibility as a mastery of specific and opposing behaviors in both the social/interpersonal domain and the functional/business domain demonstrate construct validity evidence and are highly predictive of ratings of overall effectiveness.

Introduction

The central premise behind the contingency approaches to leadership is that the effectiveness of a given behavior depends on the situation (Fiedler, 1967; Hersey & Blanchard, 1969; Vroom & Yetton, 1973). Since the managerial job is characterized by an unrelenting tempo and staccato rhythm of disparate and often unrelated episodes (Mintzberg, 1975), it follows that individuals need a flexible array of behavioral responses to effectively fulfill the role. This is probably the case more so now than ever before. As a recent *Wall Street Journal* article put it, today's managers face "a new kind of complexity and a new degree of turmoil" for all of the familiar reasons like globalization, flatter structures, rapid changes in technology, increasing speed of information flow, and so on (Murray, 2001, p. A1). According to systems theorists and the principle of requisite variety (Ashby, 1952), a controller can only deal with a system to the extent that the controller has sufficient variety to represent the complexity of that system. Modern managers, then, need a broader repertoire than their 20th century counterparts.

Scholars and practitioners alike anticipated the central role flexibility would play in management in the new millennium. The last fifteen years have seen a number of attempts in the academic literature to conceptualize, measure, and validate a construct to account for the ability of managers to adapt to rapidly changing contingencies (e.g., Dodge, Hunt, & Hooijberg, 1997; Quinn, 1988; Zaccaro, Foti, & Kenny, 1991a; Zaccaro, Gilbert, Thor, & Mumford, 1991b). The practitioner literature has also seen its share of best practice ideas about notions like versatility, adaptability, and agility (e.g., Kaplan, 1996; Lombardo & Eichinger, 2000; Sloan, 1994). These dimensions are also found regularly in the now-ubiquitous competency model and 360° survey.

Despite the flurry of research activity and widespread interest, several practical questions about managerial flexibility remain: How best to conceptualize the construct? How to assess it? How to help managers develop it? We address the first two—the related problems of definition and measurement—in the context of the third, providing developmental feedback to managers.

Conceptual Definitions

Pulakos et al. (2000) noted that, "adaptability, flexibility, and versatility are elusive concepts that have not been well defined in the psychological literature and are therefore difficult to measure, predict, and teach effectively" (p. 612). Although some models of managerial flexibility include cognitive, personality, and/or motivational components (e.g., Dodge et al., 1997; Lombardo & Eichinger, 2000; Zaccaro et al., 1991a; 1991b), our focus is on assessing it in terms of behavior. Definitional difficulties notwithstanding, there have been two dominant traditions in the conceptualization of the behavioral flexibility of managers in recent years.

The first is a general take, defining flexibility as a global construct along the lines of its dictionary meaning, such as a capability to adapt to new, different, or changing requirements. The range of specific behaviors one is capable of enacting is less important here than the generic tendency to vary one's approach. In that sense, this conceptualization is like a trait or dispositional approach. This definition is most common in the applied world—e.g., Personnel Decision International's competency of "Demonstrates Adaptability" is defined as the ability "to move quickly, deal with ambiguity, and accept change" (Davis et al., 1996, p. 648). Some researchers have studied flexibility this way, operationalizing it as a monolithic tendency to vary

behavior across contexts (e.g., Zaccaro et al., 1991a). The appeal of this kind of definition is that it is straightforward and intuitive, perhaps accounting for the fact that the assessment tools commonly used in organizations usually define flexibility this way (Leslie & Fleenor, 1998).

The second approach to conceiving of flexibility is more behaviorally specific and complex, finding its roots in the paradoxes, tensions, and trade-offs inherent in the managerial job. This tradition owes an intellectual debt to Quinn (1988) and his Competing Values Framework. To account for what separates master managers from less effective managers, Quinn offered the concept of *behavioral complexity*—"the ability to act out a cognitively complex strategy by playing multiple, even competing roles, in a highly integrated and complementary way" (Hooijberg & Quinn, 1992, p. 164). Quinn originally spoke about this in terms of *interpenetration*—the simultaneous operation of opposites. His framework represents an interpenetration called *Tough Love*, the capability to push for productivity while *also* building cohesion, and an interpenetration called *Practical Vision*, capacity to both establish stability *and* introduce change. Others have discussed managerial flexibility similarly. Sloan (1994) listed several balances to be struck, like competition and collaboration, vision and pragmatism, change and continuity, and so on. Kaplan also discussed versatility, defining it as the ability to turn freely between opposing styles like an assertive "forceful" versus a more considerate "enabling" approach or between a focus on long-range strategic needs versus near-term operational concerns (Kaplan, 1996; Kaplan & Kaiser, 2003a).

