Self-Assessment and Rater-Assessment as a Dimension of Management Development

Kenneth M. Nowack

In light of mixed findings in the literature regarding the congruence between self-assessment and those provided by others, this study investigated differences between the self-ratings of managers and others (supervisors, subordinates, peers) using an instrument assessing twenty specific management skills areas derived from job analysis procedures. Data on 335 managers working in several large organizations (health care, aerospace, utility) who recently received feedback were used to compare self-other ratings along each management practice skill area. The results indicated that the amount of agreement between managers and the others describing them was moderately low, with correlation coefficients ranging from .122 to .295 (p < .05). Paired t-tests revealed that managers consistently reported practicing specific skills more frequently than others who rated them across the majority of skill areas. Implications for practice and future research are discussed.

Self- and other ratings are typically important elements in a wide variety of human resource systems including career management, performance appraisal, succession planning, personnel selection, and management training and development. For example, it is common in many performance appraisal systems to encourage employees to rate themselves and to compare these ratings to those of their supervisors during the appraisal review discussion. Similarly, assessment centers typically use a variety of both self-assessment tools (personality, style, and career interest inventories) and trained assessor ratings to provide an overall evaluation of employees for either selection or development. Of importance in each of these human resource systems is the accuracy and congruence of both the self- and other ratings utilized. In particular, a large amount of the literature on multiple ratings has focused on the utility and validity of self-assessment.

Despite alleged gains from self-assessment processes in human resource

systems such as performance appraisal and training (for example, clarifying differences of opinion, improving communication between supervisors and subordinates), empirical research shows mixed findings regarding the congruence between self-ratings and those provided by other sources (see Harris and Schaubroeck, 1988; Mabe and West, 1982). Some studies have shown moderately high correlations in the range of .50 to .60 between self-ratings and those of others (for example, Pym and Auld, 1965; Williams and Seiler, 1973), whereas others have shown little or no significant correlations (for example, Cooke, 1989; Ferris, Yates, Gilmore, and Rowland, 1985; Klimoski and London, 1974; Shrauger and Terbovic, 1976; Huber, 1991).

In their review of over fifty studies, Shrauger and Schoeneman (1979) examined evidence relating self-perceptions to evaluations by significant others. In general, they found little evidence of congruence between self-perceptions and evaluations by others, nor did they find consistent evidence that self-evaluations are strongly influenced by others' feedback. In other words, if feedback from others is either critical or constructive, it may simply be ignored altogether. Additional evidence suggests that different raters exhibit only low to moderate agreement with each other, which may not be unique to self-ratings (for example, Lawler, 1967; Tucker, Cline, and Schmitt, 1967).

In a recent meta-analysis of self and supervisor, self and peer, and peer and supervisor ratings, Harris and Schaubroeck (1988) concluded that there is, at best, a modest correlation between different rating sources. The results of their meta-analysis indicated a relatively high correlation between peer-supervisor ratings ($r = .65$), but only a modest correlation between self-peer ($r = .36$) and self-supervisor ratings ($r = .35$). Furthermore, although rating format (global versus dimensional) and rating scale (behavioral versus trait) did not moderate the results, job type (managerial versus blue collar) did appear to moderate self-peer and self-supervisor ratings. Specifically, self-supervisor and self-peer correlations were much lower for managerial employees than for blue-collar types. However, these effects were not seen for peer-supervisor correlations.

In practice, this apparent lack of congruence may not be as unexpected or as enigmatic as it might appear. It is possible that multiple raters may actually be observing different aspects of individual performance or possess diverse standards resulting in different perceptions of skills, ability, and performance evaluations of others. Harris and Schaubroeck (1988) suggest that a number of explanations might exist to account for this lack of congruence between self-ratings and those of others in the literature, including: (1) differences in organizational level leading to differential weightings of performance factors and standards; (2) inequality of observational opportunities to rate and evaluate specific supervisory and management practices; and (3) self-rating bias, social

desirability, defensiveness, and distortion that lead to unrealistic appraisals and evaluations of job performance.

