Leadership Perception
Analysis of 360-Degree Feedback

Thomas P. Bradley, Jeff M. Allen, Scott Hamilton, and Scott K. Filgo

Introduction

The 360-degree feedback process allows the leader to compare his or her self-ratings to the ratings of subordinates, peers, and the boss. Including multiple rater groups in an analysis of leadership performance captures unique rating variance. Multiple sources are necessary because a lack of agreement often occurs when assessing overall leadership effectiveness (Tsui & Ohlott, 1988).

Typically, raters are grouped by organization level, including subordinates, boss, peers, and the individual leader being assessed. Individuals in workgroups are likely to influence each other through shared observations, perceptions, and social contagion (Lord & Maher, 1991; Meindl, 1995). Within rater groups there would be some consistency in rating the leader but between-group ratings have not had the same consistency (Penny, 2001). Only modest correlations between self and other raters have been found (Harris & Schaubroeck, 1988; Tsui & Ohlott, 1988). Each group of raters evaluates the leader from their own perspective, and each group typically has different scores for a given competency (Borman, 1974; Toegel & Conger, 2003).

Interpretation of the results has been a continuing challenge because of the differences in scores between rater groups. These differences have been the focus of much research to identify the sources of disagreement. Rater characteristics are one possible source of variation. For instance, Antonioni and Park (2001) examined interpersonal affect in a study that investigated the relationship between the raters interpersonal affect toward leader and the leniency of ratings. The results suggest that interpersonal affect does influence rater scores but that the influence is not the same for individual raters or groups. Interpersonal affect exhibits greater influence on subordinates and peers than on supervisors. The researchers wrote: "We believe that each source of 360-degree feedback may be influenced differently by the same factors: they may even be influenced by different factors" (Antonioni & Park, 2001, p 490).

Pulakos, Schmitt, and Chan (1996) investigated leader race and gender as well as organizational level effects. The primary purpose of their research was twofold. First, it investigated the differences in rating due to

Multirater feedback, often called 360-degree feedback, is a popular development and assessment tool, especially for organizational leaders. Raters from different organizational levels, including subordinates, boss, peers, and self, rate the leader's performance. However, there seldom is strong agreement across rater groups. This study used the data from a commercially available 360-degree leader development feedback instrument and a second-order confirmatory factor analysis model to try to explain the differences in ratings between the groups. Rather than an explanation of the differences, what was found were two second-order factors that may be the underlying elements that all raters consider when observing leader performance.

rater race and gender subgroups. Second, it investigated the differences for peer versus boss ratings. They found no statistically significant differences in scores based upon leader race or gender. However, the research results did suggest that the raters from different organizational levels weigh these factors differently. They conclude that "raters from different organizational levels may consider different factors in formulating their ratings of leader" (Pulakos et al., 1996, p. 117).

Leader style, leader gender, and subordinate personality were the variables under consideration in a study of leader evaluations and subordinate motivation by (Cellar, Sidle, Gourdy, & O'Brien, 2001). This study looked for interactions between rater agreeableness, leader style (autocratic or democratic), and leader gender. Rater agreeableness interacted with leader style to affect leader ratings. Agreeable participants rated democratic leaders as more effective than autocratic leaders. Leader gender had no affect on ratings of leader effectiveness.

Brutus and Fleenor (1998) investigated 360-degree feedback across several industries to determine whether the variance between industries was greater than within industry ratings. The potential differences investigated include culture, goals, technology, organizational structures, and management systems. The results indicated small but significant differences between organizational types. Leniency bias, for example, was stronger for educational institutions. Peer ratings also were sensitive to leniency effect in public sector organizations. Interrater agreement was lowest in government agencies and highest in education and manufacturing. Within-group agreement, however, did not differ among various types of industries. They concluded that organization type did affect agreement between rater groups, but not within a rater group.

Tsui and Ohlott (1988) considered three explanations for low rater agreement: informational differences, rating error tendency, and variations in performance criteria. They argued that the most plausible source of dissensus lies in the criteria used to judge the leader. The differences lie in both the type of criteria used and importance the observer attributes to the criteria. They also contended that there may be common criteria associated with leader effectiveness. Their results suggested that all three are plausible causes of rater dissensus and confirm the lack of consensus in overall ratings between the rater groups. They did not find differences in the criteria associated with overall judgment of effectiveness. Rather, the results implied that the weight assigned to a given criterion by a rater could be a cause for dissensus (Tsui & Ohlott, 1988).

