

Poster session presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA, April 2000.

Getting at Leadership Versatility: The Case of the Forceful and Enabling Polarity

Robert B. Kaiser & Robert E. Kaplan
Kaplan DeVries Inc.

Contemporary thinking about executive leadership suggests that versatility is key to effectiveness. This paper describes the development of an innovative measure and an empirical evaluation of a theory of versatility in terms of forceful and enabling leadership styles. Persuasive evidence for the construct validity of the theory and measure is presented. Forceful and enabling leadership were inversely related and few executives showed clear evidence of versatility in their use of forceful and enabling leadership. Further, versatility on the forceful and enabling polarity predicted between 16 to 35% of the variance in coworker ratings of effectiveness. Implications for leadership theory and leadership development are discussed.

Contemporary thinking about senior corporate leadership suggests that behavioral flexibility or complexity is key to executive effectiveness (Denison, Hooijberg, & Quinn, 1995; Hooijberg & Quinn, 1992; Quinn, 1988; Zaccaro, 1999). In this spirit, Kaplan (1996) discussed effective leadership as a function of versatility, the fluid ability to adapt to the competing demands tapped out by the staccato rhythm of a typical executive's day. To explain why some executives seem to lack versatility, he applied the concept of polarization—where potentially complementary roles tend to be treated as incompatible, contradictory, and mutually exclusive.

In demonstrating the usefulness of the idea of polarity in understanding versatility and executive effectiveness, Kaplan (1996) juxtaposed two familiar approaches to leadership. He described how many executives tend to identify with, value, and enact either a self-assertive forceful leadership style or a supportive and people-oriented enabling style. It was further noted that executives who favored one style tended to devalue and avoid the opposing style. In this way, forceful and enabling

leadership were described as a polarity and it was suggested that versatility on these two dimensions—that is, the ability to selectively draw from either mode rather than be habitually drawn to one while avoiding the other—contributed to leadership effectiveness.

The present paper describes an effort to construct a 360-degree approach to measuring versatility in terms of forceful and enabling leadership and to evaluate empirically important aspects of a polarity theory about how these dimensions play out in the executive suite.

Leadership and Versatility

By all accounts, leadership is a social phenomenon and most definitions of leadership have in common the central notion of social influence (Bass, 1990). Psychologists have long regarded interpersonal flexibility—the ability to adjust one's behavior to meet varied interpersonal demands (Leary, 1957)—as a hallmark of healthy, adaptive social functioning (Sullivan, 1953; Vaillant, 1993). Empirical research supports this position. For example, interpersonal flexibility is positively related to adjustment (Paulhus & Martin, 1988). Since the work of corporate leaders takes place in a complex social milieu involving a good deal of interpersonal exchanges, it follows that flexibility is essential to handling effectively the demands of the senior executive role.

Beyond the social and interpersonal nature of leadership, there are many other factors that demand flexibility in effective corporate leadership. The constant changes that characterize the

Robert B. Kaiser, Director of R&D, Kaplan DeVries Inc.; Robert E. Kaplan, Co-President, Kaplan DeVries Inc.

We gratefully acknowledge constructive input for this work from Sally Carr, Bart Craig, David DeVries, Seymour Epstein, Rob Goldberg, Bill Hodgetts, Mike Lombardo, Denise Lyons, Connie McArthur, Chuck Palus, and Amy Webb.

Requests to use the measure described herein and other correspondence about this research can be addressed to: Rob Kaiser, Kaplan DeVries Inc., 1903 G Ashwood Ct., Greensboro, NC 27455. Electronic mail can be sent to rkaiser@kaplandevries.com.

modern global economy coupled with rapid technological innovation and unprecedented levels of demographic diversity require executives to perform varied and highly complex roles. Where markets, products, operations, structures, strategies, and personnel literally can change overnight, there is a huge need for requisite variety in senior leadership skill sets and competencies. Some authorities have even wondered, "Are today's executive positions too big for one person?" (R.J. Campbell, personal communication, August 23, 1996).

Quinn and his colleagues have looked upon cognitive complexity (c.f., Jaques, 1986) as a necessary, but not sufficient, ingredient in effective senior leadership. To these researchers, cognitive complexity makes possible the necessary and sufficient requirement of behavioral complexity in meeting the varied and sometimes paradoxical demands confronted by executives (Denison et al., 1995; Hooijberg & Quinn, 1992; Quinn, 1988; Quinn, Spreitzer, & Hart, 1991). Behavioral complexity is "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). According to this view, effective senior managers have in their behavioral repertoire the ability to serve multiple functions—to wear many "hats"—and do not emphasize certain roles to the neglect of other roles.

The Forceful and Enabling Polarity

Kaplan (1996) took a position similar to Quinn's when he discussed leadership versatility in terms of polarities. Through years of action-research with senior executives, Kaplan observed a pervasive tendency for executives to embrace either an aggressive and self-assertive forceful style or a more modest and relational enabling style. Moreover, senior managers who identified with one style tended to neglect and devalue the other style. Work with extreme cases of these kinds of "lopsided" leaders revealed that aversions to the opposing style were often rooted in fear and sustained by exaggerated beliefs about the negative consequences of that style. Influenced by the work of Jung (1976) and Levinson (1978), Kaplan (1996) identified this pattern of regarding potentially complementary ways of leading as antagonistic as "polarization." He further noted how heavy reliance on one approach and avoidance of the opposing approach seemed to help account for many of his executive clients' performance problems.

Forceful and enabling are not entirely new concepts; they can be understood at the intersection between personality theory and behavioral approaches to the study of leadership. The personality characteristics of leaders described as forceful include aggressive, competitive, critical, dominant, intense, outspoken, self-assertive, and tough. In contrast, enabling leaders' characteristics include appreciative, cautious, caring, compassionate, gentle, mild, responsive to others, and understanding. These two constellations of traits parallel other polarities or dualities: status/dominance and acceptance/nurturance as discussed in interpersonal theories of personality (Hogan, 1983; Leary, 1957; Wiggins, 1991) and the

motivational drives of agency/individuality and communion/relatedness that appear in the developmental literature (Bakan, 1966; Guisinger & Blatt, 1994).

The forceful and enabling pair also share conceptual similarity with the twin pillars of the leader behavior paradigm—initiating structure and consideration, respectively. Years of research have consistently identified these two factors to broadly summarize the vast array of observable leadership behaviors (see Bass, 1990 for review). Initiating structure includes such behaviors as organizing and directing group activity, setting expectations, task-oriented problem solving, evaluating performance, and maintaining standards. Consideration include showing concern for group members' welfare, expressing appreciation, addressing subordinates' concerns, treating subordinates as equals, involving subordinates in decision making processes, putting others at ease, and facilitating team interaction (Bass, 1990).

