Assessing the Flexibility of Managers: A comparison of methods[1]

Robert B. Kaiser*, Jennifer T. Lindberg*, and S. Bartholomew Craig****

*Kaplan DeVries Inc., 1903 G Ashwood Court, Greensboro, NC 27455, USA. rkaiser@kaplandevries.com
**North Carolina State University, NC, USA

Despite keen interest, questions remain about defining and measuring the behavioral flexibility of managers. This paper reports a conceptual and empirical comparison of three alternative methods of assessing this construct. Results suggest that the way managerial flexibility is typically assessed in practice – as a trait-like characteristic with coworker ratings that describe a general tendency to vary behavior across situations – is deficient. However, more complex models that represent flexibility as a higher-order construct reflecting mastery of specific and opposing behaviors in both the social/interpersonal domain and the functional/organizational domain show promise. They demonstrate construct validity evidence, predict as much as 42% of the variance in overall effectiveness, and provide more specific diagnostic information to guide behavior change.

1. Introduction

Contingency approaches to leadership rest on the premise that the effectiveness of a given behavior depends on the situation (Fiedler, 1967; House, 1971; Vroom & Yetton, 1973). As the managerial job is characterized by an unrelenting tempo and staccato rhythm of disparate and often unrelated episodes (Mintzberg, 1975), it follows that managers need an array of flexible behaviors. This is probably the case more so now than ever before. As a *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 rapid changes in technology, more and faster information flow, flatter structures, globalization, and so forth (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 can represent the complexity of that system. Modern managers, then, need a broader repertoire than their 20th century counterparts.

Both scholars and practitioners anticipated the pivotal role flexibility would play in management in the new millennium. There have been a number of attempts in the recent academic literature to conceptualize, measure, and validate a construct to account for the ability of managers to adapt to a wide variety of changing circumstances (e.g., Briscoe & Hall, 1999; Dennison, Hooijberg, & Quinn, 1995; Hooijberg, Hunt, & Dodge, 1997; Hooijberg, & Quinn, 1992; Quinn, 1988; Zaccaro, Foti, & Kenny, 1991a; Zaccaro, Gilbert, Thor, & Mumford, 1991b). The practitioner literature has also seen similar notions like leadership versatility (Kaplan & Kaiser, 2006; Sloan, 1994) and agility (Lombardo & Eichinger, 2000). These dimensions are also common in competency models and 360° surveys (Leslie & Fleenor, 1998). The impetus for this work lies in how the increased complexity and pace of change in the modern workplace puts a premium on managerial flexibility,

making it prerequisite for effective performance. Indeed, Briscoe and Hall (1999) defined adaptability as a metacompetency – one of the key competencies that drives the development of all other competencies. Taken together, these trends reflect how the measurement and development of managerial flexibility is of both theoretical and practical significance.

Despite widespread interest and research activity, 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 feedback to managers with coworker ratings.

1.1. Conceptualizing behavioral flexibility

Pulakos, Arad, Donovan, and Plamondon (2000) noted how '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 (Streufert & Nogami, 1989; Zaccaro *et al*., 1991b), personality (Lombardo & Eichinger, 2000; Zaccaro *et al*., 1991b), or motivational components (Briscoe & Hall, 1999) – or even all of these (Hooijberg *et al*., 1997) – our focus is on assessing it in terms of behavior. There are three related reasons for this choice. First, we are concerned with measurement for the pragmatic purpose of assessment for development, as opposed to measurement for its own sake or for strictly research purposes. Second, most development activities in organizations involve assessments of behavior (e.g., 360° feedback) and feedback is more likely to lead to improved performance if it is focused on behavior (Kluger & DeNisi, 1996). Finally, we agree with Dennison *et al*. (1995) that although many cognitive, dispositional, and motivational antecedents to flexibility may be necessary for managerial effectiveness, flexible *behavior* 'must certainly be the sufficient condition' (p. 524) because performance is ultimately the result of action.

1.1.1. A case for behavioral flexibility

There is a rich tradition in the study of management that recognizes the multiplicity of roles managers must play. The seminal ethnographic studies conducted by Mintzberg (1975) may be the most influential in the modern era. Based on an in-depth, structured observation study of executives, Mintzberg delineated 10 essential roles that each fit into one of three clusters – interpersonal, informational, and decisional role sets – that managers enact on a regular basis. Moreover, Mintzberg argued that these roles are integrated, forming a necessary gestalt and set of interdependencies that define the job of manager. For instance, the interpersonal roles lead to the establishment of networks and contacts that provide the data needed to fulfill the informational roles. And the informational roles provide a basis for making and carrying out the decisional roles. Although Mintzberg did recognize that the amount of time managers spend in each role varies with the functional nature of the position (e.g., sales managers spend more time performing interpersonal roles, production managers spend more time in decision roles), he emphasized that all 10 roles are vital to effective performance. Because each of these roles requires a somewhat distinct set of behaviors, managers must be able to enact a wide range of behavior patterns to match the requirements of the different tasks that comprise the managerial job (Zaccaro, 2001).

Another revelation from Mintzberg's (1975) ethnographic study was that the managerial job is characterized by fast-paced action and discontinuity. For instance, about half of the activities the managers he observed engaged in were brief, <9 min, while only about 10% of their activities took longer than 1 h. These tasks also varied greatly from one to the next, where it was not uncommon to move quickly from a phone call about a performance problem in a business unit to an update from the personnel department on current staffing needs to making a decision about changing vendors all within the course of a half-hour. The pace and diversity of activities facing managers, then, not only means that they must have a wide behavioral repertoire but they must also be able to toggle between different behaviors on short notice (see Zaccaro, 2001, for a similar analysis). In short, managers must be flexible. There have been two dominant traditions in the conceptualization of flexibility in recent years.

1.1.2. Trait approach

The first tradition defines 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. For instance, using Mintzberg's (1975) managerial roles, this view puts less emphasis on how well or when managers perform interpersonal, informational, and decisional roles and instead emphasizes whether managers shift into and out of each. In that sense, this conceptualization could be classified as a trait or dispositional view.[2] The appeal of this kind of definition is that it is straightforward and intuitive.

A trait-like definition of managerial flexibility is frequently used in the applied world (although some researchers have employed it too, e.g., Zaccaro *et al*., 1991b). For instance, one widely used commercial

competency model defines the dimension of *Demonstrates Adaptability* as the ability 'to move quickly, deal with ambiguity, and accept change' (Davis, Skube, Hellervik, Gebelein, & Sheard, 1996, p. 648). Other instruments contain dimensions like *Flexible* defined as 'easily adjusts to change' and *Leadership-style Flexibility* defined as 'the extent to which managers treat everyone the same or treat people differently' (see Leslie & Fleenor, 1998, pp. 68 and 114, respectively). In each of these examples, it is clear that the particular behaviors that managers employ and shift between is far less important than the fact of shifting, changing, or adjusting in light of circumstances. This is, at least in spirit, consistent with Mintzberg's (1975) position that effective managers must have a flexible behavioral repertoire to fulfill the 10 roles he identified.