Perhaps the differences between these groups are a natural part of the process since raters from different levels of the organization have different perceptions of the leader or observe significantly different facets of a leader's job. If so, then the ratings should reflect these differences (Borman, 1974).

Understanding the perspective and bias of the rater can improve our interpretation of 360-degree feedback. This perspective can provide more insight into the cause of performance gaps within a workgroup and lead to more meaningful interventions. Tsui and Ohlott (1998) posited that

"By isolating the sources of dissensus, researchers can identify when and to what extent interrater agreement is possible. By clarifying the bases for disagreement, managers can determine which raters would be appropriate for what evaluation purposes" (p. 780).

Identifying and building upon the understanding of rater perspective will result in a more complete understanding of 360-degree feedback. Observers at different levels witness different behavior and therefore are likely to have a different bias from other observers (Borman, 1974). Perhaps the disagreement between rater levels is a result of different roles, orientations, and perceptions. Observers from different levels may consider different factors in formulating their ratings (Pulakos et al., 1996).

Differences Between Raters

There are several other reasons why the ratings between groups would differ. First, leadership itself is based upon contingency factors such as leader behavior, leader-follower relationship, and leadership prototypes. (Chemers, 1997; Lord, Brown, & Freiberg, 1999). Leadership contingency theories (Fiedler, 1978; House, 1971) have implied that differences in ratings are the result of effective leadership. Effective leaders will adjust behaviors depending upon situational factors relating to the followers or the task. This adjustment may result in different behaviors performed before different raters.

Understanding the perspective and bias of the rater can improve our interpretation of 360-degree feedback. This perspective can provide more insight into the cause of performance gaps within a workgroup and lead to more meaningful interventions.

Second, the quality of the relationship between the observer and the leader also impacts rater perception, performance, and subsequent ratings (Howell & Hall-Merenda, 1999; Lord, Brown, Harvey, & Hall, 2001). For example, members of the subordinate group perceive the leader from a familiar perspective since they often interact on a more frequent basis with the leader than other groups.

Third, leadership is, in part, based upon the observer's perceptual processes. Perception involves the creation of a schema or prototype of leadership in the mind of the observer. A leader who matches a perceived schema is considered effective (Lord & Maher, 1991). The perceptions or prototypes may differ between organizational levels, based upon differences in observed leader behavior and social contagion (Borman, 1974; Meindl, 1995; Penny, 2001).

The source of rater difference may exist at a level deeper than the dimensions measured by the 360-degree instrument. Perhaps an underlying construct can be identified as the source of disagreement. If that construct is identified as a source of agreement within a rater group, then this knowledge will increase the usefulness of 360-degree feedback. If no underlying construct is found, it is possible that a shared view of leadership does not exist within rater groups and that individual differences, perception, rater error, or some other antecedent can cause observer differences.

Research Variables

Two constructs were selected because of their history in leadership research and because of their probable contribution to the differences in how leaders are perceived. The constructs of interest in this study are Relationship and Results. The historical importance of the factors of Relationship and Result go back to studies conducted at Ohio State University and the University of Michigan in the 1940s and 1950s. In the Ohio State University studies the Relationship and Results were identified as (a) Consideration, which describes the leader's behavior toward subordinates and (b) Initiating Structure, which describes the leader's task orientation (Stogdill & Coons, 1957). The University of Michigan studies identified: Employee Orientation, which describes a leader who emphasizes the leader's relationship behavior; and Production Orientation, which describes the leader's technical and production behavior (Kahn & Katz, 1960).

> The source of rater difference may exist at a level deeper than the dimensions measured by the 360-degree instrument. Perhaps an underlying construct can be identified as the source of disagreement.

Howell and Hall-Merenda (1999) noted that "leader-focused research implicitly assumes a relationship of some sort between leader and follower, and that implied relationship is fundamental to the link between leader behavior and follower response" (p. 680). Conclusions from that study showed a significant connection between the quality level of leader-follower relationship and higher levels of follower performance.

In this context, Relationship as defined as mutual trust, respect, and influence between a leader and follower (Howell, & Hall-Merenda, 1999). Results are defined as what to accomplish and how to accomplish it. Achieving results means attaining targeted, expected, hoped for, and desired outcomes as measured by the organization and significant stakeholders (Ulrich, Zenger, & Smallwood, 1999).