Kaplan (1996) explained how it is useful for managers to be capable of filling both forceful and enabling roles—in other words, to be versatile:

"On the one hand, leaders need to be forceful—to assert themselves by means of their own intellect, vision, skills, and drive and to push others hard to perform. Forceful leaders take charge, very much make their presence felt, make it crystal clear what is expected, let very little deter them from achieving objectives, step up to tough objectives, and so on.

On the other hand, leaders need to be enabling—to tap into, bring out, and show appreciation for the capabilities and intensity of other people. Enabling leaders do a great job of involving their people and of opening themselves to their influence—in setting the strategic direction and in making those decisions that affect the unit as a whole. And they give their subordinates plenty of latitude to do their jobs. They invest in their people's development and make sure that they feel valued (Kaplan, 1996, p. 6)."

Polarizing and overdoing/underdoing. It is not difficult to identify leaders who value and fill either the forceful or enabling role to a fault. In these same leaders, if you look closely, you will usually find a void. The opposing role tends to get neglected—even actively avoided and devalued. This polarizing tendency can be seen as limiting a leader's range of behaviors, his versatility: one role gets chronically "overdone" or taken to an extreme while the complementary role remains chronically unfulfilled or "underdone." Kaplan (1996) described how many executive performance problems could be linked to a polarized pattern of overdoing and underdoing on forceful and enabling leadership. For example, he described an overly forceful executive who was spread thin from personally solving his unit's problems and, on the enabling side, he failed to delegate decisions or grant his subordinates a chance to lead. An overly enabling executive, on the other hand, involved others in the

decision making process too much and gave too much air-time to others; on the forceful side he wasn't decisive enough and failed to take strong stands in representing his unit's needs.

This view of performance problems squares with a wealth of data that indicates both consideration and initiating structure are positively and independently related to leader effectiveness (see summary in Bass, 1990). It is also consistent with the finding that the strengths which propel a manager up the career ladder can become overly relied on and undermine performance at more senior levels (McCall, Lombardo, & Morrison, 1988; House & Aditya, 1997). What is novel here is the explicit linkage between overdoing and underdoing, a view that simultaneously considers forceful and enabling leadership as intimately related in a unified polarity.

The link between overdoing and underdoing on forceful and enabling may seem counterintuitive in light of the frequently observed positive correlation between consideration and initiating structure (see review in Bass, 1990). However, this observation may be due to an artifact of the measurement and/or data analysis process. Almost all measures of leadership employ similar response scales, whose anchors often carry an implicit assumption of linearity. The assumption is that more of a given behavior is better (e.g., 1 = does a poor job, 5 = does very well, if not excellent). Scales of this type are adequate for measuring behaviors that are "underdone" (with a low score) but they may be ambiguous—perhaps even silent—about behaviors that are overdone. Of course, not all scales carry such implicit assumptions; some response scales are purely descriptive (e.g., 1 = does not do, 5 = does often, if not always). But scores on these scales are almost exclusively related to other variables by use of the product-moment correlation, a method that assumes linear relationships. Again, the implicitly curvilinear idea that a particular behavior can be done too little, optimally, or too much relative to a given criterion cannot be accounted for under these conditions.

For these reasons, Kaplan (1996) broke from tradition in constructing a measure to get at his polarity theory of forceful and enabling leadership. The idea, which we describe below, was to create a measurement system that could account for grounded observations of practicing executives by reflecting behaviors that range from being underdone to being overdone. Furthermore, to represent versatility on the forceful and enabling polarity, we created a measure that considers simultaneously the degree of overdoing and underdoing on both forceful leadership and enabling leadership.

Present Study

The remainder of this paper rests on the idea that measurement and theory development should go hand in hand (Cronbach & Meehl, 1955; Loewinger, 1957). We report results from an initial attempt to generate evidence of construct validity for a measure of the forceful and enabling polarity theory. First examined are the internal psychometric properties of two sets of items intended to operationalize forceful leadership and

enabling leadership. After refining the item pool, we create forceful and enabling scales with satisfactory measurement properties and use them to also derive an index of leadership versatility. These measures are then used in an empirical evaluation of the theory of the forceful and enabling polarity.

Four substantive hypotheses about the forceful and enabling polarity theory are evaluated here. *H1*: There is a strong negative relationship between forceful and enabling leadership—that is, the more forceful an executive tends to be, the less enabling he or she also tends to be (and vice-versa). *H2*: Overly forceful executives are more prevalent than overly enabling executives. *H3*: Only a minority of executives are versatile or well-balanced in their use of forceful and enabling leadership. *H4*: Finally, and perhaps most importantly, polarization, or a lack of versatility, on forceful and enabling leadership is negatively related to effectiveness.

Method

Participants

The present data were gathered as part of a leadership assessment for a total of 86 executives. Forty-seven were vice-presidents and higher management in two large U.S. firms involved in separate group-based leadership development programs ($n_s = 24$ and 23). The remaining 39 participants held titles ranging from vice-president to CEO and represented 17 different U.S. firms. Each of these participated in an individually tailored leadership development process (described in Kaplan, 1998). The participants were mostly male ($n = 76$) and between 38 to 60 years of age. Ratings on forceful and enabling leadership were collected from a total of 84 self-raters, 123 superiors (who rated 78 target leaders), 300 peers (78 targets), and 389 subordinates (82 targets).

Measures

Forceful and enabling leadership. Kaplan (1996) presented a list of behavioral descriptions indicative of forceful and enabling leadership. That original set of items has been refined and the current pool consists of 11 pairs. Items were generated by consulting actual case material from former leadership development clients to determine conceptual content and wording. Items were constructed in pairs such that the content for a forceful item seemed to complement its enabling counterpart (e.g., "Pushes people hard; holds them accountable" and "Is understanding when people are not able to deliver"). The behavioral core and descriptive statistics for the 11 item pairs, aggregated within rating sources for each target executive, are presented in Table 1.