1.1.3. Mastery of opposites

The second approach to conceiving of flexibility is more behaviorally specific and conceptually complex, finding its roots in the paradoxes, tensions, and trade-offs inherent in the managerial job. Although Mintzberg (1975) did not explicitly identify conflict among managerial roles, several writers who built off of his pioneering work have. For instance, Tsui's (1994) reputational effectiveness model emphasizes how superiors, peers, and subordinates each have unique expectations of the same target manager. Moreover, these expectations are often incongruent and in extreme cases, they may be diametrically opposed. For instance, some of the behaviors superiors consider to be effective may be negatively related to the behaviors subordinates require to consider their boss effective. As Zaccaro (2001, p. 131) noted after reviewing this literature, 'for a leader, success is likely to depend on maintaining a delicate balance of conflicting role behaviors.'

This view of managerial flexibility has been elaborated by Quinn (1988) in the competing values framework (CVF). Quinn argued that conflicting values are inherent in organizational life, as typified in the conflicting and oppositional nature of the four dominant models of organizational effectiveness described in the literature. The four primary perspectives are the human relations, open systems, internal process, and rational goal models. Moreover, these distinct views are not just different schools of scholarly thought on organizational effectiveness but are also alternative frames of reference that individual managers adopt.

The central premise to the CVF is that all four perspectives have their uses and highly effective managers toggle between them, applying the perspective best suited for the situation at hand. To account for how managers do this, 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 (Quinn, Spreitzer, & Hart, 1992). The CVF encompasses two interpenetrations. The first, *Tough Love*, represents resolution of the rational goal and human relations opposition and is defined as the capability to push for productivity while *also* building cohesion and morale. *Practical Vision*, the second interpenetration, has to do with the conflict between the open systems and internal process models and involves the capacity to both adapt and introduce change *and* establish stability and predictability. 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 discussed versatility in terms of the ability to turn freely between opposing styles like an assertive 'forceful' approach vs a more considerate 'Enabling' approach or between a focus on long range strategic needs vs near term operational matters (Kaplan, 1996; Kaplan & Kaiser, 2003a, 2006). He contrasted versatility with lopsidedness, where managers display a bias in favor of one and a prejudice against its opposite.

It is instructive to note that two distinct pairs of oppositions appear with regularity. The first concerns social/interpersonal process – for instance, Quinn's 'Tough Love' or Kaplan's 'Forceful–Enabling' pair. These distinctions are similar to such prior ones as task vs relationship orientations (Fiedler, 1967), autocratic vs participative decision making (Vroom & Yetton, 1973), and the twin pillars of the leader behavior paradigm, initiating structure vs consideration (Fleishman & Harris, 1962). The other opposition has more to do with functional organizational 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*.[3]

1.1.4. Summary

There appear to be two distinct traditions for defining managerial flexibility. We call the first one – flexibility conceived as the behaviorally abstract, general tendency to vary behavior across situations – the trait approach. We call the second tradition the 'mastery of opposites' approach, following Zaccaro who, after reviewing the literature on behavioral complexity theories, noted that 'leader effectiveness entails the *mastery* of countervailing behavior patterns' (2001, p. 134, italics added). Central here is capability and skill with contrasting behaviors that are both functional despite seeming to be mutually exclusive.

These two alternative approaches to conceptualizing flexibility have both common and unique features. Both

traditions view the essence of the construct as changing or varying behavior in response to changing circumstances. Also, both methods deemphasize circumstantial factors and focus instead on managerial actions. However, the domain of behaviors under consideration is different in the two approaches. In the trait view, this domain is broadly conceived and theoretically includes the entire universe of managerial behaviors. The core of the trait-based definition of flexibility is the fact of changing or varying behavior; what particular behaviors are varied is not of concern. As such, the trait definition refers to a general, cross-situational tendency summarized across a range of disparate behaviors. In contrast, specific and narrowly defined behaviors comprise the domain in the mastery of opposites view. There are two opposing behavioral dimensions in the interpersonal sphere – for instance, Accomplishment vs Cohesion in Quinn's model, or Forceful vs Enabling in Kaplan's model – and two in the functional/organizational sphere – Innovation vs Stability, Strategic vs Operational. However, the mastery of opposites view is a higher-order construct, because the unit of analysis is the simultaneous consideration of standing on two opposing dimensions. It is an integrative notion derived from two first-order concepts.

1.2. Measurement and assessment

Performance ratings are the most common method for assessing behavior in organizations and their frequency of use is increasing (Murphy & Cleveland, 1995; Viswesvaran, Schmidt, & Ones, 2002). Thus, although there are other methods for assessing the behavioral flexibility of managers (e.g., assessment centers, biodata, personality), our focus is on coworker ratings.

1.2.1. Trait approach

Ratings of flexibility as construed in the trait approach are usually collected by presenting raters with items that describe summaries of variations in behavior – often without reference to specific behaviors. Sample items in typical 360° surveys include 'adapts to change,' 'demonstrates flexibility,' and 'varies her approach with the situation' (Leslie & Fleenor, 1998). These items are straightforward and flow naturally from a conceptual definition that is targeted at an aggregate level to summarize a generic tendency to vary one's approach with changing circumstances.

1.2.2. Mastery of opposites

There are at least three distinct methods for measuring flexibility as conceived in the mastery of opposites tradition. One is to use a typical Likert-type scale with rating items that refer directly to opposing behaviors (e.g., 'Is tough and at the same time compassionate'; Lombardo & McCauley, 1994). This method, however, is problematic.[4] Therefore, we examine the other two approaches in the present study.

First, we consider what Quinn *et al.* (1992) called the interpenetration method. In this method, raters evaluate performance on items from two separate Likert-type scales representing theoretically opposing dimensions. For instance, Quinn *et al.* (1992) had raters evaluate target managers on items that measure Accomplishment and a separate set of items concerning Cohesion. 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 the contrasting constructs (see 'Method'). The resulting score reflecting balance across Accomplishment and Cohesion provides an operational definition of Quinn's Tough Love interpenetration.

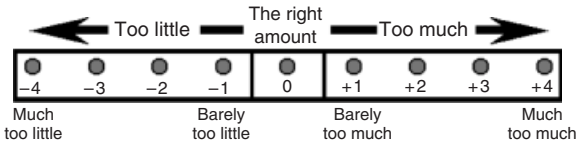
The other method in the mastery of opposites approach is what Kaplan and Kaiser (2003a, b, 2006) call a *duality-oriented* view of versatility. They note that managers tend to over-rely on one side and under-use the other side in dualities like Forceful–Enabling and Strategic–Operational. Thus, their measure uses a new type of response scale that ranges from underdoing to optimal to overdoing (see Figure 1). The reasoning for this response scale format is similar to recently advanced ideas about personality: while the relationship between constructs like general intelligence and performance may be monotonically linear (Coward & Sackett, 1990), relationships between personality and social behavior and performance criteria are likely to be curvilinear (Hogan & Kaiser, 2005; Murphy, 1996; Robie & Ryan, 1999). That is, relative to some criterion (like effectiveness), there may be too little, an optimal amount, or too much of a given managerial behavior (Kaiser & Kaplan, 2005, 2006; Kaplan & Kaiser, 2006).

Similar to the interpenetration method, separate scales are used for opposing dimensions in Kaplan's duality-oriented approach. Although items are rated one at a time on the too little/too much scale, they are scored in pairs (e.g., each Forceful item has an Enabling complement). For each pair, Pythagorean geometry is used to calculate how close the ratings are to 'optimal' on both (Kaplan & Kaiser, 2003b; Kaiser & Kaplan, 2002). 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 can indicate 'lopsidedness' (too much of one, not enough of the other) or 'disengaged/laissez faire' (too little of both).