It is instructive to note that two particular distinctions appear with regularity on most lists. The first concerns social/interpersonal process—for instance, Quinn's "Tough Love" or Kaplan's "Forceful-Enabling" pair. These oppositions are similar to such prior ones as autocratic versus consultative (Vroom & Yetton, 1973), task-oriented versus relationship-oriented (Fiedler, 1967), and initiating structure versus consideration (Fleishman & Harris, 1962). The other opposition has more to do with functional business concerns and is less common in the psychological literature. It is represented by Quinn's "Practical Vision" and Kaplan's "Strategic-Operational" pair. If the first major distinction refers to the *how* of management, the second might be said to refer to the *what*.

We call the first tradition—flexibility conceived as the behaviorally abstract and general tendency to vary behavior with changing situations—the trait approach. We call the second tradition the "mastery of opposites" approach. Central here is capability and skill with contrasting behaviors that are both important despite seeming to be mutually exclusive. The key difference is that the mastery of opposites approach specifies particular behaviors whereas the trait approach does not and instead refers to general, cross-situational tendencies summarized across behaviors.

Measurement and Assessment

Ratings are the most common method for measuring performance in organizations (Murphy & Cleveland, 1995). Thus, although there are other ways of measuring the behavioral flexibility of managers (e.g., assessment centers, biodata), we restrict our focus here to coworker ratings. Ratings of flexibility as construed in the trait approach are usually collected by presenting raters with items that describe summaries of variations in behavior—without reference to specific behaviors. Sample items in typical 360° surveys include "Adapts to change," "Demonstrates

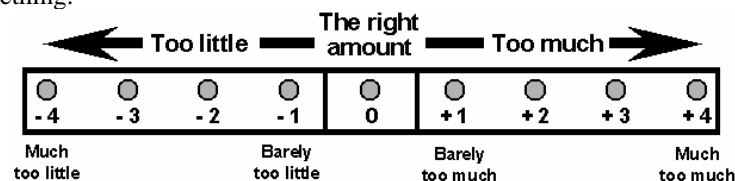
flexibility," and "Varies her approach with the situation" (e.g., see Leslie & Fleenor, 1998).

There are at least three distinct methods for measuring flexibility as conceived as a mastery of opposites. One is to use a typical scale with items that specifically refer to opposing behaviors (e.g., "Is tough and at the same time compassionate;" Lombardo & McCauley, 1994). We examine the other two in the present analyses. The first—what Quinn, Spreitzer, and Hart (1992) call the "interpenetration" method—has raters use a traditional Likert-type scale to rate performance on items from two separate scales representing opposing dimensions. Then a formula is used to combine the two scale scores into a single continuous variable where high values indicate "integrative balance" (i.e., relatively equal scores) between contrasting roles.

The final method in the mastery of opposites approach is what Kaplan and Kaiser (2003a; 2003b) call a *duality-oriented* view of versatility. They note that managers tend to over-rely on one side and under-use the other side in dualities like Forceful-Enabling and Strategic-Operational. Thus their measure uses a new type of response scale (see Figure 1) that ranges from underdoing to optimal to overdoing. Similar to the interpenetration method, separate scales are used for opposing dimensions. Although items are rated one at a time, they are scored in pairs (e.g., each forceful item has an enabling complement). For each pair, geometry is used to calculate how close the ratings are to "optimal" on both. Then this value is averaged across item pairs to arrive at a versatility score for that duality (e.g., Forceful-Enabling). High scores indicate versatility as conceived as optimal and balanced on both sides of a duality; lower scores indicate "lopsidedness" (too much of one, not enough of the other) or "disengaged" (too little of both).

Figure 1. The "implicitly curvilinear" scale in the duality-based measure of versatility.

Please note that the scale is probably different from scales that you are accustomed to using. On this scale the best score is "0," in the middle of the scale. The premise is that performance problems arise when managers either underdo or overdo something.



The Present Study

Below, we examine the convergent and discriminant validity of the trait approach to measuring the behavioral flexibility of managers and two methods of assessing the mastery of opposites approach, Quinn's interpenetration method and Kaplan's duality-based method. We also examine both the absolute and incremental validity of each approach in predicting overall effectiveness. Our central concern is the relative merits of each method for providing managers with developmental feedback.