Hypothesis

The objective of this research is to examine the result of 360-degree leadership feedback from the perspective of rater self interest. The subordinates' interests are predicted to focus on the interpersonal relationship between themselves and the leader. Members of the subordinate group perceive the leader from an interaction and relationship perspective. The relationship influences the follower's perception and schema of leadership. "Subordinates are the direct targets of the leader's behavior and thus can provide feedback to the manager about his or her behaviors from first hand experience" (Atwater, Roush, & Fischthal, 1995, p.36). Presumably subordinate ratings reflect the relative importance of the manager in their work lives, pay raises, development, assignments, and other outcomes. Subordinates may monitor manager behavior more closely in this regard (Maurer, Raju, & Collins, 1998).

Hypothesis 1: The factor loading for subordinate ratings will be statistically different between the two second order factors, Relationship and

Results, and Relationship factor loading will be greater than the Results factor loading.

At the other end of the spectrum is the leader's boss, where the subordinate-supervisor relationship is reversed. Traditionally, the boss is considered to have the best overall perspective of the leader. The interest to the boss relates to job performance, as measured by the results achieved by the leader and subordinate workgroup efforts (Lawler, 1967; Ulrich, et al., 1999). In this case the boss is in the best position to observe the results of the leader's efforts (Ulrich et al., 1999). Workgroup results are the traditional measure of leader performance, and they are of great interest to the organization and the boss. This job performance perspective is in contrast to that of the subordinates' and should produce differences between the scores for the two groups Therefore, it is suggested that the boss will have a bias toward items that measure Results.

Hypothesis 2: The factor loading for boss ratings will be statistically different between the two second order factors, and Results factor loading will be greater than the Relationship factor loading.

Peers have a singular opportunity to observe the behavior of the leader as both a rival and a colleague. Peers observe leader behavior from a perspective that includes the knowledge of, expertise in, and responsibility for organizational results similar to that of the leader (Maurer, et al., 1998). Peers see the leader's behavior when the boss does not and may observe aspects of behaviors of which the boss is unaware (Lawler, 1967; Zazanis, Zaccaro, & Kilcullen, 2001). Peer ratings, therefore, may contain a bias toward organizational results.

Hypothesis 3: The factor loading for peer ratings will be statistically different between the two second order factors, and Results factor loading will be greater than the Relationship factor loading.

Self-ratings may be good predictors of future behaviors, and leaders know more about their own behavior than anyone else (Lawler, 1967). However, self-report data tend to be unreliable, leaders are likely to select responses that put themselves in the best light with the boss and other raters (Podsakoff & Organ, 1986). They may underrate performance so as not to appear to be boasting or they may overrate their performance to present an image better of themselves (Atwater et al., 1995). Self-ratings are expected to show a bias toward Result measures.

Hypothesis 4: The factor loading for self ratings will be statistically different between the two second order factors, and Results factor loading will be greater than the Relationship factor loading.

Method

Sample

The sample comes from an existing commercial data set consisting of over 6,021 cases of feedback ratings from January 2002 to October 2003. Each rating case includes feedback results from up to 10 raters. The 10 raters include one boss and one self-rating; the remaining 8 ratings are a

combination of peers and subordinates. Peer and subordinate scores are aggregate scores calculated by simple averages.

Leaders were asked to identify their organization type. The wide range of industries represented is shown in Table 1. The percentage column indicates the percentage of total responses for that organization type. The leaders rated in this study held positions in all levels of management. Executives

TABLE 1
ORGANIZATION TYPE

Type of organization	%
Education	4.8
Military of Government	3.3
Other Public Sector	2.5
Social Services	0.7
Other Non-Public Sector	4.8
Financial Services	9.7
Manufacturing	14.8
Sales	14.9
Health Care	16.1
Utilities, Transportation and Communication	11.9
Consulting	16.4

comprised 29% of the leaders rated, middle management comprised 33%; and supervisors comprised 25%. The remaining 13% was made up of other staff positions.

There is a wide range of industries represented and a cross-section of management levels, therefore, inferences from this study should be generalizable to the larger population of leaders and managers. However, the leaders rated are all members of organizations that chose to use a leader development instrument. Differences may exist between these organizations and organizations that choose not to use 360-degree instruments. Therefore, generalizations made to the larger population of leaders and managers may include some risk of bias. Additionally, cross-cultural differences were not considered in this study since almost all of the ratings occurred in the United States and those few ratings that were done within other cultures were mostly rating American leaders.