1.3. The Present study

The purpose of the present study was to empirically compare three alternative methods for assessing the

Note that this rating 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 do too little or do too much of something.



WARNING: Some people misread this scale. Please do not mistake it for the usual type where higher scores are better.

Figure 1. The 'too little/too much' scale in the duality-based measure of versatility. Reproduced from Kaplan and Kaiser (2002), *Leadership Versatility Index*®, with permission from the authors.

behavioral flexibility of managers: the trait method and two distinct methods in the mastery of opposites approach, Quinn's interpenetration method and Kaplan's duality-based method. Below we examine the convergent and discriminant validity of the three methods and both the absolute and incremental validity of each in predicting overall effectiveness. Our interest is in the relative merits of each method for providing feedback to managers. There is virtually no extant literature theoretically comparing these alternative assessment methods. And we are aware of no empirical study comparing them either. Thus, we felt limited in articulating hypotheses about how the three methods would compare. Nonetheless, we expected the following.

First, regarding convergent and discriminant validity, we hypothesized that all three methods would be positively related. To the extent that each is a valid measure of the same phenomenon, the behavioral flexibility construct, they should evince positive correlations.

Hypothesis 1: The trait, interpenetration, and duality-based measures of managerial flexibility will all be positively correlated.

However, we expected that the two mastery of opposites methods, Quinn's interpenetration and Kaplan's duality-based methods, would demonstrate more convergence with each other than either one would converge with the trait-based method. This is because they are borne of the same conceptual foundation: considering performance on two sets of specific, narrowly defined and opposing behaviors simultaneously. In contrast, the trait-based method is more broadly conceived and thus may reflect a larger universe of managerial actions, only some of which overlap with the content of the two mastery of opposites methods.

Hypothesis 2: The interpenetration and duality-based measures will be more highly correlated than either will be correlated with the trait measure.

Another hypothesis about convergent and discriminant validity concerns the distinction between flexibility in terms of how one leads vs what one leads. We expected that Quinn's Tough Love interpenetration would correlate most highly with versatility on Kaplan's Forceful–Enabling duality. Similarly, we expected Quinn's Practical Vision interpenetration to correlate most highly with versatility on Kaplan's Strategic–Operational duality. These hypotheses are derived from the observation that Tough Love and Forceful–Enabling behaviors are in the social/interpersonal domain (how one leads) while Practical Vision and Strategic–Operational behaviors are in the functional/organizational domain (what one leads). This pattern of convergence and discrimination would suggest support for the construct validity of such a distinction.

Hypothesis 3: The interpenetration measure of Tough Love will be most strongly related to the duality measure of Forceful–Enabling versatility; Practical Vision will be most strongly related to Strategic–Operational versatility.

Finally, we had no *a priori* basis for hypothesizing which of the three methods would be most valid in predicting overall managerial effectiveness. However, there is reason to expect the trait measure to demonstrate incremental validity over the two mastery of opposites methods while the reverse may not be true. That is, the trait measure appears to reflect a more broad and general construct, potentially tapping flexibility on the entire universe of more specific managerial behaviors. In contrast, the mastery of opposites measures represent higher-order constructs derived from two narrowly defined first-order constructs – for example, accomplishment and cohesion in the Tough Love interpenetration. Thus, because it is more general, the trait measure should cover some of the same variance as the more precise mastery of opposites

measures plus additional valid variance that is independent of them.

Hypothesis 4: The trait measure of flexibility will yield incremental validity in predicting overall effectiveness above and beyond the validity of the interpenetration and duality-based measures.

2. Method

2.1. Participants

The data for this study were 360° performance ratings from 264 coworkers of 30 upper-middle managers, primarily functional heads (e.g., sales, marketing), employed in three different industrial and financial services firms owned and operated in the United States. Nine of the managers were women, 21 were men. Age and tenure data were not collected, but most were in their 40s and had been with their employer for several years. The ratings were collected as part of a leadership development program. The 264 coworkers included 33 superiors, 93 peers, and 138 subordinates. All raters were informed that the data would be used for developmental feedback and not administrative purposes. Except for superiors, the raters were assured anonymity.

2.2. Measures

The items used to measure the various conceptions of flexibility came 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 multiitem scales; the items are grouped by conceptual similarity.

We created scales out of selected *SKILLSCOPE®* items to represent the trait and interpenetration approaches to flexibility, following Nunnally's (1978, pp. 254–257) recommended method for deriving theoretically and empirically homogenous item clusters. First, we relied on the judgment of three subject matter experts (SMEs). The SMEs first reviewed previously developed scales used to measure conceptually similar constructs. They were also provided with theoretical definitions for each construct. Next they were asked to identify which of the five constructs (i.e., trait flexibility, accomplishment, cohesion, stability, or innovation) each of the 98 *SKILLSCOPE®* was most relevant to, if it was relevant to any of them at all. Finally, SMEs were presented with a list of items identified by two or more of the three judges as relevant to each construct. They were then asked to remove items that had the least in common with the items used in previous scales. Next, we vetted the items identified by the SMEs using item-total correlations and reliability analyses in a large, independent *SKILLSCOPE®* database provided by the Center for Creative Leadership. Finally, we used this independent sample to conduct factor analyses to confirm the adequacy of the theoretically derived and empirically refined measures. First, we conducted a factor analysis using maximum likelihood estimation procedures on the five items selected to measure the trait conception of flexibility. The first factor had an eigenvalue seven times larger than the second factor and accounted for 68% of the total variance, supporting a one-factor interpretation of this item set. Next, a factor analysis using maximum likelihood estimation and oblique rotation on the 28 items selected for the interpenetration method indicated a four-factor solution, according to the scree test, corresponding to the accomplishment, cohesion, stability, and innovation dimensions. The behavioral cores of all items in each scale are shown in Table 1.

The trait measure of flexibility is composed of five items from *SKILLSCOPE®* ($\alpha = .75$). Consistent with our definition and conceptual analysis, these items refer to flexibility and making adjustments in behavior in general, devoid of behavioral or situational specification. The scale is similar in length, reliability, and content to other, conceptually related scales such as the *Campbell Leadership Index* flexibility scale, the *Executive Success Profile* adaptability scale, and *The Profilor* scale, demonstrates adaptability (Leslie & Fleenor, 1998).

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 seven-item scales, labeled Accomplishment ($\alpha = .74$) and Cohesion ($\alpha = .81$), to use in the computation of the Tough Love interpenetration; for Practical Vision, we created two seven-item scales, Stability ($\alpha = .72$) and Innovation ($\alpha = .74$).