Method

Participants. The data for this study are 360° performance ratings of 29 U.S. executives who participated in a leadership development program. A total of 264 coworkers (33 superiors,

93 peers, 138 subordinates) provided the ratings. All raters were informed that the data would be used solely for developmental feedback. Except for superiors, the raters were assured anonymity.

Measures. The scales used to measure the various conceptions of flexibility come from two commercial 360° feedback instruments, SKILLSCOPE® (Kaplan, 1997) and the *Leadership Versatility Index*™ (LVI; Kaplan & Kaiser, 2002). SKILLSCOPE® is a checklist that contains 98 items rated as either "development needed" or "strength." Respondents can leave the item blank to indicate "neither." We coded responses to these items as 1 = development needed, 2 = neither, and 3 = strength. SKILLSCOPE® does not contain conventional multi-item scales; the items are grouped by conceptual similarity. We created multi-item scales for this study based on factor analyses of a larger SKILLSCOPE® database (N > 80,000 raters) and selecting items that were very similar in content to those used in prior studies.

Our trait measure of flexibility is composed of five items from SKILLSCOPE®. These items refer to flexibility and making adjustments in behavior in general, devoid of behavioral specification. The scales used to measure the two interpenetrations (Tough Love and Practical Vision) in the mastery of opposites approach also come from SKILLSCOPE®. Consistent in content and name with the measures used by Quinn et al. (1992), we created two 7-item scales, labeled Accomplishment and Cohesion, to use in the computation of the Tough Love interpenetration; for Practical Vision, we created two 7-item scales, Stability and Innovation.

To represent the duality-based approach to the mastery of opposites notion of flexibility, we used four five-item scales (forceful, enabling, strategic, and operational) from the LVI (Kaplan & Kaiser, 2002; 2003a). LVI items are rated on a unique response scale that separates underdoing and overdoing as two distinct types of ineffective performance. The scale ranges from - 4 (much too little) to + 4 (much too much), with the optimal point in the middle, 0 (the right amount). Although the items are rated one at a time, they were created in pairs reflecting complementary opposites. For instance, the forceful item "pushes people hard" is paired with the enabling item "provides support." Each forceful item has a complementary enabling item; each strategic item has an operational companion item.

Overall effectiveness was measured by asking raters, in a semi-structured interview conducted separately from the 360 ratings, to "Please rate X's overall effectiveness as an executive on a ten-point scale, where 10 is outstanding and 5 is adequate." Although single-item measures can be deficient, Kaiser and Craig (2001) offered support for the psychometric adequacy of this one. They reported estimates of inter-rater reliability and agreement and between-source (i.e., superior, peer, and subordinate) convergence correlations for this single-item rating that were comparable to meta-analytic estimates of the same statistics for multiple-item scales (cf. Conway & Huffcut, 1997).

The behavioral cores of all items in each scale are shown in Table 1. Descriptive statistics, reliabilities, and correlations are presented in Table 2. Reliability estimates for each scale exceeded the .70 minimum recommended by Nunnally (1978) for use in basic research.

Procedures. Assessing flexibility with the interpenetration and duality-based methods required further computation beyond creating scale scores. We calculated the interpenetration scores following Quinn's method (Quinn et al., 1992; Hooijberg & Quinn, 1992). Specifically, he

Table 1. Behavioral cores from the items used for each measure of managerial flexibility.

TRAIT APPROACH	
Flexibility	
Varies approach with the situation	
Makes adjustments in behavior	
Learns from experience; not set in his/her ways	
Thinks in terms of trade-offs	
Takes ideas different from own seriously	
MASTERY OF OPPOSITES APPROACH	
Interpenetration I – TOUGH LOVE	
Accomplishment	Cohesion
Driven to achieve objectives	Delegates
Presses for immediate results	Shares responsibility with subordinates
Good initiative	Recognizes and rewards people
Decisive	Cooperative
High energy level	Collaborates well
Sparks others to take action	Deals well with people's feelings
Extremely productive	Makes good use of people
Interpenetration II – PRACTICAL VISION	
Stability	Innovation
Sets priorities well	Creates significant change
Spots problems early	Generates new ideas
Gets to the heart of problems	Introduces change despite opposition
Good troubleshooter	Has vision
Optimistic that problems can be solved	Conveys a sense of purpose
Implements decisions, follows through	Seizes new opportunities
Data-based, rational	Good at promoting ideas
Duality I – FORCEFUL-ENABLING VERSATILITY	
Forceful	Enabling
1f. Decisive	1e. Participative
2f. Pushes people hard	2e. Provides support and encouragement
3f. Holds people accountable	3e. Understanding when people don't deliver
4f. Makes tough calls	4e. Compassionate
5f. Asks challenging questions	5e. Makes it easy for people to push back
Duality II – STRATEGIC-OPERATIONAL VERSATILITY	
Operational	Strategic
1o. Hands-on	1s. Sees the big picture
2o. Focused on executing	2s. Sets long-term direction
3o. Realistic; practical	3s. A visionary
4o. Internally oriented	4s. Externally oriented
5o. Uses structure, discipline	5s. Exciting view of the future