Rather than a traditional Likert-type rating scale, an implicitly curvilinear scale was created to capture underdoing and overdoing. This scale ranged from .5 to 3.5 in .5 increments with the following anchors: 1 = too little, 2 = just right amount, and 3 = too much. Respondents were alerted twice about the uniqueness of this response scale and to the fact that the highest

Table 1
Forceful and Enabling Item Descriptive Statistics and Inter-rater Agreement by Rating Source

Domain Item	Rating Source											
	Superiors (n = 78)			Peers (n = 78)			Subordinates (n = 82)			Self (n = 84)		
	M	SD	*Mr _{wg}	M	SD	*Mr _{wg}	M	SD	*Mr _{wg}	M	SD	
<i>Forceful</i>												
1f. Leads visibly	1.89	.35	.79	1.94	.38	.69	1.88	.31	.74	2.01	.44	
2f. Declares self	2.04	.49	.75	2.06	.37	.73	2.00	.36	.74	2.26	.52	
3f. Makes tough calls	1.84	.34	.77	1.93	.29	.73	1.84	.29	.77	1.85	.40	
4f. Makes judgments	1.93	.32	.77	1.95	.31	.74	1.96	.26	.70	1.93	.40	
5f. Competitive	2.13	.32	.75	2.18	.28	.71	2.15	.29	.78	2.27	.43	
6f. Pushes people hard	1.90	.38	.70	1.98	.37	.65	2.01	.35	.78	2.06	.51	
7f. Forces issues	1.88	.38	.76	1.92	.36	.65	1.88	.34	.71	1.97	.52	
8f. Decisive, quick	1.89	.33	.77	1.90	.27	.73	1.84	.32	.68	1.99	.47	
9f. Confident	2.00	.30	.71	2.06	.27	.67	2.05	.19	.79	2.04	.33	
10f. Speaks up to superiors	1.90	.39	.86	1.92	.31	.72	1.89	.25	.82	2.07	.48	
11f. Builds own capability	1.94	.28	.92	1.93	.20	.77	1.95	.20	.83	1.82	.37	
<i>Enabling</i>												
1e. Enables subordinates	1.83	.40	.85	1.76	.35	.68	1.80	.29	.77	1.85	.41	
2e. Receptive	1.77	.29	.82	1.74	.26	.75	1.80	.21	.79	1.87	.34	
3e. Compassionate	1.86	.33	.84	1.79	.29	.73	1.73	.29	.73	1.93	.40	
4e. Shows appreciation	1.86	.27	.85	1.81	.23	.79	1.74	.27	.79	1.82	.37	
5e. Team player	1.83	.32	.86	1.78	.24	.76	1.93	.18	.85	2.00	.30	
6e. Understanding	1.97	.41	.75	1.85	.31	.73	1.84	.32	.74	1.83	.50	
7e. Fosters harmony	1.76	.40	.78	1.71	.34	.63	1.71	.28	.74	1.77	.39	
8e. Gets input	1.89	.27	.84	1.85	.22	.76	1.85	.19	.79	1.90	.32	
9e. Modest	1.85	.32	.82	1.80	.30	.66	1.82	.26	.79	1.88	.41	
10e. Responsive to superiors	2.03	.31	.72	2.05	.27	.75	2.08	.25	.76	1.95	.44	
11e. Recruits strong people	1.83	.29	.85	1.83	.23	.79	1.88	.15	.87	1.96	.28	

Note: Item pair complements have a common number, letter *f* represents Forceful and *e* Enabling domains. Only the behavioral core of items are presented. *Mr_{wg} is the average inter-rater agreement estimate calculated across ratings for each focal manager who was rated by two or more coworkers representing the respective rating source (superiors, n = 25; peers, n = 70; subordinates, n = 81).

rating is not necessarily the best rating. The scale and rater instructions are shown in Figure 1.

It is worth noting that this scale requires raters to make a value judgment as to what constitutes “too much” or “too little” for a particular item. Since managers vary in the implicit mental models of effective leadership used to guide their judgments (Lord & Maher, 1993; Sivasubramaniam, Kroeck, & Lowe, 1997), it is possible that these value-laden assessments are purely “in the eye of the beholder.” If this were the case, then our measurement system would be of little value as a research or feedback instrument because the ratings would say more about the raters than the focal leader. Thus, our analysis included close attention to inter-rater agreement.

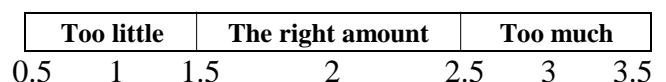
Whereas inter-rater reliability—the most commonly used index of similarity of ratings within a group—assesses rating congruence in rank-ordering or correlational terms, inter-rater agreement provides information on how similar ratings are in terms of overall level (Fleenor, Fleenor, & Grossnickle, 1996; James, Demaree, & Wolf, 1984; 1993). To illustrate the difference, consider two hypothetical raters who evaluated the same manager on three items. Suppose the first rater gave scores of 1, 1.5, and 2 and the second rater gave scores of 2, 2.5, and 3. The inter-rater reliability of these two sets of ratings (i.e., the correlation between them) would be perfect, 1.00. However,

inter-rater agreement (i.e., level of agreement) would tell a different story: the mean of the first rater's evaluations would be 1.5 whereas that for the second rater would be 2.5. On our scale, the first set of ratings would indicate underdo whereas the second set would suggest overdo. We reasoned that items should

Figure 1. *Implicitly “curvilinear” response scale and instructions used to tap underdo and overdo.*

Please rate the manager in question on each of the following aspects of executive leadership.

Please note that the scale is probably different from scales that you are accustomed to using. On this scale the best score is “2,” smack in the middle of the scale. The premise is that performance problems arise when managers either underdo or overdo something.



WARNING: Some people misread this scale. Please do not mistake it for the usual type where a high score is the best.

only be used in this survey if raters of the same manager give ratings that are roughly equivalent in overall level (high inter-rater agreement). This could be taken as evidence that, despite differences among raters in their personal theories of effective leadership, they are able to reach a reasonable degree of consensus about whether the focal leader does too much, optimally, and too little with respect to a given executive role.

Effectiveness. Coworker ratings of overall effectiveness were used to test the hypothesis that versatility on the forceful and enabling polarity is related to higher performance. Raters were asked in a semi-structured interview conducted at a different time from the forceful and enabling rating task to “Please give a rating of X’s overall effectiveness as an executive on a ten-point scale, where 10 is outstanding and 5 is adequate.” Effectiveness ratings were only available for a subsample of 62 participants because the 24-person group program did not include these assessments. Effectiveness ratings were collected from a total of 58 self-raters, 98 superiors (who rated 54 target leaders), 218 peers (53 targets), and 299 subordinates (56 targets).