In fact, a growing body of empirical evidence suggests that most individuals possess an overly positive, albeit unrealistic, view of themselves (see Greenwald, 1980; Kagan, 1988; Taylor and Brown, 1988). Of particular concern is the tendency for people to present themselves in socially desirable or self-deceptive ways (for example, Schwartz, 1990; Crowne and Marlowe, 1964; Edwards, 1957). For example, when asked to indicate how accurately positive and negative personality adjectives describe the self, individuals tend to judge positive characteristics of self as being more descriptive than negative attributes (Alicke, 1985; Brown, 1986; Lewinsohn, Mischel, Chaplin, and Barton, 1980). Additionally, limited evidence exists to suggest that even when negative feedback is presented and acknowledged, it is typically rationalized as inconsequential or diminished in terms of its accuracy (Campbell, 1986; Lewicki, 1984). Distrust regarding the predictive utility of self-reports has undoubtedly been enhanced by critiques of trait approaches to personality measurement, which have consistently shown only modest associations between self-report measures and specific behaviors (for example, Mischel, 1972, 1979).

Although previous research raises legitimate causes for questioning the validity of self-reports as they compare with other commonly used assessment methods, the significance of these studies may have been largely overestimated, especially in light of the many advantages individuals may have as judges of themselves. Shrauger and Osberg (1981) reviewed the validity of individuals' self-ratings with other assessment procedures widely used in psychological and job performance evaluations. The authors conclude that self-assessments are at least as predictive of specific criteria (including job performance and peer ratings) as are other assessment methods against which they have been compared. The authors suggest that the relative accuracy of self- versus other assessment procedures seem to vary somewhat with the type of construct being predicted. For some criteria, such as peer judgments of personality traits, self-reports have been consistently superior to assessment by others. However, the results for psychotherapy outcomes and academic achievement tend to show that self-reports are only slightly better predictors. Finally, studies of job performance and post-hospitalization adjustment have been largely inconclusive, with neither self-report nor external judgments clearly being more accurate.

Despite evidence suggesting that individuals tend to hold unrealistically positive views of the self, some recent studies by Howard and others (1980), Cole, Howard, and Maxwell (1981), Gabbard, Howard, and Dunfee (1986), and Howard, Conway, and Maxwell (1985) dispute the notion that self-report ratings may be less valid than other assessment methods. In these studies exploring the consensual validity between self-report ratings and other measurement methods (such as observational ratings by others),

· when controlling for the influence of method variance, far greater evidence of attribute-method agreement was consistently observed. The critical point to be learned from these studies is that previous failures to find strong evidence of the construct validity of self-reports might well be due to errors in the criterion variables as opposed to the measures being validated.

In general, the majority of organizational studies to date investigating the congruence between self- and other ratings have tended to focus on measures used in performance evaluation and appraisal systems. It can be argued that the relationship between self- and other ratings have not been as adequately explored within the context of other human resource systems such as assessment centers, management training, and succession planning. Although an impressive literature already exists, it would appear that additional research exploring the association between self- and other ratings with alternate human resources systems would provide even greater clarity of the utility and validity of such assessment approaches.

In light of the mixed literature findings on the efficacy and convergence of self-ratings with those of others, this study attempted to replicate and extend previous research investigating the congruence between self-perceptions of specific supervisory and management practices and behaviors to those of others in diverse organizational settings. Specifically, this study compared self-report perceptions of specific management practices in twenty skill areas to a pooled rating of up to five others (such as subordinates, superiors, peers) who had the opportunity to observe and evaluate the behaviors. This study extended earlier research by using a behaviorally focused management practices questionnaire derived from job analysis procedures of supervisory and management positions and by using pooled "other" ratings to minimize potential error variance attributed to differences in organizational level and observational opportunities. Recent evidence suggests that the validity of others' assessments is enhanced by pooling or averaging across a number of different raters (Horowitz, Inouye, and Siegelman, 1979). Consistent with prior research findings suggesting that individuals possess exaggerated self-evaluations, it was hypothesized that only modest correlations in the range of .10 to .40 would be found between self- and other ratings of specific supervisory behaviors and management practices.