Note: Forceful (f) and Enabling (e) items (and Strategic [s] and Operational [o] items) with the same number are pairs. Items used in the trait approach and for the interpenetration method are from SKILLSCOPE® (Kaplan, 1997) and are reproduced here with permission from the Center for Creative Leadership.

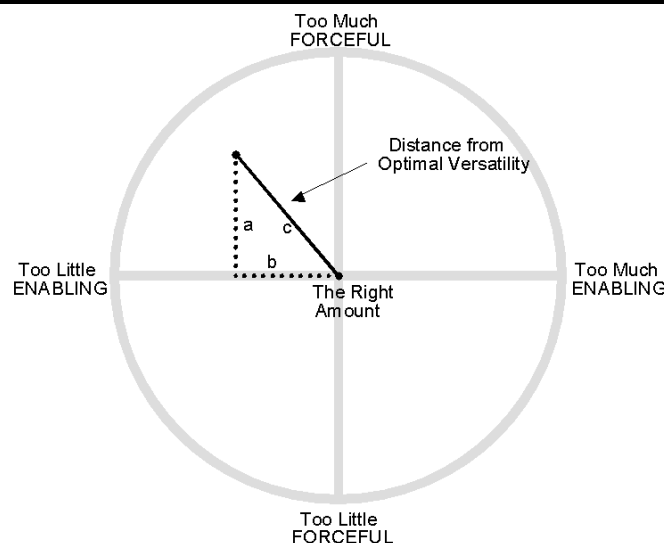
recommends the formula provided by Bobko and Schwartz (1984) for integrating bi-polar constructs. This method was developed as a way to construct a single continuous variable to represent the integrative balance of conceptually opposing constructs (e.g., androgyny as balance on the two separate dimensions of masculinity and femininity). The equation is:

$$\text{Interpenetration} = [(k-1) - (|X - Y|)] * [(X + Y)/2]$$

where X and Y are opposites to be integrated and measured on a scale ranging from 1 to k . With $k = 3$, as is here, values can range from 0 to 6. High scores indicate managers who are rated high but relatively equal on contrasting concepts. Lower scores reflect high on one, low on the other. We used this procedure to derive scores for two interpenetrations, Tough Love (balance on Cohesion and Accomplishment) and Practical Vision (Stability and Innovation) and computed the mean on these two for an overall interpenetration score (c.f. Quinn et al., 1992).

We followed the procedure described by Kaiser and Kaplan (2002) to compute versatility scores for the duality-based measures. These scores consider jointly the extent to which a manager is rated as using opposing behaviors optimally versus doing too much of one and too little of the other. For example, in the upper-left-hand corner of Figure 2 is a plot for ratings in the "too much" region on forceful and in the "too little" region on enabling. The distance these ratings are from the right amount (0) on both items can be derived from the Pythagorean theorem. It is calculated as: $c^2 = a^2 + b^2$ where $a = (\text{forceful score} - 2)$, $b = (\text{enabling score} - 2)$, and $c = \text{distance from optimal on both}$. A versatility score for each item pair is calculated as the ratio of the observed *distance from optimal* to the maximum possible distance from optimal (i.e., scores on the extreme ends of the scale, - 4 and + 4). The average of these scores is calculated across the five item pairs for that duality. So that higher scores indicate more versatility, this average value is subtracted from 100%. Versatility scores can range from 0% (most lopsided) to 100% (perfectly versatile). We computed versatility scores for Forceful-Enabling, Strategic-Operational, and an overall summary (the mean of the first two scores).

Figure 2. How the Pythagorean theorem is used to compute the "distance from optimal" for a pair of items in the duality-based method.



Results

All analyses were conducted on rater-level data. Researchers sometimes use either one rater per focal manager or the average rating from all individuals in a given group (e.g., subordinates) when analyzing multisource ratings. But these techniques limit statistical power unnecessarily (Craig & Kaiser, 2003). Furthermore, there is an emerging literature suggesting that rating source isn't a significant factor and that individual rater effects account for over half of the variance in the latent structure of multisource ratings (Mount et al., 1998; Scullen, Mount, & Goff, 2000). These findings not only support but recommend rater-level analyses for multisource ratings data.