Although single-item measures are not inherently flawed (Judge & Ferris, 1993), they are often suspected to lack adequate measurement characteristics. We looked at the psychometric properties of the effectiveness ratings in terms of inter-rater agreement and inter-rater reliability within rating sources (cf. Fleener et al., 1996) and convergent validity between rating sources. Inter-rater agreement for each rating source was assessed with James’ r_{wg} statistic (James et al., 1984;

1993), calculated separately for each target where two or more raters provided data. This index is appropriate when a group of raters rate a single target on a single variable or construct and the researcher wants to know the extent to which the overall level of ratings is similar across the individual raters. Similar to the interpretation of indices of reliability, r_{wg} values closer to 1.00 indicate better measurement properties.

Inter-rater reliability was estimated with intraclass correlations (ICCs; Shrout & Fleiss, 1979). ICCs were calculated for the median number of raters per source. ICCs were computed as the reliability of the mean rating for a random sample of two superiors per target (where possible) and random samples of four peers and five subordinates—for superiors (ICC[2,2]), for peers (ICC[4,4]) and for subordinates (ICC[5,5]) (see Shrout & Fleiss, 1979). Estimates for the reliability of a single rater (e.g., ICC[1,4] in the case of peer data) are also presented. As reported in Table 2, this aggregated single item rating easily exceeded minimum standards of agreement and reliability (i.e., $> .70$; Nunnally, 1978; James et al. 1993).

Convergent validity evidence is demonstrated in the between-rating source correlations (all are positive and significant). Further, these measurement properties are in line with meta-analytic estimates of ratings of middle managers on multiple item scales (Conway & Huffcut, 1997). Taken together, we interpreted this data as support for the validity of this multiple source, single-item rating as a general perceived effectiveness construct.

Table 2
Descriptive Statistics and Validity Evidence for Effectiveness Ratings

Rating Source	<i>N</i>	<i>M</i>	<i>SD</i>	ICC (single) ¹	ICC (mean) ²	Self	Superiors	Peers	Subs
Self	58	7.54	1.18	--	--	--			
Superiors	54	8.02	1.21	.57***	.84***	.33**	(.81)		
Peers	53	7.72	1.01	.43***	.75***	.31*	.67***	(.80)	
Subordinates	56	7.98	.87	.56***	.87***	.31*	.43***	.29*	(.81)

Note. Coefficients along the diagonal are Mr_{wg} , computed as the average r_{wg} across target executives within coworker rating sources (superiors, $n = 24$; peers, $n = 51$; subordinates, $n = 59$). Intraclass correlation coefficients are presented as estimates for ¹the reliability of a single rating and ²for the mean of n raters (superiors $n = 2$, peers $n = 4$, subordinates $n = 5$).

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

Results

Measurement Properties

The strategy used to evaluate and refine our measure of forceful and enabling leadership involved considerations of inter-rater agreement, internal consistency (α) reliability, “complement correlations” (described below), and the clarity of underlying factor structures. All psychometric analyses were conducted separately on ratings from the self, superiors, peers, and subordinates. Results for self-ratings are presented, but our focus was on coworker ratings because coworker ratings tend to be the more valid (Conway & Huffcut, 1997) and the survey is designed to be used primarily as a 360-degree feedback tool.

All 22 items were first evaluated separately in terms of inter-rater agreement to determine whether raters agreed on what the focal leader does too little, just right, and too much. James’ r_{wg} statistic was used for this purpose (James et al., 1984; 1993). Separate r_{wg} values were computed for each focal leader for each rating source where two or more coworkers provided ratings. The mean r_{wg} for each item computed across rating targets is presented in Table 1. According to conventional standards (e.g., James et al. 1984; 1993; Nunnally, 1978), these values generally indicate an acceptable level of agreement within rating groups. Despite the unconventional and highly subjective nature of this rating scale, coworkers reached a good deal of consensus about the extent to which target leaders overdo and underdo forceful and enabling leadership.

In the next step, we evaluated the extent to which the item pairs functioned as a “sub-polarity”—that is, the extent to which each pair of items could be said to complement one another yet tend to be enacted in an either-or way by practicing executives. Two criteria were adopted, one considered necessary and the second a more lofty ideal that was relaxed. First, we retained only item pairs that demonstrated a moderate to high negative inter-correlation which would suggest that managers who overdo one of the pair tend to underdo the other. Second, we desired items to be most negatively related to their complements, but pairs were retained if their inter-item correlation was within .10 points of the highest cross-scale correlate. This would suggest that we had a fairly solid matching of forceful and enabling items and that feedback with this instrument could be used at the item pair level of analysis to more clearly target and hone in on developmental needs. These *complement correlations* are available from the first author upon request.

An iterative series of internal consistency (α) reliability estimates were calculated separately for the forceful and enabling scales. Items that either lowered the α value and/or demonstrated an item-total $r < .30$ averaged across the three coworker rating sources were dropped and the analysis was performed again. This was done twice, with results suggesting five items (and their complements, to maintain balance in the total item pool) should be dropped.

The results of the complement correlation analyses and the reliability analyses yielded a good deal of convergence in suggesting which item pairs worked in a way consistent with the underlying theory and also passed psychometric standards. A total of six item pairs were dropped because at least one of the pair did not meet our criteria of contributing to the scale’s α reliability and/or exhibiting a moderate to high complement correlation across rating sources.

Table 3
Forceful and Enabling Item Factor Structures Across Rating Sources

<i>Superiors</i>			<i>Peers</i>		
Item	Factor		Item	Factor	
	I	II		I	II
1f. Leads visibly	<u>.81</u>		1f. Leads visibly	.40	<u>.82</u>
2f. Declares self	<u>.65</u>		2f. Declares self	.52	<u>.86</u>
3f. Makes tough calls	<u>.69</u>	-.44	3f. Makes tough calls	.44	<u>.61</u>
6f. Pushes people hard	<u>.81</u>		6f. Pushes people hard	<u>.71</u>	<u>.63</u>
7f. Forces issues	<u>.80</u>		7f. Forces issues	.63	<u>.67</u>
1e. Enables subordinates		<u>.46</u>	1e. Enables subordinates	<u>-.55</u>	-.41
2e. Receptive		<u>.62</u>	2e. Receptive	<u>-.49</u>	
3e. Compassionate		<u>.85</u>	3e. Compassionate	<u>-.82</u>	
6e. Understanding	-.54	<u>.57</u>	6e. Understanding	<u>-.74</u>	-.47
7e. Fosters harmony		<u>.74</u>	7e. Fosters harmony	<u>-.69</u>	-.45
Eigenvalue	3.78	1.56	Eigenvalue	4.91	1.22
% of Variance	37.8	15.6	% of Variance	49.1	12.3