Instrument

The instrument selected was the CheckPoint 360° Competency Feedback System® because of its history as a developmental tool and its wide acceptance among practitioners. The instrument is a 70-item self-administered survey. The rater scores the item by indicating the frequency of

observed leader behavior on a 5-point Likert scale. A score of 1 indicates that the behavior is never observed. A score of 5 indicates that the behavior is always observed. The 70 items define eight core competencies of:

1. Communication—including the skills of listening to others, processing information and communicating effectively.
2. Leadership—covering the abilities of instilling trust, providing direction, and delegating responsibility.
3. Adaptability—encompassing the skills of adjusting to circumstances and thinking creatively.
4. Relationships—assessing the capabilities of building relationships and building team success.
5. Task Management—gauging the level of aptitude for working efficiently and competently.
6. Production—appraising the ability to initiate action and achieve results.
7. Development of Others—measuring proficiencies in cultivating individual talents and in motivating successfully.
8. Personal Development—including the behaviors of displaying commitment and seeking improvement (Profiles International, 2001, p. 4, used with permission).

The intended purpose of the instrument is to provide feedback on the strengths and developmental needs of the leader. The instrument is administered online, and confidentiality is assured for peer and subordinate raters. The competencies are validated periodically by comparing competency ratings to actual performance reviews.

Three studies of reliability have been reported for the CheckPoint 360° Competency Feedback System®. The internal reliability test was conducted between 1994 and 1996. The resulting reliability scores ranged from .12 to .85. A test-retest reliability study for a sample of 49 managers over a time interval of 8 weeks had scores from .21 to .83. The third study was a factor analytic study that resulted in a seven-factor model, with alpha scores ranging from .75 to .90 (Profiles International, 2001).

Data Analysis

The data were processed using a second-order confirmatory factor analysis (CFA) model as shown in Figure 1. The model shows the eight leader competencies, identified by the rectangles, for each rater group. The observed variables, or in this case competencies, are coded DR1 through DR8 for the subordinate ratings. Peers boss and self ratings are coded with P, B, and S respectively. The eight latent factors were identified by loading the appropriate competency factors onto latent factors that define Relationship and Result by rater group. For example: the first order factor Peer Relationship (PeerRel) is defined by the competencies of: P1. Communication; P2. Leadership; P4. Relationship; and P7. Develop Others. The factor for Peer Results (PeerRes) is defined by the competencies of: P3. Adaptability; P5.
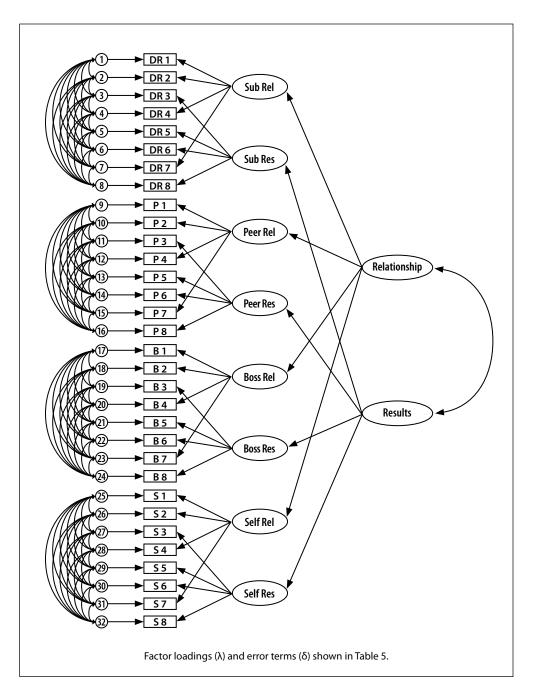
Factor loadings (λ) and error terms (δ) shown in Table 5.

FIGURE 1.
Second-order factor model with correlated error terms.

Task Management; P6. Production; and P8. Personal Development. The latent factors, for each rater group, in turn define the second-order factors Relationship and Result.

Confirmatory factor analysis is iterative in that after the data are analyzed, post hoc modifications may be made to improve model to data fit.

Two conditions must be met before considering changes to the model. First, modifications must be defensible based upon proposed theory. Second, modifications must be verified with another sample (Bandalos, 1996).

A panel of practicing managers enrolled in a graduate management class assisted in determining which competencies would load on which second-order factors for the initial trials of the model. The final factor loading configuration was refined based upon analysis of model to data fit while complying with the two conditions above. The analysis was accomplished using the structural equation modeling program Lisrel 8.5 (Jöreskog & Sörbom, 2000a).

Factor loadings (λ) identify the extent to which a given observed variable is able to define the latent variable. In confirmatory factor analysis, latent variables are not directly measured; they are inferred from the relationships with the observed variables. Factor loadings between the observed variables and the latent variables indicate their correlation (Schumacker & Lomax, 1996).