Convergent/Discriminant Validity. We first assessed the convergent and discriminant validity of each way to measure the behavioral flexibility of managers. The pattern of correlations in Table 2 suggests moderate to high convergence across methods, with correlations between the three of .31, .51, and .53. The interpenetration method showed the strongest relationships with the other two; the trait and duality methods were least related. The trait approach was more strongly related to Tough Love than Practical Vision and Forceful-Enabling than Strategic-Operational. Raters evidently weigh interpersonal behaviors more than functional business skills to judge flexibility in the abstract. Finally, the two measures in the mastery of opposites tradition showed the most convergence. And there was further construct validity evidence for both methods in that Tough Love and Forceful-Enabling were most related to one another and Practical Vision and Strategic-Operational were most related to each other.

A final observation merits attention. The mastery of opposites approach assumes that contrasting dimensions are in tension. This suggests a negative relationship (Kaplan & Kaiser, 2003b). However, in the interpenetration data, the opposites were actually positively related ($r = .33$ for Accomplishment and Cohesion, $r = .54$ for Stability and Innovation). The duality-oriented measures were more consistent with theoretical expectations in that Forceful and Enabling were inversely related ($r = -.49$) while Strategic and Operational were independent.

Concurrent Validity. We were next interested in comparing the ability of each method of assessing flexibility to predict ratings of overall effectiveness. The trait measure was least valid, ($r = .28$), while the two mastery of opposites methods were quite valid ($r = .52$ for the overall interpenetration score, $r = .64$ for the overall duality-based versatility score). To further understand the validity of the two mastery of opposites approaches, we next conducted multiple regression analyses using the two component scores as separate predictors. These results appear in Table 3. The interpenetration model accounted for 29% of the variance in overall effectiveness and the duality-based versatility model accounted for a notably higher 42%. Interestingly, in both models, flexibility in functional/business terms (Practical Vision, Strategic-Operational) was almost twice as strongly related to overall effectiveness than was flexibility in social/interpersonal terms (Tough Love, Forceful-Enabling). This may help explain why the trait measure showed such little validity—it is primarily saturated with interpersonal flexibility.

Incremental Validity. Finally, we examined the incremental validity of each of the three approaches relative to the other two. Since all three approaches overlap with one another empirically, we wanted to know what unique contribution each could make in predicting overall effectiveness. To do this, we constructed three hierarchical regression models, first entering the

Table 2. *Descriptive statistics and correlations for all study variables.*

	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Flexibility (trait)	2.47	.47	(.74)															
2. Interpenetration (overall)	4.34	.99	.51	(.75)														
3. <i>Interpenetration I (Tough Love)</i>	4.13	1.36	.49	.87	(.78)													
4. Accomplishment	2.71	.34	.24	.55	.40	(.74)												
5. Cohesion	2.43	.51	.56	.78	.89	.33	(.81)											
6. <i>Interpenetration II (Practical Vision)</i>	4.55	1.04	.32	.76	.33	.52	.32	(.72)										
7. Stability	2.65	.36	.33	.61	.37	.64	.34	.68	(.70)									
8. Innovation	2.59	.40	.30	.64	.31	.50	.36	.79	.54	(.74)								
9. Duality-based Versatility (overall)	.83	.12	.31	.53	.40	.46	.44	.48	.51	.54	(.74)							
10. <i>Forceful-Enabling Versatility</i>	.82	.13	.30	.49	.43	.40	.47	.35	.43	.45	.89	(.73)						
11. Forceful	-.30	.63	-.20	.06	-.10	.39	-.11	.25	.35	.27	.37	.35	(.71)					
12. Enabling	-.36	.70	.37	.29	.44	-.05	.50	-.02	-.06	.06	.17	.20	-.49	(.77)				
13. <i>Strategic-Operational Versatility</i>	.84	.13	.26	.46	.29	.42	.32	.50	.48	.52	.90	.59	.31	.10	(.79)			
14. Operational	-.25	.50	-.06	.20	.16	.24	.14	.17	.29	.12	.36	.22	.34	.04	.41	(.71)		
15. Strategic	-.38	.58	.15	.22	.02	.21	.05	.39	.22	.49	.54	.32	.22	.05	.65	-.01	(.76)	
16. Overall Effectiveness	8.15	1.15	.28	.52	.37	.41	.36	.49	.51	.48	.64	.52	.18	.05	.61	.31	.38	(.72)

Notes: $N = 264$. Correlations $> |.20|$ are significant at $p < .001$, $> |.16|$ are significant at $p < .01$, $> |.12|$ are significant at $p < .05$.