To represent the duality-based approach in the mastery of opposites model of flexibility, we used four five-item scales from the *LVI*, a valid and reliable measure developed over a series of psychometric studies (Kaplan & Kaiser, 2002, 2003a, 2006). The four scales were Forceful, Enabling, Strategic, and Operational (αs = .71, .77, 76, and .71, respectively). *LVI* items are rated on a unique response scale that separates underdoing and overdoing as two distinct types of ineffective performance (see Figure 1). The scale ranges from −4 (much too little) to +4 (much too much), with the optimal point in the middle, 0 (the

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

Trait approach	
Flexibility	
1. Varies approach with the situation	
2. Makes adjustments in behavior	
3. Learns from experience; not set in his/her ways	
4. Thinks in terms of trade-offs	
5. Takes ideas different from own seriously	
Mastery of opposites approach	
Interpenetration I: Tough Love	
Accomplishment	*Cohesion*
1. Driven to achieve objectives	1. Delegates
2. Presses for immediate results	2. Shares responsibility with subordinates
3. Good initiative	3. Recognizes and rewards people
4. Decisive	4. Cooperative
5. High energy level	5. Collaborates well
6. Sparks others to take action	6. Deals well with people's feelings
7. Extremely productive	7. Makes good use of people
Interpenetration II: Practical Vision	
Stability	*Innovation*
1. Sets priorities well	1. Creates significant change
2. Spots problems early	2. Generates new ideas
3. Optimistic that problems can be solved	3. Introduces change despite opposition
4. Good troubleshooter	4. Has vision
5. Implements decisions, follows through	5. Conveys a sense of purpose
6. Data-based, rational	6. Seizes new opportunities
7. Gets to the heart of problems	7. 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 do not 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. Visionary
4o. Internally oriented	4s. Externally oriented
5o. Uses structure, discipline	5s. Compelling view of the future

Note: Items used in the trait approach and for the interpenetration methods are from *SKILLSCOPE*® Kaplan (1997) and are reproduced here with permission from the Center for Creative Leadership. Items used in the duality methods from the *Leadership Versatility Index*® Kaplan and Kaiser (2002) and are reproduced with permission from the authors. Forceful (f) and Enabling (e) items (and Strategic [s] and Operational [o] items) with the same number are pairs.

right amount). Although the items are rated one at a time, they were created in pairs to reflect 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.

Scale scores were calculated as the average rating across all items on the given scale. Descriptive statistics, interrater reliabilities, internal consistency reliabilities, and intercorrelations for these scales are presented in Table 2. Internal consistency reliability estimates for each exceeded the .70 minimum recommended by Nunnally (1978). Across all study variables, the average interrater reliability (ICC[1]; Shrout & Fleiss, 1979) was .44, which is comparable with metaanalytic estimates of .39 (Conway & Huffcut, 1997) and .49 (Viswesvaran, Ones, & Schmidt, 1996) for interrater reliability.

Overall effectiveness was the criterion used to compare the validity of the three different methods for measuring flexibility. Overall effectiveness was measured by asking raters, in a semistructured interview conducted separately from the 360° ratings, to 'Please rate this person's overall effectiveness as a manager on a ten-point scale, where 5 is adequate and 10 is outstanding.' Although single-item measures can be deficient and unreliable, the interrater reliability in the present sample was ICC(1) = .51. In conceptual defense of this measure of overall effectiveness, it is worth noting that a global, single-item rating seems befitting as a measure of

Table 2. Descriptive statistics and correlations for study variables

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

Note: $N = 264$. Correlations $> |.20|$ are significant at $p < .001$, $> |.16|$ are significant at $p < .01$, $> |.12|$ are significant at $p < .05$. ICC(1) values are the estimated reliability of ratings from a single rater. Internal consistency reliability estimates appear on the diagonal. Reliabilities for the three interpenetration variables and the overall duality-based versatility variable were calculated using the formula provided by Mosier (1943) for the reliability of a composite; α is presented for all other scales. Scale scores can range from 1 to 3 for Flexibility, Accomplishment, Cohesion, Stability, and Innovation; Forceful, Enabling, Strategic, and Operational scores can range from −4 to +4. See text for computation of scores on Interpenetration (range of 0 to 6) and Versatility (range of .00–1.00) variables.

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

a general, overall evaluation construct. Further, in multi-item scales that measure the overall effectiveness of an individual manager, the highest loading item is usually phrased very similarly to this one (e.g., Hooijberg & Choi, 2000). Thus, the most likely shortcoming of this single-item measure is less reliability than a multiitem scale. The effect of this on the data analysis would be to attenuate external correlations (Hunter & Schmidt, 1990), providing for underestimated relationships between flexibility and overall effectiveness. However, as the purpose of this study is to compare three alternative measures of flexibility, this seems to be a minor concern: criterion unreliability will affect all three correlations equally.

2.3. 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 recommends the formula provided by Bobko and Schwartz (1984) for integrating bipolar 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:

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

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 the 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 (cf. Quinn *et al*., 1992). We used Mosier's (1943) formula to estimate the reliability of these three composite variables, which was .77, .73, and .75 for Tough Love, Practical Vision, and overall interpenetration, respectively.

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 vs 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 this pair of ratings is from the right amount (0) on both items can be derived from the Pythagorean theorem (see Kaplan & Kaiser, 2003b; Kaiser & Kaplan, 2002). It is calculated as:

$$c^2 = a^2 + b^2$$

where *a* is the Forceful rating, *b* the Enabling rating, and *c* the *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% (least versatile) to 100% (most versatile). We computed versatility scores for Forceful–Enabling, Strategic–Operational, and duality-based versatility overall (the mean of the first two scores). We computed the reliability of the versatility scores by calculating the internal consistency of the versatility scores based on the five-item pairs for each duality separately; for Forceful–Enabling, $\alpha = .73$, for Strategic–Operational, $\alpha = .79$. The reliability of the overall versatility composite variable (cf. Mosier, 1943) was .76.

3. Results

The statistical analyses reported below were computed on rater-level data with each variable rated by $N = 264$ raters of a total of 30 target managers. However, because each target manager was rated by an average of nine raters, we first conducted a within- and between-entities analysis (WABA; Dansereau, Alutto, & Yammarino, 1984; Yammarino, 2003) to determine if rater-level analyses were justified. The WABA computations were performed using a computer program developed by O'Connor (2004). According to Yammarino (2003) and Markham and Halverson (2002), WABA procedures are ideally suited to determine whether empirical relationships at multiple levels of analysis are best viewed as a function of whole entities (all raters of a common target manager) or parts contained within those entities (individual raters). Individual rater-level analyses were justified by the WABA results as the average ratio of between-target eta to within-target eta across all study variables was .70, with an average *F*-ratio of $F(234, 29) = .28$, NS, suggesting that differences among individual raters accounted for more covariance among the study variables than did the rating target. This is consistent with findings that individual rater effects are the primary determinant of the latent structure of multisource ratings (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen, Mount, & Goff, 2000).[5]

3.1. Convergent/discriminant validity

We first assessed the convergent and discriminant validity of each method of measuring managerial flexibility. The pattern of correlations in Table 2 is consistent with Hypothesis 1, which predicted that all three methods would be positively related, with observed correlations of .35 (between trait flexibility and overall versatility), .46 (between trait flexibility and overall interpenetration), and .53 (between overall versatility and overall interpenetration; all p's $< .001$). Correcting for unreliability using the ICC(1) values for both variables increased these values, respectively, to .85, 1.16, and 1.14.[6] In other words, there was a high degree of association among the constructs assessed by the three different methods, although the trait measure showed the least overlap with the duality-based versatility measure. Moreover, the trait measure was more strongly related to Tough Love than Practical Vision and to Forceful–Enabling versatility than Strategic–Operational versatility. Raters evidently weigh interpersonal behaviors more heavily than functional business skills to judge flexibility in the abstract.[7]

As predicted in Hypothesis 2, the two measures in the mastery of opposites tradition showed the most convergence ($r = .53$, $p < .001$). However, according to Fisher's *r*-to-*Z* transformation (formula 2.8.5 from Cohen & Cohen, 1983, p. 54), this correlation was not significantly higher than the correlation between the trait and interpenetration measures ($r = .46$; $Z = 1.06$, ns). There was further construct validity evidence for both mastery of opposite methods in that Tough Love and Forceful–Enabling versatility were most strongly

related to one another ($r = .43$; correcting for unreliability, $\rho = 1.03$) while Practical Vision and Strategic–Operational versatility were most related to each other ($r = .50$; $\rho = 1.10$). Consistent with Hypothesis 3, this pattern provides support for the conceptual distinction between the how vs what of leadership style using conventional multitrait/multimethod criteria.