Measurement error (δ) is that portion of the observed variable that measures something other than the latent variable. Error terms are similar to residuals on multiple regression in that they are the part of the observed variable not explained by the latent factor (Bandalos, 1996; Schumacker & Lomax, 1996).

Results

The total sample from the database was 6,021 cases. Missing values occurred for self-ratings in fewer than 1% of the cases; boss ratings were missing in 4% of the cases; peer ratings were missing in 5% of the cases; and subordinate ratings were missing in 15% of the cases. Rather than eliminate the cases that had missing data, values were imputed using the estimation and maximization (EM) method in Lisrel 8.5 (Jöreskog & Sörbom, 2000a). EM imputation is a two-step iterative process. The first E-step finds the current estimated value. The second M-step maximizes the expected log likelihood as if it were based upon complete data. This maximization results in the next estimate for the missing parameter. Using the new value, the E asnd M steps are repeated until it converges to a maximum likelihood for the missing data (Rubin & Thayer, 1982).

Descriptive Statistics

Structural equation models are sensitive to data extremes; non-normality, outliers, and missing data can all prevent proper analysis of structural equation models. Prelis 2 a component of Lisrel 8 analysis software, checks for data anomalies and allows for numerous data transformations (Jöreskog & Sörbom, 1993). Prelis also provides common descriptive statistics including the means and standard deviations by rater group as shown in Table 2.

Reliability

Internal reliability was calculated using the entire data set of ratings. The Cronbach alpha values for each of the variables are shown in Table 3.

TABLE 2
MEAN AND STANDARD DEVIATION FOR IMPUTED DATA

Competencies	Self		Boss		Subordinates		Peers	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Communication	4.00	0.41	3.91	0.52	4.05	0.48	3.93	0.45
Leadership	4.22	0.39	4.07	0.50	4.14	0.46	4.06	0.43
Adaptation	3.95	0.50	3.82	0.60	4.01	0.50	3.85	0.48
Relationship	4.11	0.44	3.99	0.55	4.08	0.50	3.97	0.48
Task	4.02	0.49	4.06	0.57	4.18	0.47	4.07	0.45
Production	4.12	0.50	4.02	0.60	4.14	0.49	4.00	0.47
Develop Others	4.08	0.51	4.01	0.59	4.03	0.56	3.98	0.50
Personal Development	4.13	0.46	4.12	0.56	4.16	0.48	4.04	0.45

TABLE 3
CRONBACH ALPHA FOR OBSERVED VARIABLES BY RATER

Competencies	Raters				
	Self	Boss	Peer	Subordinate	Overall
Communication	0.84	0.89	0.91	0.93	0.92
Leadership	0.85	0.88	0.91	0.92	0.91
Adaptability	0.83	0.87	0.90	0.90	0.90
Relationships	0.82	0.86	0.89	0.91	0.89
Task Management	0.79	0.83	0.88	0.88	0.88
Production	0.86	0.90	0.92	0.92	0.92
Development of Others	0.82	0.86	0.88	0.89	0.88
Personal Development	0.79	0.84	0.84	0.85	0.84

Alpha values were calculated for each competency by rater, and an overall alpha value was calculated for each competency for all raters.

Model Development

Structural equation model development was an iterative process in which modifications were made to improve model to data fit assuming the modification can be supported. The model presented in Figure 1 represents the best fit of leader competencies to rater perceptions of the Relationship and Results factors tested n this study.

The model was evaluated with multiple random samples of 400 cases. Then the complete data set including all 6,021 cases was tested against the model. The large difference in chi-square is due the sensitivity of chi-square

to the larger sample size for the complete data set. A summary of the fit statistics is also shown in Table 4 along with the critical values to aid in interpretation. Multiple fit statistics have been included for analysis since no single test of significance has been shown to be conclusive for structural equation modeling (Schumacker & Lomax, 1996).

The factor loadings (λ), shown in Table 5, were all significant at an alpha level of 0.05, which indicates that the observed variables actually measured

TABLE 4
COMPARISON OF MODEL FIT STATISTICS

	Critical values	Sample data[1]	Complete data[2]
Chi Square	Nonsignificant	1356	13697
RMSEA	<.06	0.089	0.088
RMR	<.06	0.053	0.031
NFI	>.95	0.92	0.94
CFI	>.95	0.92	0.94
GFI	>.95	0.81	0.85

Notes: [1] n=400; [2] n=6021.