Reliability estimates appear on the diagonal (interpenetration reliabilities were estimated conservatively as the average α of constituent scales, α is presented for all other scales; the reliability for overall effectiveness is the estimate reported by Kaiser & Craig, 2001). Scale scores were computed as the average rating across the respective items (can range from 1 to 3 for Flexibility, Accomplishment, Cohesion, Stability, and Innovation; Forceful, Enabling, Strategic, and Operational scales can range from -4 to $+4$). See text for computation of Interpenetration (possible range of 0 to 6) and Versatility (possible range of .00 to 1.00).

Table 3. Summary of multiple regression analyses predicting overall effectiveness with the two mastery of opposites models of flexibility.

<i>Interpenetration Method</i>	
	β
Tough Love	.24***
Practical Vision	.41***
Model R^2	.29***
<i>Duality-based Method</i>	
	β
Forceful-Enabling Versatility	.25***
Strategic-Operational Versatility	.47***
Model R^2	.42***

Note: *** $p < .001$

terms for the first two methods, then testing the additional variance accounted for by the focal method. These results appear in Table 4. The trait measure yielded no incremental validity over the two mastery of opposites approaches. Said differently, the mastery of opposites measures contain all of the criterion-related variance of the trait measure and then some.

The interpenetration method enhanced the prediction of variance in overall effectiveness by a significant 4.2% beyond the 42.5% accounted for by the trait and duality-based approaches combined. Finally, the duality-based method accounted for an additional 17.4% of the variance in effectiveness beyond the 29.4% attributable to the combination of the trait and interpenetration methods. Thus, although the two mastery of opposites methods showed substantial convergence, it appears that the duality-based method contains about four times more unique variance related to overall effectiveness (i.e., $17.4/4.2 = 4.14$).

Discussion

The purpose of this study was to determine the relative merits of three methods of assessing behavioral flexibility for use in management development efforts. The results offer support for the more complex and behaviorally specific "mastery of opposites" approach. Although the trait approach is notably more simple and intuitive, making it attractive for use in practice, it appears deficient on several counts. First, it taps only a fraction of what is measured by the mastery of opposites approach. Second, the mastery of opposites approach measures all of the criterion-related variance contained in the trait measure—and then some. Third, the trait approach appears to mostly tap flexibility in terms of social/interpersonal process. Lacking is flexibility in terms of functional/business matters, which appears to be the opposition most related to perceptions of overall effectiveness. And on a conceptual note, feedback in the form of the trait measure is difficult to act on. In effect, low scores suggest improvement lies in "being more flexible." This

is behaviorally ambiguous and ignores the conventional—and empirically corroborated (Kluger & DeNisi, 1996)—wisdom of providing specific, behavior-focused feedback. By contrast, feedback in the form suggested by the mastery of opposites methods, along the lines of "do less of X and more of Y," is more instructive.

Table 4. Summary of hierarchical regression analyses testing incremental validity of three methods of assessing flexibility relative to each other.

Test Trait Approach		Test Interpenetration Approach		Test Duality-Based Approach	
	β		β		β
<i>Step 1</i>		<i>Step 1</i>		<i>Step 1</i>	
Tough Love	.12*	Flexibility (trait)	.10*	Flexibility (trait)	.05
Practical Vision	.20***	Forceful-Enabling	.23***	Tough Love	.22***
Forceful-Enabling	.18**	Strategic-Operational	.46***	Practical Vision	.41***
Strategic-Operational	.37***				
<i>Step 1 R²</i>		<i>Step 1 R²</i>		<i>Step 1 R²</i>	
		.467***		.425***	
				.294***	
<i>Step 2</i>		<i>Step 2</i>		<i>Step 2</i>	
Flexibility (trait)	.01	Tough Love	.12*	Forceful-Enabling	.18**
		Practical Vision	.20***	Strategic-Operational	.37***
<i>Step 2 ΔR^2</i>		<i>Step 2 ΔR^2</i>		<i>Step 2 ΔR^2</i>	
		.000		.042***	
				.174***	

Note: *** $p < .001$ ** $p < .01$ * $p < .05$

There were several noteworthy results concerning the two approaches in the mastery of opposites tradition. First, there was substantial convergence between the interpenetration and duality-based methods. This is true for the overall scores as well as for the component scores. The two social/interpersonal-oriented oppositions, Tough Love and Forceful-Enabling versatility, bore a stronger relationship with one another than with either of the two functional/business-oriented oppositions, Practical Vision and Strategic-Operational versatility. The converse was also true. This provides construct validity evidence for the conceptual distinction between flexibility in terms of the *how* and *what* of management. It is also evidence for the validity of both the interpenetration and duality-based conceptions of flexibility—the two methods appear to converge on the same phenomena.