Method

Participants and Procedures. Data on 335 supervisors and managers working in several diverse organizations, including aerospace, utility, and health care, were used in this study. These individuals were selected from a larger population of managers who had recently completed a management practices questionnaire as part of a company sponsored management training program. The cases selected for analysis were those that had been processed most recently and had no missing data on the self-description

instrument. Each manager was asked to select up to five others to complete a parallel form of the instrument soliciting feedback on the same supervisory and management practice skill areas.

Approximately two weeks after all self- and other ratings were collected, each manager received a confidential feedback report summarizing twenty distinct managerial skill areas and comparing self-perceptions of performance to those of the others. The identity of the others providing feedback to the manager was not revealed in an attempt to reduce desirability bias and increase the objectivity of the other ratings. A letter was included with the feedback reports thanking the respondents for their participation and containing a brief explanation of how to interpret the report.

Descriptions of a subset of these managers (335) were provided by a total of 1,205 "others," who included peers, subordinates, and superiors. The mean age of the sample was 40.4 (SD = 9.31) with ages ranging from 25 to 67 years. The sample was composed of 73.1 percent males and 26.9 percent females, including entry-level to midlevel supervisors and managers. The mean age of the others ($N = 1205$) was 42.87 (SD = 10.06) and was composed of 34.6 percent males and 65.4 percent females.

Measurement of Management Practices. In this study, the Management Practices Questionnaire (MPQ) was used to assess supervisory and management skills (Nowack, 1989). The MPQ was designed to provide participants in supervision and management with specific feedback on twenty specific job skills and behaviors. The instrument is completed by the manager and up to five other individuals who are in a position to accurately evaluate his or her work performance. The combined perceptions of the five others (including peers, subordinates, superiors) are compared to the manager's own responses, in a comprehensive computerized feedback report, to identify strengths and weaknesses and provide a base for individualized professional development efforts.

The MPQ is composed of 100 items measuring twenty specific management practices areas organized into four general areas: (1) *communication skills*—listening, two-way information, written communication, oral communication, oral presentation; (2) *administrative skills*—goal setting, planning/organizing, delegation, follow-up, performance appraisal; (3) *interpersonal skills*—team building, recognizing/rewarding, empathy/sensitivity, conflict resolution, discipline, coaching, leadership, involvement orientation; and (4) *decision-making skills*—problem-analysis, decision making.

The MPQ was based on an extensive job analysis of supervisory and managerial positions in several large service, manufacturing, and aerospace companies (Nowack, 1989). Incumbents were interviewed regarding the major job tasks associated with their positions. A comprehensive job-task activity questionnaire was constructed that obtained ratings of both importance and frequency of each job task relative to the targeted supervisory and management positions. The job-task questionnaires were distributed to

incumbents and statistically analyzed. A set of critical supervisory and managerial dimensions were derived in four general areas: (1) interpersonal, (2) communication, (3) administration, and (4) decision making. Items were rationally constructed to measure the full range of supervisory and management skills in each of these four areas. A total of twenty separate scales were derived, each measured by five separate questions each. Each question was behaviorally focused, and respondents rated the frequency or practice of each behavioral skill using a seven-point scale, where 1 equals "to an extremely small extent" and 7 equals "to an extremely large extent."