TABLE 5
FACTOR LOADING AND ERROR TERMS FOR FACTOR DIAGRAMS

Competencies	Subordinate		Peer		Boss		Self	
	λ	δ	λ	δ	λ	δ	λ	δ
Relationship								
Communication	0.951	0.10	0.944	0.11	0.780	0.39	0.834	0.31
Leadership	0.954	0.09	0.906	0.18	0.823	0.32	0.857	0.27
Relationships	0.953	0.09	0.955	0.09	0.942	0.11	0.860	0.26
Develop Others	0.915	0.16	0.886	0.22	0.880	0.23	0.754	0.43
Variance Explained	0.890		0.852		0.737		0.685	
Results								
Adaptability	0.908	0.18	0.879	0.23	0.841	0.29	0.809	0.35
Task Management	0.909	0.17	0.893	0.20	0.808	0.23	0.786	0.38
Production	0.938	0.12	0.905	0.18	0.882	0.22	0.872	0.24
Personal Development	0.905	0.18	0.861	0.26	0.792	0.37	0.790	0.38
Variance Explained	0.837		0.783		0.691		0.664	

Note: Variance explained (h^2) is the percentage of variance that the latent variables explain for each of the second-order factors. The value of h^2 is the sum of the square of the factor loadings divided by the number of factor loadings.

what they were intending to measure. The error or correlated uniqueness terms (δ) indicate relatively high correlation between errors for the observed variables. Methods effects are the probable cause of the correlated error terms. The effects of the high correlation were minimized by allowing the error terms to correlate for each rater group.

The second-order path diagram, Figure 2, shows the latent variables for each rater based upon the competencies related to the second-order factors, Relationship and Results. For example, the peer ratings for the competencies, Leadership, Communication, Relationship, and Develop Others, were loaded on a factor called PeerRel, for Peer Relationship. The high values indicate the correlation between the latent variable and the second order factor.
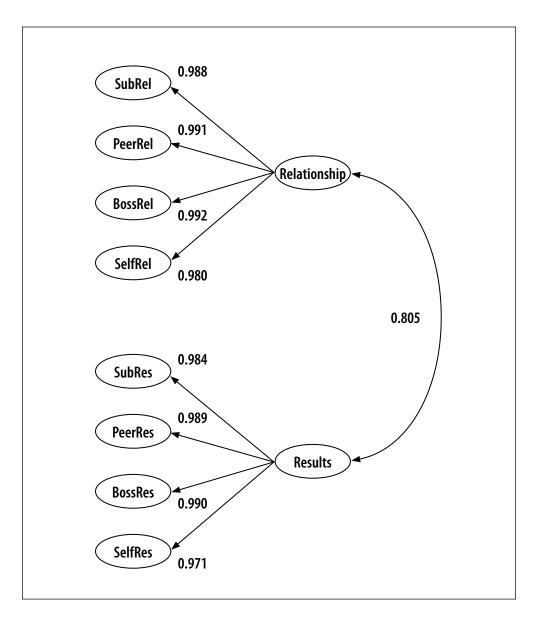
Discussion

The model demonstrates a reasonable fit to the data when model complexity is taken into consideration. Values above the critical value of 0.95 are considered acceptable for normed fit index (NFI), comparative fit index (CFI), and goodness of fit (GFI), while root mean square error of approximation (RMSEA) and root mean square residual (RMR) should be less than the critical value of 0.06 (Bentler & Bonett, 1980). Traditionally, critical values have been considered acceptable if they were larger than 0.90 for GFI and CFI.

Model fit was evaluated for both a small sample of 400 cases and the complete data set of 6,021 cases. The reason for this was to examine the effects of sample size on chi-square. In Table 4 the chi-square value for the model with a sample of 400 was 1,356. With the sample of 6,021, chi-square became 13,697. Both values were significant, so chi-square is not a good indicator of fit for the selected sample sizes.

An unexpected finding was the higher factor loading values for all raters on the second-order factor Relationship. Figure 2 shows the second-order factor loadings. In each case the factor loadings are statistically significant, but the difference between Results and Relationship is not statistically significant. Three of the four hypotheses for this study posited that factor loadings for Results would be greater than Relationship. It was expected that only the subordinates would have higher factor loadings on Relationship due to their direct day-to-day contact with the leader. Boss, self, and peers sould indicate Results as higher.

Three important observations can be made form the analysis of Figure 2. First, the only statistically significant difference between the second-order factors was with the boss group of raters. The actual difference was very small and raises the question of practical significance. It was expected that the differences would be much greater. The fact that the values were close indicates that, although there is a statistically measurable difference between the two constructs, both are important to the boss in providing feedback to a leader.