<i>Subordinates</i>			<i>Self</i>		
Item	Factor		Item	Factor	
	I	II		I	II
1f. Leads visibly		<u>.74</u>	1f. Leads visibly	<u>.69</u>	
2f. Declares self		<u>.80</u>	2f. Declares self	<u>.71</u>	
3f. Makes tough calls		<u>.76</u>	3f. Makes tough calls	<u>.54</u>	-.37
6f. Pushes people hard	.47	<u>.47</u>	6f. Pushes people hard	<u>.38</u>	<u>-.42</u>
7f. Forces issues		<u>.91</u>	7f. Forces issues	<u>.72</u>	
1e. Enables subordinates	<u>-.58</u>		1e. Enables subordinates		<u>.56</u>
2e. Receptive	<u>-.78</u>		2e. Receptive		
3e. Compassionate	<u>-.92</u>		3e. Compassionate		<u>.47</u>
6e. Understanding	<u>-.58</u>	-.35	6e. Understanding		<u>.70</u>
7e. Fosters harmony	<u>-.72</u>		7e. Fosters harmony		
Eigenvalue	5.47	1.51	Eigenvalue	2.18	1.53
% of Variance	54.7	15.1	% of Variance	29.3	16.1

Note: Factor structures based on Principal Axis Factoring with oblique rotation. Factor loadings $< |.35|$ are suppressed.

The five item pairs remaining were factor analyzed using a principal axis method. The factor solution was rotated using an oblique method to facilitate interpretation. Two- and three-factor solutions appeared tenable by Scree tests and the Kaiser criterion (i.e., retain eigenvalues > 1.0). Across all rating sources, the two-factor solutions were most easily interpreted (see Table 3; note that the solution for the peer data is less clear). Forceful items defined one and enabling items defined the second factor. The pattern of factor loadings clearly suggests two distinguishable factors, but there are several moderate to high negative cross-loadings. This was expected for two reasons. First, polarization assumes inverse relations between the factors. Second, with our criteria for complement correlations, we preselected item pairs that were negatively related which, given the high alpha reliability, necessarily translated into negative relations between items and opposing factors.

In light of the above analyses, we concluded that of the original 11 item pairs, five (see Table 4) demonstrated solid structural validity evidence for scales measuring forceful and enabling leadership. Averaging across the three coworker rating sources, inter-rater agreement [$r_{wg(j)}$] at the five-item scale level was .88 for both scales, the mean internal consistency reliability estimate (α) was .84 (forceful) and .78 (enabling), and the two factors—roughly equivalent in size after rotation—accounted for approximately 62% of the common variance among the 10 items.

Table 5
Forceful and Enabling Scale Descriptive Statistics, Reliabilities, Inter-rater Agreement, and Correlations

Rating Source	N	Forceful					Enabling				r Forceful & Enabling
		M	SD	α	$Mr_{wg(j)}$	M	SD	α	$Mr_{wg(j)}$		
Superiors	78	1.91	.30	.81	.85	1.84	.25	.70	.80	-.45***	
Peers	78	1.97	.28	.83	.89	1.77	.23	.80	.90	-.70***	
Subordinates	82	1.92	.28	.89	.91	1.77	.22	.85	.93	-.62***	
Self	84	2.03	.33	.73	--	1.85	.25	.56	--	-.33**	

** $p < .01$ *** $p < .001$

Theory Evaluation

Coworker ratings for the five forceful and five enabling items were averaged (separately for superiors, peers, and subordinates) to create overall forceful and enabling scale scores. These scores were used to further evaluate the construct validity of the present measures vis-à-vis the forceful and enabling theory. Recall that a score of 2.0 on the response scale indicates an optimal amount of the behavior; scores lower than 2.0 represent underdoing and scores higher than 2.0 indicate overdoing. Table 5 contains descriptive statistics, α reliability estimates, $Mr_{wg(j)}$ inter-rater agreement values, and inter-correlations between these two scales.

Hypothesis 1. Consistent with Hypothesis 1, there was a strong negative correlation between the forceful and enabling scales. Correlations ranged from -.45 to -.70 across coworker rating

sources (p 's < .001), suggesting a fairly strong effect. Managers' self-ratings also showed this polarizing tendency, but here it was less pronounced ($r = -.33, p < .01$). In a sense, we stacked the deck to favor this result by imposing the complement correlation criteria on the selection of items. Using scales created from the original 11 item pairs, the same pattern of correlations were observed but at a lower magnitude (r s ranging from -.35 to -.56, p s < .002, for coworker data, $r = -.20, ns$, for self ratings).

Table 4
Items Retained After Psychometric Analyses

Forceful	Enabling
1f. Leads visibly	1e. Enables subordinates
2f. Declares self	2e. Receptive to others' ideas
3f. Makes tough calls	3e. Compassionate
6f. Pushes people hard	6e. Understanding about limits
7f. Forces issues	7e. Fosters harmony

Hypothesis 2. To determine whether or not overly forceful executives are more prevalent than overly enabling executives, the present sample was categorized into leadership types. The operational definition of these types were: *Too Forceful* = Forceful > 2.0 and Enabling < 2.0; *Too Enabling* = Forceful < 2.0 and Enabling > 2.0. Many of the present sample had scores lower than 2.0 on both scales—underdoing both forceful and

enabling—and thus were classified as “disengaged.” This was not expected; earlier writing did not discuss managers who tend to underdo on both dimensions. No executives were rated by coworkers as “overdoing” on both forceful and enabling leadership, which constitutes further evidence for the construct validity of the measure and theory.

As can be seen in the style frequency counts in Table 6, overly forceful types outnumber overly enabling types as rated by all three coworker groups. Chi-square analyses revealed that the observed frequencies are significantly different for superior, peer, and subordinate ratings ($\chi^2(2) = 7.00, p < .03, \chi^2(2) = 13.23, p < .01, \text{ and } \chi^2(2) = 8.37, p < .02, \text{ respectively}$). These data provide clear support for Hypothesis 2, that forceful executives are the more common and enabling executives the more rare exception. The presence of many disengaged leaders in this sample poses a set of questions that are addressed in the discussion section.