In an attempt to maximize inter-rater reliability, respondents were given explicit instructions on completing the MPQ. These instructions stated: "The Management Practices Questionnaire contains 100 items. You will provide your response to these items using the 1–7 scale next to each item. Circle the number that reflects how *frequently* you have practiced (or observed the other person practicing) the specific skill or behavior described in the item. Please circle the 1 or 7 *only* when the behavior is *almost never* or *almost always* practiced or observed. As a frame of reference, think how you or the person you are describing compares with other supervisors or managers that you work with within your organization. If you have not been able to observe the behavior or it is not applicable to the position, please circle Not Applicable. Be sure to provide a response for every item in the questionnaire. There is no time limit. Take your time with each item and be certain that you are rating yourself or others on what is specifically stated. There are no right or wrong answers. You will be most helpful if you answer as accurately and as honestly as possible. You may notice that some items appear similar to each other. Actually, no two items are identical. They differ, even if only to a small degree. This similarity in the item pool provides for a comprehensive and reliable assessment of supervisory and management skills and behaviors."

Two parallel forms of the instrument were provided, one for self-description and the second for description by others. The two parallel instruments and items are identical. Sample items in the MPQ include: (1) Waits out silences and listens patiently without interrupting others; (2) Gives employees feedback about how they are performing their job on a regular basis; (3) Conveys an appreciation for what employees are experiencing or how they are feeling; (4) Seeks employees' input about decision making, planning, and problem-solving processes; and (5) Clearly communicates performance expectations and standards to employees. Administration of the MPQ takes approximately thirty minutes.

Analysis. Scale scores were computed by taking the average of the responses to the five items associated with each of the MPQ twenty scales. Means and standard deviations were obtained for each scale, separately for self-descriptions and the descriptions of others. Internal consistency reliability was estimated for each scale by means of Cronbach's alpha. Compar-

ison of inter-rater agreement was assessed using a series of paired *t*-tests comparing self and other ratings in an attempt to minimize error variance in the other's ratings (stratifying the others by focal manager). In these analyses, descriptions by others were aggregated to the focal manager level. These aggregated (averaged) descriptions were also correlated with the self-descriptions to provide a measure of consensual validity.

Item-total scale correlations were also run at the self-description level to ascertain whether the MPQ items correlated more strongly with their own scales than with any of the other scales. Finally, principal components analysis with varimax rotation was done on the MPQ using the sample data to explore the exact factor structure of the instrument.

Results

The descriptive statistics for self- and other ratings on the twenty MPQ scales are summarized in Table 1 below. Responses by the managers as well as the other respondents are skewed toward the favorable side, indicating frequent practice of specific management skills. Mean scores for self-descriptions range from 23.44 (performance appraisal) to 27.54 (decision making); means for descriptions by others range from 22.57 (performance appraisal) to 25.38 (written communication). For all scales, the means of the manager's self-descriptions are higher (that is, more frequent practice of specific behaviors) than the means of the descriptions of others. However, the standard deviations for the others' descriptions are consistently higher than those of the self-descriptions.

The internal consistency reliability (alpha) of the MPQ scales is fairly high, ranging from .71 to .91 as measured by Cronbach's alpha (Table 1). The MPQ has shown acceptable test-retest reliability ranging from .45 to .82 over a three-month period in previous research (Nowack, 1989). The average test-retest reliability across all twenty scales was .65.

Item-total analyses indicate that items within each scale are highly intercorrelated with each other; however, items across certain scales also correlate modestly with each other (data not shown). A principal components factor analysis across the entire sample, with varimax rotation, yielded a three-factor solution explaining a total of 74.2 percent of the common variance. These three factors included (1) task management skills, (2) people management skills, and (3) communication skills that correspond to the general managerial skill areas identified in the original job analysis. However, the first factor (people management skills) was clearly the largest factor, accounting for 61.4 percent of the common variance.