A final observation merits attention. The mastery of opposites approach assumes that contrasting dimensions are in conflict (Zaccaro, 2001). This suggests a negative relationship – a normative trend indicating that as one increases, the other decreases (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-based measures were more consistent with theoretical expectations in that Forceful and Enabling were inversely related ($r = -.49$) while Strategic and Operational were unrelated.[8]

3.2. Concurrent validity

We next compared the three methods of assessing flexibility in terms of predicting ratings of overall effectiveness. The trait measure was least valid ($r = .28$; $p < .001$; 90% Confidence Interval: .19–.37), while the two mastery of opposites methods were quite valid (for overall interpenetration, $r = .52$, $p < .001$; 90% confidence interval: .45–.58; for overall versatility, $r = .64$, $p < .001$; 90% confidence interval: .59–.69). According to Fisher's *r*-to-*Z* tests and 90% confidence intervals, these latter two correlations were significantly higher than the correlation for the trait-based measure ($Z = 3.30$, $p < .001$; $Z = 5.38$, $p < .001$, respectively). And the validity for overall versatility was significantly higher than the validity for overall interpenetration ($Z = 2.08$, $p < .05$). The operational validity for each of these measures (i.e., correlations with overall effectiveness corrected for unreliability in the criterion only) was .38 (trait measure), .71 (overall interpenetration), and .88 (overall versatility). To rule out different levels of unreliability as the reason for the disparities in validity coefficients, we also corrected for the ICC(1) values of the predictors, resulting in ρ estimates of .64 (trait), 1.08 (interpenetration), and 1.25 (versatility). These results suggest that the lower level of validity found for the trait measure is not an artifact of measurement error and is likely due to deficiencies with that method of measuring flexibility.

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; the duality-based versatility model accounted for 42% of that variance. In both, flexibility in functional/organizational terms (Practical Vision, Strategic–Operational) was more related to overall effectiveness than was flexibility in social/interpersonal terms (Tough Love, Forceful–Enabling). A follow-up dominance analysis (Budescu, 1993) quantified this difference as nearly twice as much, indicating that Practical Vision and Strategic–Operational together accounted for 30.0% of the variance in overall effectiveness, whereas Tough Love and Forceful–Enabling accounted for only 16.7%. This may help explain why the trait measure showed comparatively lower validity – it is primarily saturated with social/interpersonal flexibility and less so with the relatively more predictive functional/organizational flexibility.

Table 3. Predicting overall effectiveness with two mastery of opposites methods of measuring flexibility

	β
Interpenetration method	
Tough Love	.24***
Practical Vision	.41***
Model Statistics	$R^2 = .29$ (adjusted $R^2 = .28$), F (2, 261) = 53.92***
Duality-based method	
Forceful–Enabling Versatility	.25***
Strategic–Operational Versatility	.47***
Model Statistics	$R^2 = .42$, (adjusted $R^2 = .41$), F(2, 261) = 92.86***

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

3.3. Incremental validity

Finally, to test Hypothesis 4, we examined the incremental validity of each of the three approaches relative to the other two in predicting overall effectiveness. As all three approaches overlap with one another empirically, we wanted to determine the unique contribution each could make in predicting overall effectiveness. Our prediction was that the broader, more general trait measure would add incremental validity over the mastery of opposites measures, but the reverse would not be true. To test this, we examined three hierarchical regression models, first entering the terms for the first two methods, then testing the additional variance accounted for by the focal method. These results appear in Table 4. Contrary to expectations, the trait measure yielded no incremental validity over the two mastery of opposites approaches. Said differently, the mastery of opposites measures contained all of the criterion-related variance of the trait measure.

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. And the

Table 4. Incremental validity of three methods for assessing flexibility

Test trait approach			Test interpenetration approach			Test duality-based approach		
	β	ΔR^2 (Δadjusted R^2)		β	ΔR^2 (Δadjusted R^2)		β	ΔR^2 (Δadjusted R^2)
Step 1								
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***							
		.467*** (.459***)			.425*** (.418***)			.294*** (.286***)
Step 2								
Flexibility (trait)	.01		Tough Love	.12*		Forceful–Enabling	.18**	
			Practical Vision	.20***		Strategic–Operational	.37***	
		.000 (.002)			.043*** (.039***)			.174*** (.171***)

Note: ***$p<.001$; **$p<.01$; *$p<.05$. Statistics for the full model, including all five predictors: $R^2=.467$, (adjusted $R^2=.457$), $F(5, 258)=45.30$, $p<.001$.

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 demonstrated appreciable convergence, the duality-based method contained about four times more unique variance related to overall effectiveness (i.e., 17.4/4.2 = 4.14). Further, this difference is significant according to a Fisher's *r*-to-*Z* test ($Z=.270$, $p<.01$).

4. Discussion

We compared two conceptual approaches and three measurement methods for assessing the flexibility of managers with coworker ratings for the purpose of developmental feedback. The results offered support for the more complex and behaviorally specific 'mastery of opposites' approach. Although the more commonly used trait approach is simple, shorter, and intuitively appealing it appears deficient in some regards. First, it overlaps only somewhat with the mastery of opposites methods. Specifically, only 27.5% of the observed variance in the trait measure was accounted for by the mastery of opposites measures (see Footnote 2). Moreover, the trait measure had no incremental validity over the mastery of opposites measures. In fact, these more complex measures contained all of the criterion-related variance captured by the trait measure – and then some. Thus, part of our reasoning for Hypothesis 4 seems to hold: the trait measure does indeed capture something other than social/interpersonal flexibility and functional/organizational flexibility. However, whatever the substance of this remaining 72.5% of observed variance is, it is not related to overall effectiveness. Moreover, correcting for measurement error did not change these conclusions, ruling out differences in reliability as an explanation for the weaker performance of the trait approach. Taken together, this suggests that the trait approach is an empirically inferior method.

On a practical note, the trait approach requires only a fraction of the number of items required by either measure of the mastery of opposites model. And even with the smaller number of items, it was significantly correlated with overall effectiveness, although less so than the mastery of opposites measures. However, feedback provided with trait measures is difficult to act on. In effect, low scores suggest improvement lies in 'being more flexible.' This is ambiguous and ignores the conventional, and empirically corroborated, wisdom of providing specific, behavior-focused feedback (Kluger & DeNisi, 1996). 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 precise and, thus, is likely to be more instructive and actionable (Kaplan & Kaiser, 2003b).

There is an interesting theoretical difference between the two conceptual traditions. In the mastery of opposites model, flexibility is viewed as a higher-order construct composed of discreet, narrowly defined, lower-level behaviors. This makes it a more complex construct. The trait model is a less complex, first-order conception because it requires raters to make a direct generalization across behaviors and situations. Of course, which behaviors and how they are combined are not specified, leaving this to raters to

idiosyncratically determine. This difference may help account for the fact that the interrater reliability of the trait measure was the lowest across all study variables. Furthermore, the explicit specification of the behavioral components in the mastery of opposites model makes it the more rich and theoretically elegant approach. Evidently this degree of specificity and richness also translates into more explanatory power as suggested by the stronger correlations with overall effectiveness.