Also relevant to the question of construct validity is the relationships between putatively opposing dimensions. The underlying notion in the mastery of opposites approach is that contrasting dimensions are in tension with one another (Quinn, 1988), implying a negative relationship between the two. However, using traditional Likert-type rating scales, we found positive relationships between opposites in the interpenetration data. This is not unique to our data set and the present scales: the scales used in Quinn's program of research consistently demonstrate positive correlations with each other of a magnitude similar to the present results (e.g., Dennison, Hooijberg, & Quinn, 1995; Hooijberg & Choi, 2000). This inconsistency is provocative and begs explanation. It would seem that either the theory of "tension between opposites" is wrong or the traditional approach to assessing them is flawed. To this point, the

duality-based measures with the too little/too much response scale did not reveal positive associations between opposites. In fact, the hypothesized inverse relationship was found between the forceful and enabling dimensions. It may be the case that this new rating scale confers certain advantages over the typical response scale (Kaiser & Kaplan, 2004). We leave it to future research to explicate those advantages as well as the limitations to that type of scale.

Relative to the interpenetration method, the duality-based approach both out-predicted overall effectiveness and yielded more incremental validity. This suggests that it is the better approach to use for diagnosing strengths and weaknesses in the context of management development. Of course, further research with a wider range of effectiveness criteria (e.g., unit productivity, unit climate, collective team efficacy, team viability, subordinate role clarity and conflict, job satisfaction, organizational commitment, and so forth) and with alternative measures (e.g., those developed by Quinn and his associates) is needed to make any definitive claim on this point.

On the grounds of face validity, both approaches in the mastery of opposites tradition are somewhat opaque because of the statistical procedures needed to derive a single index of flexibility from two separate measures. However, the duality-based computation of versatility seems easier to explain, at least visually, than the Bobko formula used in the interpenetration method. Where user acceptance is a critical factor, this distinction warrants consideration.

There is at least one finding in our results with major implications for the study of managerial effectiveness. We found clear evidence that flexibility in terms of functional/business issues—the *what* of management—was far more related to how managers gauge each others' overall effectiveness than was flexibility in terms of social/interpersonal behavior, the *how* of management. This is provocative for several reasons. First, psychologists have put a great deal of attention on the *how* of management and leadership. The first major paradigm in modern leadership theory was focused on various forms of this (Hunt & Dodge, 2000). And each of the major contingency theories focused on it as well—Fiedler's (1967) explication of which situations favor task or relationship behavior, Hersey and Blanchard's (1969) situational model of when task-oriented or people-oriented behaviors are best, and Vroom and Yetton's (1973) normative model for when to make decisions autocratically or with a consultative approach. Rare is the model that explicitly focuses on the dilemmas between such business issues as short-term profits versus long-term viability, efficiency versus experimentation, tactical discipline versus visionary inspiration, and so forth (for an exception, see Kotter, 1990). This represents a ripe opportunity for students of management to make their work more relevant to industry.