Separate factor analyses were run to compare the factor structure of the MPQ for both self- and other ratings in this study. Principal component analysis, with varimax rotation, yielded a similar three-factor solution, with a single interpersonal factor accounting for the majority of common vari-

Table 1. MPQ Scales: Descriptive Statistics and Internal Consistency Reliability of the MPQ

MPQ Scale	Mean (Self/Others)	SD (Self/Others)	Alpha
Communication Skills			
Listening	24.87/23.99	3.89/5.32	.84
Two-way information	26.30/24.00	3.41/5.12	.76
Written communication	27.05/25.38	4.78/5.55	.90
Oral communication	26.12/25.25	3.75/5.16	.82
Oral presentation	25.36/24.59	3.84/5.19	.73
Administrative Skills			
Goal setting	24.92/23.67	3.54/4.92	.83
Planning/organizing	24.48/23.89	3.99/4.93	.71
Delegation	26.47/24.49	3.20/4.98	.81
Follow-up	24.26/23.14	3.77/4.93	.87
Performance appraisal	23.45/22.59	4.59/4.89	.83
Interpersonal Skills			
Team building	26.69/23.98	3.53/5.28	.86
Recognizing/rewarding	25.39/23.38	3.85/5.16	.91
Empathy/sensitivity	26.08/23.73	4.19/5.45	.87
Conflict resolution	25.98/23.33	3.58/5.08	.81
Discipline	25.78/23.71	3.68/4.96	.87
Coaching	26.19/23.69	3.39/5.07	.78
Leadership	25.59/23.25	3.37/4.95	.80
Involvement orientation	26.68/23.89	3.52/5.01	.81
Problem Solving			
Problem analysis	25.83/24.16	3.79/4.99	.78
Decision making	27.54/25.03	3.15/5.12	.80

Note: N = 335.

ance in both the self- and other ratings (56.3 percent and 63.3 percent, respectively). In this study, interpersonal dimensions such as rewarding others, team building, listening, giving feedback, leading, resolving conflict, and so on are rather stable and salient interpersonal dimensions being assessed by respondents and others in the MPQ.

Inter-Rater Agreement and Convergent Validity. A series of paired *t*-tests were used to assess differences in agreement between the managers and the others, describing them across all twenty MPQ scales as seen in Table 2. Paired *t*-tests were used to minimize for potential nuisance factors related to individual rater characteristics and possible unequal observational opportunities on the part of the others providing feedback. With paired *t*-tests, the observed differences between the two groups are more likely to be directly attributable to differences in management practices in specific skill areas rather than observer characteristics. Across all the MPQ scales,

Table 2. MPQ Scales: Inter-Rater Agreement and Convergent Validity

MPQ Scale	Mean Differences	t-value	p	r
Communication Skills				
Listening	0.87	2.74	.006	.237[c]
Two-way information	2.23	7.60	.000	.206[c]
Written communication	1.67	4.90	.000	.274[c]
Oral communication	0.87	2.81	.005	.218[c]
Oral presentation	0.77	2.43	.015	.214[c]
Administrative Skills				
Goal setting	1.25	4.05	.000	.144[b]
Planning/organizing	0.59	1.92	.058	.218[c]
Delegation	1.99	6.49	.000	.122[a]
Follow-up	1.22	3.92	.000	.164[b]
Performance appraisal	0.88	2.77	.006	.252[c]
Interpersonal Skills				
Team building	2.72	8.87	.000	.240[c]
Recognizing/rewarding	2.00	6.72	.000	.300[c]
Empathy/sensitivity	2.35	7.37	.000	.288[c]
Conflict resolution	2.65	8.55	.000	.175[b]
Discipline	2.07	6.99	.000	.238[c]
Coaching	2.51	8.18	.000	.166[a]
Leadership	2.34	7.95	.000	.207[c]
Involvement orientation	2.79	9.08	.000	.166[a]
Problem Solving				
Problem analysis	1.67	5.45	.000	.204[c]
Decision making	2.51	8.11	.000	.130[a]

[a] $p < .05$.
[b] $p < .01$.
[c] $p < .001$.

except planning and organizing, managers reported significantly more frequent practice of specific skills than the others (aggregated by focal manager) describing them (Table 2).