There were several noteworthy results for the mastery of opposites methods. First, there was convergence between the interpenetration and duality-based measures. This was true for the overall scores and 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/organizational-oriented oppositions, Practical Vision and Strategic–Operational versatility (and *vice versa*). This provides construct validity evidence for the mastery of opposites model of flexibility as well as the interpenetration and duality-based measures – the two methods showed the expected pattern of convergence and discrimination.

Both mastery of opposites methods demonstrated substantial validity in predicting ratings of effectiveness. The overall measures correlated with effectiveness between .52 and .64. Moreover, the two component oppositions (representing *how* and *what*) in the interpenetration method as well as in the duality-based method each made unique contributions in predicting effectiveness, together accounting for 29% and 42% of the observed variance, respectively. These qualify as large effects (Cohen, 1988) and indicate that flexibility in terms of social/interpersonal behaviors as well as functional/organizational skills account for much of what it means to be regarded as an effective manager.

4.1. Future research

The results of this single study cannot provide a definitive statement on how to measure managerial flexibility. Taken together, however, these results are suggestive that the mastery of opposites model of flexibility has much to offer practitioners and students of management. We encourage replications and extensions to this research. In particular, there are three areas where further study is needed to advance our understanding of this approach.

4.1.1. Comparing methods

The interpenetration and duality-based methods appear to be superior to the trait-based method. But how do they compare with each other? There are two bases for comparing the measures empirically: in terms of internal structure and relations with external variables (Drasgow, 1984). With regard to external relationships, in this study the duality-based measure both outpredicted overall effectiveness and yielded four times more incremental validity. Of course, more research with a wider range of managerial effectiveness criteria, particularly team process and unit outcome variables, is needed to render a definitive judgment.

Concerning internal structure is the question of how theoretically opposing dimensions are related. An assumption in the mastery of opposites model is that contrasting dimensions are in conflict with one another (Dennison *et al*., 1995; Kaplan & Kaiser, 2003a, 2006; Quinn, 1988; Zaccaro, 2001), which implies a negative relationship. However, we found positive relationships between opposites in the interpenetration data. This is not unique to our data set; research inspired by Quinn's theory consistently reports positive correlations between opposites (e.g., Dennison *et al*., 1995; Hooijberg & Choi, 2000). It seems that either the assumption of a 'conflict between opposites' is wrong or the positive relations are artifacts of the measurement method. To this point, the duality-based measures with the too little/too much response format did not demonstrate positive relationships between opposites. In fact, the theoretical inverse relationship was found between the Forceful and Enabling dimensions. It may be the case that this new response format confers certain advantages over the typical kind (Kaiser & Kaplan, 2005, 2006). We leave it to future research to explicate those advantages as well as the limitations of this type of rating scale.

4.1.2. How vs what distinction

The present data offer validity evidence for a distinction between flexibility in the interpersonal/social domain and flexibility in the functional/organizational domain. A movement is afoot in the leadership literature to recognize and elaborate this fundamental difference (e.g., Antonakis & House, 2002; Zaccaro, 2001). Our results support such a conceptual development.

In two separate measures, we found that flexibility in the functional/organizational domain was related nearly twice as strongly as was interpersonal/social flexibility to overall effectiveness. This suggests that the former is the more important determinant of how managers evaluate one another overall. And it is consistent with Zaccaro and Horn's (2003) observation that 'leadership involves fundamentally the process of direction setting and operational management' (p. 785). Yet, leadership research is dominated by the interpersonal/social *how* and has barely examined the functional/organizational *what*. Hunt (1991), for instance, determined that 90% of this voluminous literature was focused on the former. And the major contingency theories all concern social/interpersonal processes – Fiedler's (1967) explication of which situations favor task or relationship orientations; House's (1971) path–goal model of

when directive, achievement, participative, or supportive behaviors are best; and Vroom and Yetton's (1973) normative model for when to make decisions oneself or to involve others in varying degrees. The field lacks a contingency model relevant to the dilemmas between such organizational issues as efficiency vs innovation, execution vs strategic redirection, short-term results vs long-term viability, and so forth. This represents a ripe opportunity for future research.

4.1.3. Antecedents of flexible behavior

Our focus has been explicitly on using coworker ratings to assess the *behavioral* flexibility of managers. However, the broader topic of managerial flexibility involves more than behavioral considerations. For instance, in the literature on managerial complexity, a distinction is made between cognitive complexity – having highly differentiated and integrated schemata to represent and understand the external environment – and behavioral complexity – having a varied and yet integrated behavioral repertoire (Dennison *et al*., 1995; Hooijberg *et al*., 1997; Hooijberg, 1996; Hooijberg & Quinn, 1992; Zaccaro, 2001). Thus, flexibility involves correctly identifying when to use what behavioral approach and corresponding skill at enacting those responses.

This distinction suggests two future directions. First is the matter of construct validity – articulating a nomological net through empirical associations among theoretically related constructs (Cronbach & Meehl, 1955). The theoretical underpinning of the case for managerial flexibility is complexity theory and its principle of requisite variety (Ashby, 1952): managerial effectiveness is thought to depend on how well managers can mentally represent the complexity of their environment and whether they have a diverse enough repertoire to respond appropriately (Hooijberg, 1996; Zaccaro, 2001). According to the aforementioned theories of managerial complexity, cognitive complexity, the understanding component, is a precursor to behavioral flexibility, the responding component. Thus, the two should be related. Moreover, a distinction has been made between how socially complex and how informationally complex managerial environments are (Osborn, Hunt, & Jauch, 2002; Zaccaro, 2001). Informational complexity involves the number of dimensions requiring attention, the number of sources of data, and the number of alternative data sources whereas social complexity concerns the number and diversity of interpersonal constituents to interact with, coordinate, and influence. Thus, cognitive complexity of the social kind should be most related to flexibility in terms of how one manages in the interpersonal/social domain and cognitive complexity of the informational kind should be most related to flexibility in terms of what one attends to in the functional/organizational sphere.

The distinction between cognitive complexity and behavioral complexity is also important in the practice of management development. Helping managers enhance their behavioral flexibility involves the development of a more sophisticated mental model and the establishment of behavioral skills, two distinct foci for management learning and education (Hogan & Warrenfeltz, 2003). Work on mental models involves enhancing the accuracy of social perceptiveness and judgment – knowing when to use which approach and to what degree (e.g., as specified in contingency theories or local cultural norms; Zaccaro *et al*., 1991b). Skill development involves the routinization and practiced execution of such behaviors (e.g., as established in simulations, practice, and role playing; Wexley, 1984). It follows that a more complete assessment of managerial flexibility for the purpose of development would involve an assessment of managers' knowledge structures, social perceptiveness, and social judgment in addition to their behavioral flexibility. For instance, if behavioral flexibility is less than desired, it is likely that an ill-conceived or incomplete mental model of managing is also in operation (Kaplan & Kaiser, 2006).

4.2. Limitations

Four limitations to the present effort should be noted. First, this study is properly regarded as illustrative and certainly not definitive, if for no other reason than the fact that the data used in these analyses were based on ratings for only 30 target managers. At a minimum, a replication study with a larger sample is warranted. Given the degree of sampling error to be expected in a sample of the present size, point estimates of population parameters reported here (e.g., validity coefficients) should especially be treated as tentative.