References

- Ashby, W.R. (1952). *Design for a brain*. Wiley & Sons, New York.
- Bobko, P. & Schwartz, J. (1984). A metric for integrating theoretically related but statistically uncorrelated constructs. *Journal of Personality Assessment*, 48, 11-18.
- Conway, J.M. & Huffcut, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of supervisor, peer, subordinate, and self-ratings. *Human Performance*, 19, 331-360.
- Craig, S.B. & Kaiser, R.B. (2003). Applying item response theory to multisource performance ratings: What are the consequences of violating the independent observations assumption? *Organizational Research Methods*, 6, 41-58.
- Davis, B.L., Skube, C.J., Hellervik, L.W., Gebelein, S.H., & Sheard, J.L. (1996). *Successful manager's handbook: Development suggestions for today's managers*. Minneapolis, MN: Personnel Decisions International.
- Dennison, D.R., Hooijberg, R., & Quinn, R.E. (1995). Paradox and performance: Toward a theory of behavioral complexity in managerial leadership. *Organizational Science*, 6, 524-540.
- Dodge, G.E., Hunt, J.G., & Hooijberg, R. (1997). Leadership complexity and development of the leaderplex model. *Journal of Management*, 23, 375-408.
- Fiedler, F.E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.
- Fleishman, E.A., & Harris, E.F. (1962). Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, 15, 43-56.
- Hersey, P. & Blanchard K. (1969). Life cycle theory of leadership. *Training and Development Journal*, 2, 6-34.
- Hooijberg, R., & Choi, J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *Leadership Quarterly*, 11, 341-364.
- Hooijberg, R. & Quinn, R.E. (1992). Behavioral complexity and the development of effective managerial leaders. In R.L. Phillips & J.G. Hunt (Eds.) *Strategic management: A multiorganizational-level perspective*. New York: Quorum.
- Hunt, J.G. & Dodge, G.E. (2000). Leadership deja vu all over again. *Leadership Quarterly*, 11, 435-458.
- Kaiser, R. B., & Craig, S. B. (2001). *Leadership Versatility Index™ technical report: Item selection and validation*. Greensboro, NC: Kaplan DeVries Inc.
- Kaiser, R.B. & Kaplan, R.E. (2004, April). Overlooking overkill: On the folly of linear rating scales for a non-linear world. In S.B. Craig (Chair) 360, *The Next Generation: Innovations in Multisource Performance Assessment*. Symposium at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.
- Kaiser, R.B. & Kaplan, R.E. (2002). *Leadership Versatility Index™ user's guide*. Greensboro, NC: Kaplan DeVries Inc.
- Kaplan, R.E. (1997). *SKILLSCOPE® for Managers*. Greensboro, NC: Center for Creative Leadership.
- Kaplan, R.E. (1996). *Forceful leadership and enabling leadership: You can do both*. Greensboro, NC: Center for Creative Leadership.
- Kaplan, R. E., & Kaiser, R. B. (2003a). Developing versatile leadership. *MIT Sloan Management Review*, 44, 19-26.
- Kaplan, R. E., & Kaiser, R. B. (2003b). Rethinking a classic distinction in leadership: Implications for the assessment and development of executives. *Consulting Psychology Journal: Practice and Research*, 55, 15-25.
- Kaplan, R.E., & Kaiser, R.B. (2002). *Leadership Versatility Index™*. Greensboro, NC: Kaplan DeVries Inc.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.

- Kotter, J.P. (1990). *A force for change: How leadership differs from management*. New York: Free Press.
- Leslie, J.B., & Fleenor, J.W. (1998). *Feedback to managers: A review and comparison of multi-rater instruments for management development*. Greensboro, NC: Center for Creative Leadership.
- Lombardo, M.M. & Eichinger, R.W. (2000). High potentials as high learners. *Human Resource Management, 39*, 321-329.
- Lombardo, M. & McCauley, C.D. (1994). *Benchmarks: A manual and trainer's guide*. Greensboro, NC: Center for Creative Leadership.
- Mintzberg, H. (1975). The manager's job: Folklore and fact. *Harvard Business Review, 53*, 100-110.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557-576.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Murray, M. (2001, February 8). Critical mass: As huge companies keep growing, CEOs struggle to keep pace. *The Wall Street Journal*, pp. A1, A6.
- Nunnally, J. (1978). *Psychometric theory* (2nd Ed.). McGraw-Hill, New York.
- Pulakos, E.D., Arad, S., Donovan, M.A., & Plamondon, K.E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*, 612-624.
- Quinn, R.E. (1988). *Beyond rational management*. San Francisco: Jossey-Bass.
- Quinn, R. E., Spreitzer, G. M., & Hart, S. L. (1992). Integrating the extremes: Crucial skills for managerial effectiveness. In S. Srivastva and R. E. Fry (Eds.). *Executive and organizational continuity: Managing the paradoxes of stability and change* (pp. 222-252). San Francisco: Jossey-Bass.
- Scullen, S.E., Mount, M.K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.
- Sloan, E. B. (1994). Assessing and developing versatility: Executive survival skill for the brave new world. *Consulting Psychology Journal: Practice and Research, 46*, 24-31.
- Vroom, V.H. & Yetton, P.W. (1973). *Leadership and decision-making*. New York: Wiley.
- Zaccaro, S.J., Foti, R.J., & Kenny, D.A. (1991a). Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situations. *Journal of Applied Psychology, 76*, 308-315.
- Zaccaro, S.J., Gilbert, J.A., Thor, K.K., & Mumford, M.D. (1991b). Leadership and social intelligence: Linking social perceptiveness and behavioral flexibility to leader effectiveness. *Leadership Quarterly, 2*, 317-342.