The amount of agreement between the managers and the others describing them is moderately low (Table 2, last column). Correlation coefficients range from .12 ($p < .05$) to .30 ($p < .001$), with an average correlation across all twenty MPQ scales of .21 ($p < .001$). Strongest agreement between self and others was in the rewarding and recognizing scale-interpersonal skills, which are relatively behavioral in nature and generally observable. The lowest correlations were for the delegation and decision making scales-administrative skills that are less behavioral and can easily be misinterpreted (for example, viewing an employee-involved decision-making style of a manager as reflecting a relative lack of decisiveness;

viewing the distribution of assignments to others as displays of authoritative, rather than participative, behavior on the part of a manager).

Discussion

This study examined the congruence between self-ratings of managers and those of others (aggregated ratings of peers, subordinates, and superiors) on twenty specific management practices skill areas. In general, managers consistently had inflated perceptions of how frequently they practice specific behaviors and skills compared to others who provided feedback to them. Evidence for concurrent validity was weak, with correlations ranging from .12 to .30 between self- and other ratings.

In this study, when incumbents and others (supervisors, subordinates, peers) who provided ratings were used as the unit of analysis, incumbents reported significantly higher levels of management practices than the pooled ratings of others (Table 2). In comparison, when the average aggregated ratings of all incumbents ($N = 335$) were directly compared to the aggregate ratings of all others ($N = 1205$), no significant differences were observed across any of the twenty management practices skill areas in an additional series of paired t-tests (all p's $> .05$). The difference in results for aggregate ratings versus self-other ratings supports the recent findings of Huber (1991), who suggests that important information can be lost when data is simply averaged across all group members.

The results of this study support previous research by Harris and Schaubroeck (1988), Ferris, Yates, Gilmore, and Rowland (1985), and Klimoski and London (1984) who found only modest agreement between self- and other ratings. This study also provides limited support for the existence of a tendency for individuals to maintain positive self-evaluations compared to others (Taylor and Brown, 1988). In their review of illusion and well-being, Taylor and Brown (1988) suggest that rather than being attentive to both the favorable and unfavorable aspects of self, individuals tend to be more cognizant of their strengths and assets and considerably less aware of their weaknesses and faults. In fact, some recent evidence exists that individuals who are low in self-esteem or moderately depressed may actually have a more realistic and balanced outlook on life (for example, Watson and Clark, 1984; Coyne and Gotlieb, 1983). It might appear that this capacity to maintain inflated views of the self might actually interfere with professional growth and development. That is, if managers are skilled at maintaining overly optimistic perceptions of self, even in the face of critical or negative feedback from others, what might account for the motivation to change and continue to develop professionally? As such, it would appear important to identify and understand employees who might possess poor self-insight leading to possibly unrealistic and erroneous appraisals of skills and abilities.

Recent findings by Campbell (1991) provide some limited evidence that samples of managers in a business school advanced management program ($N = 154$), plateaued marketing managers ($N = 16$), and "up and coming" marketing managers within the same organization ($N = 24$) appeared to have significantly higher and less accurate ratings of their leadership skills compared to those of others (peers, subordinates, supervisors). In his analysis of over 9,000 employees with thirty independent standardization samples for the Campbell Leadership Index (CLI), Campbell (1991) found a wide variety of significant perceptual gaps between self- and other ratings among the thirty independent samples studied. These findings suggest that differences in perception of skills, ability, and performance may be characteristic of some type of employees but not of others. Future research is obviously needed to replicate the Campbell (1991) findings regarding plateaued managers and to explore what individual and organizational factors may contribute to the accuracy of both self- and other ratings used in a variety of human resource systems.

It might be argued that managerial behavior change is a function of at least three distinct factors: (1) awareness of individual strengths and areas for improvement, (2) motivation and self-efficacy related to changing specific behaviors, and (3) ability and capability to alter specific managerial skills and behaviors. Therefore, the capacity to interpret critical feedback from others in a positive light may, in fact, prove to be a necessary condition for initiating and sustaining successful behavioral change efforts. In this way, the ability to alter perceptions of others' feedback in an empowering manner might actually enhance professional growth, learning, and development. Conversely, overly optimistic and unrealistic self-appraisals may have important negative repercussions for managers functioning in political environments. For example, managers might rely on skills that are actually ineffective or pursue career options for which they are poorly suited. Managers might also fail to perceive and accurately interpret specific negative feedback from their superiors, subordinates, or customers, leading to the maintenance of behaviors that are largely dysfunctional or resulting in organizational "derailment" (see McCall and Lombardo, 1983). Additional research evidence on these points is obviously needed.