Second, the measures used for the trait and interpenetration methods were assembled *ad hoc* from an existing instrument. In one sense, this might not be a limitation – and in fact could be considered as ecological validity – in that it closely resembles how these measures might be used in practice. Nevertheless, it is possible that our results would have been different had we used the scales created by Quinn and colleagues (see Dennison *et al*., 1995) to measure the interpenetration model.

Third, the criterion variable used in the absolute and incremental validity analyses is limited. For one, ratings of overall effectiveness reflect the perceived effectiveness of the individual manager. They do not directly measure the hallmark of effective leadership, the performance of the team or organizational unit over time (Hogan & Kaiser, 2005). Another problem is that we obtained the criterion data from the same source as the predictor data, albeit using a separate method and at a different time. This likely inflated predictor–criterion

relationships. Yet, whatever limitation this poses, it would appear to do so equally in the analysis of each measure. It is difficult to articulate how using separate sources for both predictor and criterion data might produce different comparative results than reported here. Nonetheless, future work on measuring flexibility could benefit from examining a range of objective and subjective effectiveness criteria.

The fourth limitation pertains to the generalizability of our results. The population under study was relatively homogenous – all rating targets were upper-middle managers with functional responsibility in American companies. The hypothesis that flexibility is central to managerial effectiveness is based on the principle of requisite variety (Ashby, 1952). As the nature of managerial work becomes increasingly more socially and informationally complex as one ascends the hierarchy (Osborn *et al*., 2002; Zaccaro, 2001), the strength of the flexibility–effectiveness relationship should, in theory, increase with organizational level. Thus, the magnitude of predictive validity may be notably smaller for supervisors and higher for senior executives. Future research is needed to establish the moderating effect of organizational level or job complexity on the relationships between managerial flexibility and indices of effectiveness.

Acknowledgements

We are grateful to the Center for Creative Leadership for granting permission to reproduce items from their *SKILLSCOPE*® 360° feedback instrument. We also thank Bob Kaplan, David DeVries, Robb Most, Joyce Hogan, Deniz S. Ones, and two anonymous reviewers for constructive suggestions on earlier drafts.

Notes

1. An earlier version of this study was presented at the 19th annual meeting of the Society for Industrial and Organizational Psychology in Chicago, IL, April 2004.
2. We should alert the reader to a potential confusion in labeling this definition of flexibility as a 'trait approach.' Typically, the term trait is used to describe a self-assessed individual difference variable, such as a self-reported personality characteristic or tested intelligence. We have applied the term trait to this particular method of rating behavioral flexibility because the nature of the construct has the core defining characteristics of a trait: an abstract summary of a tendency to behave in a particular manner over time and across diverse situations (James & Mazerolle, 2002). We recognize that the term trait could be assumed to mean a self-reported personality characteristic and emphasize that it does not in this case.
3. This distinction is in the spirit of a growing recognition among leadership scholars. For instance, Zaccaro (2001) has contrasted the direct influence leaders have in social exchanges and interpersonal dynamics with the indirect influence leaders exert through decisions about direction, organizational structure, and objectives. Antonakis and House (2002) discussed the difference between inspirational leadership, which is an interpersonal matter, and instrumental leadership, which concerns setting direction and facilitating the accomplishment of goals. We prefer the more generic labels, the *how* and the *what* of leadership, because they seem to be the more inclusive terms while still describing the essence of this distinction.
4. Low ratings on these 'double-barreled' items can indicate either not tough enough, not compassionate enough, or not enough of either. Thus, ratings on these kinds of items tend to be ambiguous and interrater reliability tends to be dubiously low because different raters attach different levels of significance to the two behavioral components (Kaiser & Craig, 2005).
5. As a follow up to the WABA results, we also conducted all analyses at both the individual rater level ($n = 264$) and at the target level ($n = 30$) by aggregating across all raters of a common target. The substantive pattern of results was nearly identical, although effect sizes were larger in the target-level analyses because of increased reliability due to aggregation. These results are available from the first author upon request.
6. Correcting for unreliability resulted in some values that exceeded unity. This is likely due to sampling error associated with the estimation of interrater reliability values and interrelationships in smaller sample sizes. Corrected correlations greater than 1.00 indicate that a very high degree of association between the measures is likely but precise population estimates are not calculable given the unavoidable sampling error in the study (D. S. Ones, personal communication, October 13, 2006).
7. To further understand the nature of the trait measure, we identified the relative contributions of Tough Love, Practical Vision, Forceful–Enabling, and Strategic–Operational in predicting Flexibility using dominance analysis (Budescu, 1993). This procedure compares partial and semipartial correlations to isolate the unique contribution each variable makes in multivariate prediction. Results indicated that the trait measure is more saturated with variance associated with the interpersonal side of leadership (the *how*) rather than the functional/organizational side (the *what*). Specifically, Tough Love and Forceful–Enabling together accounted for 20.7% of the variance in the trait measure whereas Practical Vision and Strategic–Operational accounted for only 6.8% of that variance.
8. Others have noted the troubling finding that the relationship between supposedly opposing, or at least theoretically independent, leadership behaviors is regularly found to be strong and positive. For instance, Judge, Piccolo, and Ilies (2004) report the corrected true correlation between consideration and initiating structure as rated on the most valid measure of those constructs, the *Leader Behavior Description Questionnaire, Form XII* (Stogdill, 1963), to be $\rho = .46$. Kaiser and Kaplan (2005, 2006) point out that traditional response formats do not

distinguish between too little and too much as unique classes of ineffective performance or between often and *too* often in frequency of occurrence. But when raters use a response scale that does make these distinctions, like in Figure 1, theoretical opposites are often negatively correlated (Kaiser & Kaplan, 2006; Kaplan & Kaiser, 2006, appendix).

References

Antonakis, J. and House, R.J. (2002) An analysis of the full-range leadership theory: the way forward. In: Avolio, B.J. and Yammarino, F.J. (eds), *Transformational and charismatic leadership: the road ahead*. Greenwich, CT: Elsevier Science/JAI Press, pp. 3–33.

Ashby, W.R. (1952) *Design for a brain*. New York: Wiley & Sons.

Bobko, P. and Schwartz, J. (1984) A metric for integrating theoretically related but statistically uncorrelated constructs. *Journal of Personality Assessment*, **48**, 11–18.

Briscoe, J.P. and Hall, D.T. (1999) Grooming and picking leaders using competency frameworks: do they work? An alternative approach and new guidelines for practice. *Organizational Dynamics*, **28**, 37–52.

Budescu, D.V. (1993) Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, **114**, 542–551.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. and Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Conway, J.M. and 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.

Coward, W.M. and Sackett, P.R. (1990) Linearity of ability – performance relationships: a reconfirmation. *Journal of Applied Psychology*, **75**, 297–300.

Cronbach, L.J. and Meehl, P.E. (1955) Construct validity in psychological tests. *Psychological Bulletin*, **52**, 281–302.

Dansereau, F., Alutto, J.A. and Yammarino, F.J. (1984) *Theory testing in organizational behavior: the varient approach*. Englewood Cliffs, NJ: Prentice-Hall.

Davis, B.L., Skube, C.J., Hellervik, L.W., Gebelein, S.H. and Sheard, J.L. (1996) *Successful manager's handbook: development suggestions for today's managers*. Minneapolis, MN: Personnel Decisions International.