Second, the factor scores for all raters are very close for both second-order factors. It was not the purpose of this analysis to find agreement

SubRel 0.988
PeerRel 0.991
BossRel 0.992
SelfRel 0.980
Relationship
0.805
SubRes 0.984
PeerRes 0.989
BossRes 0.990
SelfRes 0.971
Results

between rater groups but the data suggest that there is more agreement between raters than expected given 360-degree feedback literature (e.g. Harris & Schaubroeck, 1988; Maurer et al., 1998; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Tsui & Ohlott, 1988).

Third, the two second-order factors Relationship and Results correlate highly in this model, with a correlation value of .805. Typically, when a high correlation exists between second-order factors the question should be asked do raters perceive two factors? Or, are the two factors actually indicating only one factor? If the assumption is made that raters are considering one factor, Leadership, when performing their ratings then this explanation is valid.

FIGURE 2.
Second-order factor path diagram with factor loading (λ).

However, there are two reasons why the theoretical foundation for this analysis precludes combining the Relationship and Results factors. First, the purpose of the research was to determine whether there is a difference between these factors in the perceptions of the rater groups. Second, it can be argued that Relationship and Result are distinct constructs that happen to have a strong correlation. Effective leaders must achieve results through the efforts of others. It becomes necessary to develop a relationship with those others in order to exert influence upon their behaviors and the quality of that relationship is predictive of workgroup outcomes (Gerstner & Day, 1997). Similarly, positive results achieved by a workgroup will have a positive impact upon that group's perception of the leader and therefore result in a stronger perceived relationship (Lord et al., 1999).

Conclusions

An attempt was made in this study to contribute to the knowledge of between-group rating differences by suggesting that the differences are the result of observer perception based upon organization level. Differences were also believed to be influenced by leader behaviors toward members at that level. The data demonstrate the contrary and imply that all raters are influenced by the leader's human relations ability as well as performance results.

Model fit statistics imply that generalizable conclusions can be made with caution. The fit is acceptable but not robust; there are several reasons for this. First, the high correlations in the error factors within rater groups suggest method effects influence the analysis. Method effects are a theoretical possibility within rater groups for this study. Social desirability, halo effect, and response sets may well result in covariance among the error terms for the individual competencies and thus reduce the significance of the findings.

Second, model complexity has a negative impact upon goodness of fit statistics. Goodness of fit compares the data from the observed mode to an estimated model to evaluate statistical significance. The larger the number of coefficients estimated the greater the likelihood of reducing the fit of the observed model to the estimated model

Third, the instruments used in 360-degree feedback may elicit different responses from different rater groups. Consistency of data is important to model fit, if raters interpret the competency items differently that would reduce model fit.

Fourth, there may be other second-order factors that the competency variables are attempting to measure. Structural equation modeling does not preclude the possibility of other models. Other models and variables may fit the data as well or better than the model developed here. Given these issues, the model fit is sufficient to support previous research and to encourage further research.

These findings reinforce the research that between-group raters do not strongly agree on the observed dimensions used in the 360-degree feedback (e.g. Harris & Schaubroeck, 1988; Maurer et al., 1998; Mount et al., 1998; Tsui & Ohlott, 1988).

However, rather than finding the differences between groups on the second-order factors, fundamental constructs were found that form the foundation for agreement. The high percentage of variance explained reinforces the conclusion that Results and Relationship are important factors in leadership feedback. It also supports previous research assumptions that leader-follower relationships are linked to follower responses (Howell & Hall-Merenda, 1999).

Raters' perception of Relationship was found to be highly correlated to the perception of Results. This research suggests that Relationship and Results underlie the competency dimensions of leader feedback and that the two constructs are mutually dependant. Human relationships are necessary to achieve organizational results, and good results strengthen leader-follower relationships.

It appears, from this study, that raters value a leader with a balance of human relations and performance results focus at all levels. Both factors ranked high and were highly correlated. In practice this suggests that 360 degree evaluations should be considered from a macro level considering relationships and results as well as the traditional competency level. If any rater group identifies a weakness in the relationship or results constructs then improvements in that area is likely to have a positive impact upon the other. Leader development is also impacted by these findings; an organization seeking to improve leadership results can not overlook the impact of human relations.

Author's Note

References

Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management, 27*, 479-495.

Atwater, L., Roush, P., & Fischthal, A. (1995). The influence on upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.

Bandalos, B. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), *Applied multivariate statistics for the social sciences* (3rd ed., pp. 389-420). Mahwah, NJ: Erlbaum.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Borman, W. C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Performance, 15*, 105-124.