The present findings are somewhat contrary to the conclusions of Shrauger and Osberg (1981) and Howard (1990), who argue for the utility and validity of self-reports. As Harris and Schaubroeck (1988) point out, the general lack of congruence between different raters (self-others) is not that surprising nor necessarily problematic. It is possible that the use of the particular management assessment instrument employed in this study may have led to individual definitions of effective performance and hence to different perceptions of the manager's practice of specific skills and behaviors. However, in their meta-analysis of rater agreement, Harris and Schaubroeck (1988) did not find that rating format had any impact as a

moderator for either managerial or nonmanagerial positions. Nevertheless, subsequent research using behavioral observation and skill rating scales might provide a more definitive test of the hypothesis of whether alternative rating scales and definitions of performance affect rater agreement. Future research is also recommended to more fully explore the possible role of other potential moderating factors in clarifying the dynamics of rater agreement, such as personality characteristics of the rater, rater training, use of the rating, and observational opportunity (see McIntyre, Smith, and Hassett, 1984).

Given the mixed findings regarding the accuracy of self-ratings and extensive research suggesting that a general lack of convergence between raters is fairly common, it seems useful to consider the conditions under which self-assessment might be more practical, useful, and valid than alternative methods. Shrauger and Osberg (1981) provide a series of questions that can assist practitioners regarding the use of self-assessment ratings as well as specific recommendations for facilitating the accuracy of self-ratings. These include (1) specifying the decisions/actions to be made, (2) examining relevant empirical data on the accuracy of self- versus other assessments, (3) specifying what variables are relevant to the decisions/actions to be made, (4) determining the best source of information about the relevant predictor variables, (5) phrasing questions to maximize accuracy, (6) facilitating recall of relevant previous experiences, and (7) maximizing self-assessor's motivation for honest and accurate reports. Future research is obviously required to better understand the conditions under which self-ratings might be better than other assessment methods and to explore procedures for maximizing the construct and predictive validity of self-assessments.

In terms of practical implications, the results of this study support previous research suggesting that self-ratings will generally show only modest associations and concurrent validity with ratings by others. In previous research, this appears particularly true of managerial, supervisory, and professional positions in which tasks and performance outcomes appear to be more ambiguous than those of hourly, nonexempt, and blue-collar jobs (for example, Thorton, 1980; Mabe and West, 1982; Heneman, 1974, 1986). Practitioners utilizing self-ratings (for example, performance appraisals, management training and development) should expect some degree of disagreement and incongruence between individuals and others providing observational feedback ratings. Although the exact meaning and cause of the common disparity among raters may be unclear, practitioners should take caution not to consider one source of data as being necessarily more valid than another. In fact, some researchers have suggested that managers may have a very different, and even unique view, of their own job performance compared to any other observational data source (see Thorton, 1980; Levine, 1980; Bem, 1967).

HRD practitioners should continue to gather, compare, and analyze data from multiple sources in various human resource systems. When possible, self- and other ratings should be compared to more objective performance criterion and productivity measures. When incongruence exists between self- and other ratings on specific human resource systems (for example, performance appraisal, assessment centers), practitioners should attempt to explore the nature and meaning of these perceptual gaps and encourage discussion among all individuals providing ratings. In this way, perceptual differences can be acknowledged, clarified, and understood, paving the way for constructive behavior change and performance improvement.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*, 1621–1630.

Bem, D. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183–200.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*, 353–376.

Campbell, D. P. (1991, August). *Assessing leadership characteristics using a multi-rater self versus observer survey.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Campbell, J. D. (1986). Similarity and uniqueness: The effects of attribute type, relevance, and individual differences in self-esteem and depression. *Journal of Personality and Social Psychology, 50*, 281–294.