Dennison, D.R., Hooijberg, R. and Quinn, R.E. (1995) Paradox and performance: toward a theory of behavioral complexity in managerial leadership. *Organizational Science*, **6**, 524–540.

Drasgow, F. (1984) Scrutinizing psychological tests: measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, **95**, 134–135.

Fiedler, F.E. (1967) *A theory of leadership effectiveness*. New York: McGraw-Hill.

Fleishman, E.A. and Harris, E.F. (1962) Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, **15**, 43–56.

Hogan, R. and Kaiser, R.B. (2005) What we know about leadership. *Review of General Psychology*, **9**, 169–180.

Hogan, R. and Warrenfeltz, R. (2003) Educating the modern manager. *Academy of Management Learning and Education*, **2**, 1, 74–84.

Hooijberg, R. (1996) A multidirectional approach toward leadership: an extension of the concept of behavioral complexity. *Human Relations*, **49**, 917–946.

Hooijberg, R. and 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., Hunt, J.G. and Dodge, G.E. (1997) Leadership complexity and development of the Leaderplex Model. *Journal of Management*, **23**, 375–408.

Hooijberg, R. and Quinn, R.E. (1992) Behavioral complexity and the development of effective managerial leaders. In: Phillips, R.L. and Hunt, J.G. (eds), *Strategic management: a multiorganizational-level perspective*. New York: Quorum, pp. 161–176.

House, R.J. (1971) A path-goal theory of leader effectiveness. *Administrative Science Quarterly*, **16**, 321–338.

Hunt, J.G. (1991) *Leadership: a new synthesis*. Newbury Park, CA: Sage Publications.

Hunter, J.E. and Schmidt, F.L. (1990) *Methods of meta-analysis: correcting for error and bias in research findings*. Newbury Park, CA: Sage.

James, L.R. and Mazerolle, M.D. (2002) *Personality in work organizations*. Thousand Oaks, CA: Sage.

Judge, T.A., Piccolo, R.F. and Ilies, R. (2004) The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, **89**, 36–51.

Kaiser, R.B. and Craig, S.B. (2005) Building a better mousetrap: item characteristics associated with discrepancies in 360° ratings. *Consulting Psychology Journal: Research and Practice*, **57**, 235–245.

Kaiser, R.B. and Kaplan, R.E. (2002) *Leadership Versatility Index® user's guide*. Greensboro, NC: Kaplan DeVries Inc.

Kaiser, R.B. and Kaplan, R.E. (2005) Overlooking overkill? Beyond the 1-to-5 rating scale. *Human Resources Planning*, **28**, 3, 7–11.

Kaiser, R.B. and Kaplan, R.E. (2006) *Are all scales created equal? Response format and the validity of managerial ratings*. Presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX, April.

Kaplan, R.E. (1996) *Forceful leadership and enabling leadership: you can do both*. Greensboro, NC: Center for Creative Leadership.

Kaplan, R.E. (1997) *SKILLSCOPE® for managers*. Greensboro, NC: Center for Creative Leadership.

Kaplan, R.E. and Kaiser, R.B. (2002) *Leadership Versatility Index®*. Greensboro, NC: Kaplan DeVries Inc.

Kaplan, R.E. and Kaiser, R.B. (2003a) Developing versatile leadership. *MIT Sloan Management Review*, **44**, 4, 19–26.

Kaplan, R.E. and 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. and Kaiser, R.B. (2006) *The versatile leader: make the most of your strengths – without overdoing it*. San Francisco: Pfeiffer.

Kluger, A.N. and 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.

Leslie, J.B. and 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. and Eichinger, R.W. (2000) High potentials as high learners. *Human Resource Management*, **39**, 321–329.

Lombardo, M. and McCauley, C.D. (1994) *Benchmarks®: a manual and trainer's guide*. Greensboro, NC: Center for Creative Leadership.

Markham, S.E. and Halverson, R.R. (2002) Within- and between-entity analyses in multilevel research: a leadership example using single level analyses and boundary conditions (MRA). *Leadership Quarterly*, **13**, 35–52.

Mintzberg, H. (1975) The manager's job: folklore and fact. *Harvard Business Review*, **53**, 100–110.

Mosier, C.I. (1943) On the reliability of a weighted composite. *Psychometrika*, **8**, 161–168.

Mount, M.K., Judge, T.A., Scullen, S.E., Sytsma, M.R. and Hezlett, S.A. (1998) Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, **51**, 557–576.

Murphy, K.R. (1996) Individual differences and behavior in organizations: much more than *g*. In: Murphy, K.R. (ed.), *Individual differences and behavior in organizations*. San Francisco: Jossey-Bass, pp. 3–30.

Murphy, K.R. and 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*, A1–A6.

Nunnally, J. (1978) *Psychometric theory* (2nd edn). New York: McGraw-Hill.

Osborn, R., Hunt, J.G. and Jauch, R. (2002) Toward a contextual theory of leadership. *Leadership Quarterly*, **13**, 797–837.

O'Connor, B.P. (2004) SPSS and SAS programs for addressing interdependence and basic levels-of-analysis issues in psychological data. *Behavior Research Methods, Instrumentation, and Computers*, **36**, 17–28.

Pulakos, E.D., Arad, S., Donovan, M.A. and 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. and Hart, S.L. (1992) Integrating the extremes: crucial skills for managerial effectiveness. In: Srivastva, S. and Fry, R.E. (eds), *Executive and organizational continuity*. San Francisco: Jossey-Bass, pp. 222–252.

Robie, C. and Ryan, A.M. (1999) Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment*, **7**, 157–169.

Scullen, S.E., Mount, M.K. and Goff, M. (2000) Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, **85**, 956–970.

Shrout, P.E. and Fleiss, J.L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420–428.

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.

Stogdill, R.M. (1963) *Manual for the leader behavior description questionnaire, form XII*. Columbus, OH: Bureau of Business Research, Ohio State University.

Streufert, S. and Nogami, G.Y. (1989) Cognitive style and complexity: implications for I–O psychology. In: Cooper, C.L. and Robertson, I. (eds), *International review of industrial and organisational psychology*. Chichester: Wiley, pp. 93–143.

Tsui, A.S. (1994) Reputational effectiveness: toward a mutual responsiveness framework. *Research in Organizational Behavior*, **16**, 257–307.

Viswesvaran, C., Ones, D.S. and Schmidt, F.L. (1996) Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, **81**, 557–574.

Viswesvaran, C., Schmidt, F.L. and Ones, D.S. (2002) The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, **87**, 345–354.

Vroom, V.H. and Yetton, P.W. (1973) *Leadership and decision-making*. New York: Wiley.

Wexley, K.N. (1984) Personnel training. *Annual Review of Psychology*, **35**, 519–551.

Yammarino, F.J. (2003) Modern data analytic techniques for multisource feedback. *Organizational Research Methods*, **6**, 6–14.

Zaccaro, S.J. (2001) *The nature of executive leadership: a conceptual and empirical analysis of success*. Washington, DC: American Psychological Association.

Zaccaro, S.J., Foti, R.J. and 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. and Mumford, M.D. (1991b) Leadership and social intelligence: linking social perceptiveness and behavioral flexibility to leader effectiveness. *Leadership Quarterly*, **2**, 317–342.

Zaccaro, S.J. and Horn, Z.N.J. (2003) Leadership theory and practice: fostering an effective symbiosis. *Leadership Quarterly*, **14**, 769–806.