Table 6
Frequency Counts for Leadership Types

Leadership Style	Rating Source							
	Superiors		Peers		Subordinates		Self	
	n	%	n	%	n	%	n	%
Too Forceful	36 ^a	46.2	40 ^a	51.3	33 ^b	40.2	39 ^a	48.8
Disengaged	21 ^b	26.9	24 ^b	30.8	34 ^b	41.5	20 ^b	25.0
Too Enabling	21 ^b	26.9	14 ^c	17.9	15 ^a	18.3	21 ^b	26.3
<i>N</i>	78		78		82		80*	

Notes: Frequencies with different superscripts within rating groups are significantly different ($p < .05$). *Four self-raters scored > 2.0 (overdo) on both Forceful and Enabling and are not included in this analysis.

Hypothesis 3. Table 7 contains descriptive data for an inverse measure of versatility on the forceful and enabling polarity. This *polarity index* is not measured directly; it is derived from joint consideration of forceful and enabling scale scores in two-dimensional space. Conceptually, it represents the “distance” an executive is from being perfectly balanced on forceful and enabling (a score of 2.0 on each scale). For example, in the upper left hand corner of Figure 2 is a plot for a leader who scored in the “too much” region on forceful and in the “too little” region on enabling. The geometric distance this person is from being perfectly balanced on the polarity—that is, from a score of 2.0 on forceful and 2.0 on enabling—can be derived from the Pythagorean theorem and computed as: $c^2 = a^2 + b^2$ where $a = (\text{Forceful} - 2.0)$, $b = (\text{Enabling} - 2.0)$, and $c = \text{polarity index}$. We interpret the polarity index as a measure of lopsidedness on the forceful and enabling polarity because it represents the degree of imbalance on the dimensions.

Variance on the polarity index can be recovered through quadratic curvilinear regression analysis using the forceful and enabling scales and the cross-product interaction term as predictors (R^2 values range from .85 to .99; models available upon request). In other words, this simple distance metric can be used as an alternative in conveying the information contained in a much more sophisticated—and difficult to communicate—statistical process. This is especially important to keep in mind

when it comes to interpreting and making sense of feedback on this instrument with executives. The remaining analyses were conducted using correlations with the polarity index and using curvilinear regression with forceful and enabling scores and their cross-product. Since essentially the same substantive results were observed, only the polarity index results are reported.

Polarity indices can range from 0.0 (2.0 on both scales) to 2.12 (3.5 on one and 0.5 on the other scale). Inspection of the polarity index distributions (see Table 7 and Figure 3) revealed a pattern consistent with Hypothesis 3, that few executives are versatile in their ability to enact forceful and enabling leadership. Across rating sources, the mean polarity index was about 1.5 Standard Deviations higher than 0.00 (not overdoing or underdoing forceful or enabling) and the distributions were positively skewed, meaning that only a minority of executives in this sample had scores that were close to 2.0 (i.e., “the right amount”) on both forceful and enabling leadership.

We looked at this data in another way to more clearly understand this point. This was done by classifying as versatile all executives whose polarity index was less than six Standard Errors of Measurement greater than 0.0, which approximates the point below which one can conclude with 99% statistical confidence that the leader’s “true score” (the score she would receive if there were no measurement error) on the polarity index may be 0.0 (Nunnally, 1978), in the dead center of Figure 2. Across all rating sources, the number of executives meeting this criterion was very low. For superior, peer, subordinate, and

Figure 2. Two-dimensional conceptual space defined by forceful and enabling leadership with example of polarity index.

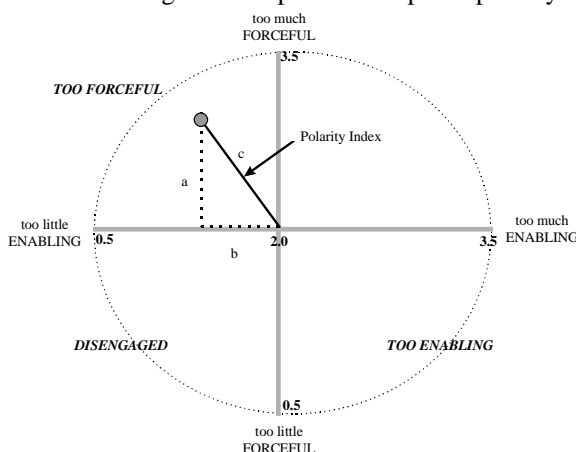
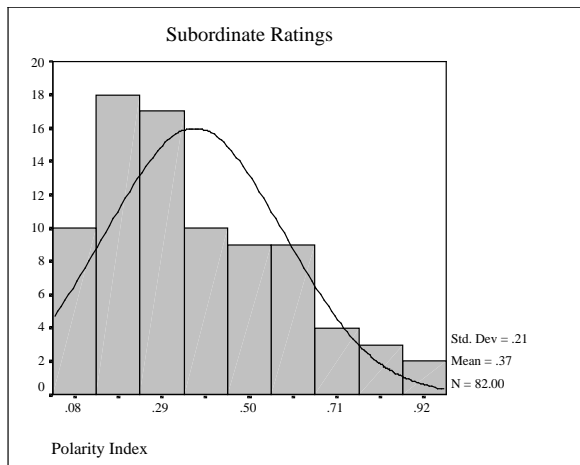


Table 7
Polarity Index Descriptive Statistics and Correlations with Effectiveness

Rating Source	Polarity Index			N^1	$r_{\text{polarity \& effectiveness}}$
	<i>N</i>	<i>M</i>	<i>SD</i>		
Superiors	82	.39	.20	54	-.58***
Peers	78	.36	.22	53	-.41***
Subordinates	78	.37	.22	56	-.53***
Self	84	.37	.21	58	-.09

Notes: Polarity Index descriptive statistics based on full sample. ¹Correlations between Polarity Index and effectiveness based on subsample for which effectiveness ratings were available. *** $p < .001$

Figure 3. Frequency distribution of Polarity Index for subordinate ratings.



self-ratings the number of executives included were 7 (9.0%), 4 (5.1%), 3 (2.4%), and 0, respectively. We take this as evidence supporting hypothesis 3, as it indicates that most executives are to some extent lopsided in their leadership style.