Brutus, S., & Fleenor, J. W. (1998). Does 360-degree feedback work in different organizations? A between industry comparison. *Journal of Management Development, 17*(2/3), 177-191.

Cellar, D. F., Sidle, S., Gourdy, K., & O'Brien, D. (2001). Effects of leader style, leader sex, and subordinate personality on leader evaluations and future subordinate motivation. *Journal of Business and Psychology, 16*(1), 61-72.

Chemers, M. M. (1997). *An integrative theory of leadership*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fiedler, F. E. (1978). The contingency model and the dynamics of the leadership process. In L. Berowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press.

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology, 82*(6), 827-844.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

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

Howell, J. M., & Hall-Merenda, K. E. (1999). The ties that bind: The impact of leader-member exchange, transformational and transactional leadership, and distance on predicting follower performance. *Journal of Applied Psychology, 84*(5), 680-694.

Jöreskog, K. G., & Sörbom, D. (2000a). Lisrel 8.50 [Computer software]. Lincolnwood, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom D. (2000b). Prelis 2.50 [Computer software]. Lincolnwood, IL: Scientific Software International.

Kahn, R., & Katz, D. (1960). Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (2nd ed.). Evanston, IL: Row and Peterson.

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51*(5), 369-381.

Lord, R. G., Brown, D. J., & Freiberg, S. J. (1999). Understanding the dynamics of leadership: The role of follower self-concepts in the leader/follower relationship. *Organizational Behavior and Human Decision Processes, 78*(3), 167-203.

Lord, R. G., Brown, D. J., Harvey, J. L., & Hall, R. J. (2001). Contextual constraints on prototype generation and their multilevel consequences for leadership perceptions. *Leadership Quarterly, 12*, 311-338.

Lord, R. G., & Maher, K. J., (1991). *Leadership and information processing: Linking perceptions and performance.* London: Routledge.

Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance measurement equivalence. *Journal of Applied Psychology, 83*(5), 693-702.

Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *Leadership Quarterly, 6*(3), 329-341.

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

Penny, J. (2001). Differential item functioning in an international 360-degree assessment: Evidence of gender stereotype, environmental complexity, and organizational contingency. *European Journal of Work and Organizational Psychology, 10*(3), 245-271.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research. *Journal of Management, 12*, 531-544.

Profiles International. (2001). *Profiles CheckPoint 360º Competency Feedback System™: History, development, and research information.* (Available from Profiles International, 5205 Lakeshore Drive, Waco, TX, 76710)

Pulakos, E. D., Schmitt, N., & Chan, D. (1996). Models of job performance ratings: An examination of leader race, leader gender, and rater level effects. *Human Performance, 9*(2), 103-119.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika, 47*(1), 69-76.

Schumacker, R. E., & Lomax, R. G. (1996). A beginner's guide to structural equation modeling. Mahwah, NJ: Lawrence Erlbaum Associates.

Stogdill, R. M., & Coons, A. (1957). *Leader behavior: Its description and measurement* (Research Monograph No. 88). Columbus: Ohio State University, Bureau of Business Research.

Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning & Education, 2*(3), 297-311.

Tsui, A. S., & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Inter-rater agreement and consensus in effectiveness models. *Personnel Psychology*, 41, 779-803.

Ulrich, D., Zenger, J., & Smallwood, N. (1999). *Results-based leadership.* Boston: Harvard Business School Press.

Zazanis, M. M., Zaccaro, S. J., & Kilcullen, R. N. (2001). Identifying motivation and interpersonal performance using peer evaluations. *Military Psychology, 13*(2), 73-88.

THOMAS P. BRADLEY

Thomas P. Bradley, Ph.D. is an assistant professor at Tarleton State University in the Department of Management Marketing and Administrative Systems. His research interests include leadership, organizational behavior, and human performance technology. *Email*: tbradley@tarleton.edu

JEFF ALLEN

Jeff Allen, Ph. D. is the Program Coordinator at the University of North Texas for Applied Technology, Training and Development. His research interests include human performance technology, strategic planning, and future learning. *Email*: jallen@unt.edu

SCOTT HAMILTON

Scott Hamilton, Ph.D. is the Senior Vice President at Profiles International in Waco, Texas. *Email*: scott.hamilton@profilesinternational.com

SCOTT K. FILGO

Scott K. Filgo, M.S. is a senior research associate at Profiles International in Waco, Texas. *Email*: scott.filgo@profilesinternational.com