Hypothesis 4. If versatility is key to effectiveness and many performance problems are the result of a habitual reliance on one way of leading to the neglect of other methods, then the polarity index should display a negative correlation with coworker ratings of effectiveness. This pattern is clearly evident across all coworker rating sources with correlations ranging from $-.41$ to $-.58$ (see last column in Table 7), thus supporting Hypothesis 4. This indicates that between 16 to 35% of the variance in coworker ratings of overall effectiveness are directly linked to versatility on the forceful and enabling polarity. Interestingly, self-ratings of effectiveness are not related to versatility derived from self-ratings of forceful and enabling leadership, suggesting that executives tend to not see a link between their own lopsidedness and performance problems. This is consistent with Kaplan's (1996) observation that lopsided leaders rationalized their preference for forceful or enabling leadership.

Discussion

We undertook this empirical study to serve two purposes. The first was to validate a measure of forceful and enabling leadership for developmental work and applied research with executives. The second goal was to put the qualitative analysis and inductive theory laid out by Kaplan (1996) under the lens of quantitative scrutiny. These data satisfied both goals and also provide useful information to guide further articulation of the forceful and enabling theory and the present measure.

Before drawing conclusions from these results, we want to emphasize two limitations in this study. First, sample sizes were relatively small. This is less problematic in the analyses of the structural properties of the forceful and enabling scales where approximately ten coworkers provided ratings for each of 78 or so executives. These data should be reasonably reliable. But the

sample sizes in the 50s for the polarity index and effectiveness correlations do impose limits on inferences drawn from these statistics. Because of small n s, confidence intervals around the point estimates are large, even though they do not include zero. Despite the consistent theory-confirming patterns, we urge caution in making broad generalizations from this data.

Secondly, the sample was not randomly selected—participants were selected to go through a developmental process. We are not fully aware of the criteria that organizations used to nominate executives for this activity. For example, although we found overly enabling executives to be rare in our sample, this may not be as much the case in the population of executive leaders—although our experience suggests that it is. The same goes for the minority of versatile executives found in this sample.

Measurement of Forceful and Enabling

The two refined five-item scales demonstrated solid measurement characteristics. However, it is likely that these scales suffer from what is technically referred to as construct under-representation (Messick, 1995). That is, forceful and enabling represent a large range of leadership; surely there are other important components of this polarity that are not represented in the current item pool. To better cover the forceful and enabling content domains, future work will need to explicate the facets of each and then generate item pairs to operationalize them. It would also be helpful to determine the convergence between measures of forceful and enabling leadership and the polarity index and conceptually similar constructs, like those used by Quinn and his colleagues (e.g., see Hooijberg & Quinn, 1992; Quinn, 1988). To better understand the characterological bases of leadership style, it will be necessary to see how versatility on forceful and enabling relates to traits such as agency, communion, openness, resilience, cognitive complexity, constructive thinking, and ego development.

We did not explore the measurement equivalence of superior, peer, subordinate, and self-ratings of forceful and enabling leadership in this study. It is reassuring that for all three coworker rating groups we found evidence for the polarity effect and the link between lopsidedness and effectiveness. However, comparing the frequencies of too forceful, too enabling, and disengaged leaders observed across rating sources suggests that superiors may hold different views than peers and subordinates do. Future work will need to evaluate the degree of congruence between different rating sources and identify the more important causes of discrepancies. This could shed important light on social dynamics in perceptions of leadership. This kind of research may also uncover important effects that may help executives in making sense of their developmental feedback.

Theory Evaluation and Development

We found strong empirical support for four substantive hypotheses derived from the forceful and enabling polarity theory (Kaplan, 1996). Using an instrument that measures both underdoing and overdoing, we found a strong negative relationship between forceful and enabling. This is promising because previous research under the linear assumption that

“more is better” would suggest that these constructs are positively related (see summary in Bass, 1990). Perhaps this polarity has been overlooked because it was obscured by the methods used to measure and statistically model them.

Also confirmed was the prediction that forceful leaders outnumber enabling leaders in executive ranks. This is also consistent with the idea that organizations in modern America highly value and promote leaders with an individualistic, self-defining orientation (Drath, 1990), despite all of the talk about empowerment and participative management. It seems that the majority of U.S. executives could benefit from development on the relational side of leadership. An interesting question concerns the cross-cultural generalizability of this finding. In more collectivist societies, it may be the case that enabling leaders are the norm and forceful leaders the exception.

The data were clear that many executives could be considered lopsided on the forceful and enabling polarity. This raises an important question: Is balance obtainable by all managers? One of our clients put it this way: “The flaw in 360-degree feedback is the idea of the renaissance man—that you can do everything.” Perhaps it is asking too much of these highly accomplished managers to “do everything.” But the present cross-sectional data suggests that movement toward increased versatility may lead to improved performance. In other words, coming to value and making strides toward better use of an undeveloped side will likely enhance effectiveness. Perfect balance and complete versatility could be viewed as the ultimate goal, but it is the journey toward that goal, not its realization, that is most important to development. We leave it to longitudinal research to determine just how much movement toward better versatility is necessary for achieving a noticeable improvement in performance.

Most noteworthy was the strong inverse relationship between lopsidedness (as measured by the polarity index) and coworker ratings of effectiveness. This offers support for the forceful and enabling theory as well as the position that “managers who have the behavioral complexity to balance competing demands will do better than managers who focus on one demand over another” (Hooijberg & Quinn, 1992, p.166). It might be argued that our response scale is a de facto measure of effectiveness. Even to the extent that our scale cues effectiveness, we find it compelling that the polarity index accounted for 16 to 35% of the variance in coworkers ratings of overall effectiveness. This is surely a sizable effect for versatility based on only two dimensions considering all of the possible personal, performance, and contextual factors that drive perceptions of effectiveness.

The finding of what we called a “disengaged” leadership style poses a new set of questions to be addressed by the forceful and enabling theory. Given that our measure is linked to effectiveness, the finding of many disengaged executives may reflect the results from a recent survey that indicated nearly half of *Fortune* 1,000 companies regard their leadership capacity as fair to poor (Csoka, 1997).

The “disengaged” phenomenon appears to be very similar to Bass and Avolio's idea of laissez-faire leadership (reported in Bass, 1990, ch. 25). Laissez-faire leadership is characterized by

inactivity, by negligence in both structuring task-oriented and relations-oriented aspects of a unit or group's work. According to Bass and Avolio, leaders who do not engage either a directive or participatory style are much less effective than leaders who engage one, the other, or both. In exploring our data set, we found the same pattern—the average effectiveness rating for disengaged leaders was about 1 standard deviation lower than that for too forceful and too enabling leaders. We intend to further understand disengaged executives by building on Bass and Avolio's work.

We believe that a polar conception of forceful and enabling leadership could represent a step forward in leadership theory. This conception has promise for integrating under one heading many of the ideas and data provided by research on consideration and initiating structure, McGregor's (1960) popular Theory X and Theory Y of managers' belief systems, managerial decision-making styles, and the link between personality and leadership. Moreover, the notion of polarity is also useful for understanding how core strengths and core weaknesses are related (see Kaplan, 1998).

We also believe that the implicitly curvilinear response scale ranging from underdo to overdo is an innovation in measurement technology. Our experience with practicing managers, however, suggests that the underlying concept is not a new one in the executive suite. Senior managers, like the rest of us, respond intuitively to the idea of “going overboard” and being “undergunned” as two kinds of performance problems. Now there is a measure of it. This type of rating scale also has a straightforward appeal in developmental feedback. When coworkers agree that a manager does something too much or too little, he or she often gets the message clearly.

Implications for Practice

The theory and measure presented here are based on several years of action-research with executives. Naturally, there are suggestions for the practice of executive leadership and its development. We limit our present discussion to two that seem most useful. First, an important finding in this study was that coworker ratings showed a fairly strong link between versatility and effectiveness while the data on self-ratings showed no correspondence. This suggests that leaders—perhaps because of their prejudices, biases, attitudes, defenses, and fears—are blind to the source of many of their own performance barriers. This highlights the importance of coworker feedback in directing developmental efforts. To that end, the measure of forceful and enabling leadership, combined with additional feedback data, holds promise for helping executive clients to recognize certain assumptions and blind spots that get in the way of their leadership.

The evidence of pervasive polarizing tendencies supports the view that performance problems are often mirrored between forceful and enabling dimensions. It may be the case that understanding this dynamic can motivate developmental work on both forceful and enabling fronts—cutting back on one while simultaneously stepping up on the other (or stepping up both in the disengaged case)—to improve versatility. The developmental pathway to increased versatility is a different

road for forceful, enabling, and disengaged executives. Understanding these different developmental needs (see literature reviewed in Guisinger & Blatt, 1994; see also Kaplan, 1996) is critical to helping executives improve their leadership. Of course, individual development is an intense and long-term process. The cornerstone to development is the recognition of one's limits—a first step toward growth that the theory and measure of the forceful and enabling polarity may facilitate.

Final Comment

We titled this paper "getting at leadership versatility" and feel that this work does that. However, we do not believe that the forceful and enabling polarity can account for all that is meant by the concept of versatility. The roles senior executives are called to perform are far too varied and complex to be reduced to two dimensions. However, we do believe that the notion of polarity can provide a fruitful way of understanding versatility. With that thought in mind, we have begun working with the idea of a strategic and operational polarity. Unpublished preliminary research is promising, but far from conclusive, in suggesting that this polarity complements the forceful and enabling polarity in understanding versatility and executive effectiveness.

References

- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Chicago: Rand McNally.
- Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd Ed.). New York: Free Press.
- Csoka, L. S. (1997). *Bridging the leadership gap*. New York: The Conference Board.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- 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.
- Dennison, D.R., Hooijberg, R., & Quinn, R.E. (1995). Paradox and performance: Toward a theory of behavioral complexity in managerial leaderships. *Organizational Science*, 6(5), 524-540.
- Drath, W. H. (1990). Managerial strengths and weaknesses as functions of the development of personal meaning. *Journal of Applied Behavioral Science*, 26, 483-499.
- Fleenor, J. W., Fleenor, J. B., & Grossnickle, W. F. (1996). Interrater reliability and agreement of performance ratings: A methodological comparison. *Journal of Business and Psychology*, 10, 367-380.
- Guisinger, S. & Blatt, S.J. (1994). Individuality and relatedness: Evolution of a fundamental dialectic. *American Psychologist*, 49, 104-111.
- Hogan, R. (1983). A socio-analytic theory of personality. In M.M. Page (Ed.), *Nebraska Symposium on Motivation, 1982*, 55-89. Lincoln: University of Nebraska Press.
- 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.
- House, R. J. & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management*, 23, 409-473.
- James, L. J., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98.
- James, L. J., Demaree, R. G., & Wolf, G. (1993). r_{wg} : An Assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78, 306-309.
- Jaques, E. (1986). The development of intellectual capability: A discussion of Stratified Systems Theory. *Journal of Applied Behavioral Science*, 22(4), 361-383.
- Jung, C.G. (1976). *The portable Jung*. New York: Penguin Books.
- Kaplan, R. E. (1996). *Forceful leadership and enabling leadership: You can do both*. Greensboro, NC: Center for Creative Leadership.
- Kaplan, R. E. (1998). Getting at character: The simplicity on the other side of complexity. In R. Jeanneret and R. Silzer (Eds.), *Individual Assessment: The Art and Science of Personal Psychological Evaluation in an Organizational Setting*. San Francisco: Jossey Bass.
- Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald Press.
- Levinson, D.J. (1978). *The seasons of a man's life*. New York: Ballantine Books.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory [Monograph No. 9]. *Psychological Reports*, 3, 635-694.
- Lord, R.G. & Maher, K.J. (1993). *Leadership and information processing: Linking perceptions and performance*. Boston, MA: Rutledge.
- McCall, M. & Lombardo, M., & Morisson, A.M. (1988). *The lessons of experience: How successful executives develop on the job*. Lexington, MA: Lexington Books.
- McGregor, D. (1960). *The human side of enterprise*. New York: McGraw-Hill.

- Messick, S. (1995) Validation of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Quinn, R.E. (1988). *Beyond rational management*. San Francisco: Jossey-Bass.
- Quinn, R.E., Spreitzer, G.M., & Hart, S. (1991). Challenging the assumptions of bipolarity: Interpenetration and managerial effectiveness. In S. Srivastva & R. Fry (Eds.) *Executive and organizational continuity* (pp. 222-252). San Francisco: Jossey-Bass.
- Shrout, P. & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Sivasubramaniam, N., Kroeck, K.G., & Lowe, K.B. (1997). "In the eye of the beholder": A new approach to studying folk theories of leadership. *Journal of Leadership Studies*, 4, 27-42.
- Sullivan, H.S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.
- Zaccaro, S.J. (May, 1999). *Leader flexibility and the assessment of leader performance and executive potential*. Symposium presented at the 14th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Vaillant, G.E. (1993). *The wisdom of the ego*. Cambridge, MS: Harvard University Press.
- Wiggins, J.S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In W. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 2, pp.89-113). Minneapolis: University of Minnesota Press.