Cole, D., Howard, G., & Maxwell, S. (1981). Effects of mono- versus multiple operationalization in construct validation efforts. *Journal of Consulting and Clinical Psychology, 49*, 395–405.

Cooke, R. (1989). Assessing managerial skills with the management effectiveness profile system. *Educational and Psychological Measurement, 49*, 723–732.

Coyne, J. C., & Gotlieb, I. H. (1983). The role of cognition in depression: A critical appraisal. *Psychological Bulletin, 94*, 472–505.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Ferris, G., Yates, V., Gilmore, D., & Rowland, K. (1985). The influence of subordinate age on performance ratings and causal attributions. *Personnel Psychology, 38*, 545–557.

Gabbard, C., Howard, G., & Dunfee, E. (1986). Reliability, sensitivity to measuring change, and construct validity of a measure of counselor adaptability. *Journal of Consulting and Clinical Psychology, 33*, 377–386.

Greenwald, A. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*, 603–618.

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

Heneman, H. G. (1974). Comparisons of self- and superior ratings of managerial performance. *Journal of Applied Psychology, 59*, 638–642.

Heneman, H. G. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology, 39*, 811–826.

Horowitz, L. M., Inouye, D., & Siegelman, E. (1979). On averaging judge's ratings to increase

their correlation with an external criterion. *Journal of Consulting and Clinical Psychology, 47,* 453-458.

Howard, G. (1990). On the construct validity of self-reports: What do the data say? *American Psychologist, 45,* 292-294.

Howard, G., Conway, C., & Maxwell, S. (1985). Construct validity measures of college teaching effectiveness. *Journal of Educational Psychology, 77,* 187-196.

Howard, G., Maxwell, S., Weiner, R., Boynton, K., & Rooney, W. (1980). Is a behavioral measure the best estimate of behavioral parameters? Perhaps not. *Applied Psychological Measurement, 4,* 293-311.

Huber, V. (1991). Comparison of supervisor-incumbent and female-male multidimensional job evaluation ratings. *Journal of Applied Psychology, 76,* 115-121.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist, 43,* 614-620.

Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59,* 445-451.

Landy, F., & Farr, J. (1980). Performance rating. *Psychological Bulletin, 87,* 72-107.

Lawler, E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51,* 369-380.

Levine, E. L. (1980). Organizational applications of self-appraisal and self assessment: Another look. *Personnel Psychology, 33,* 259-262.

Lewicki, P. (1984). Self-schema and social information processing. *Journal of Personality and Social Psychology, 45,* 384-393.

Lewinsohn, P. M., Mischel, W., Chaplin, W., & Barton, R. (1980). Social competence and depression: The role of illusory self-perceptions. *Journal of Abnormal Psychology, 20,* 63-81.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

McCall, M. W., & Lombardo, M. M. (1983). *Off the track: How and why successful executives get derailed* (Technical report no. 21). Greensboro, NC: Center for Creative Leadership.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69,* 147-156.

Marks, G. (1984). Thinking one's abilities are unique and one's opinions are common. *Personality and Social Psychology Bulletin, 10,* 203-208.

Mischel, W. (1972). Direct versus indirect personality assessment: Evidence and implications. *Journal of Consulting and Clinical Psychology, 38,* 319-324.

Mischel, W. (1979). On the interface of cognition and personality: Beyond the person-situation interface. *American Psychologist, 34,* 740-754.

Nowack, K. (1989). Administration and software manual for the management practices questionnaire. Woodland Hills, CA: Organizational Performance Dimensions.

Pym, D., & Auld, H. (1965). The self-rating as a measure of employee satisfactoriness. *Journal of Occupational Psychology, 39,* 103-113.

Schwartz, G. (1990). Psychobiology of repression and health. In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality theory* (pp. 405-434). Chicago: University of Chicago.

Shrauger, J. S., & Osberg, T. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin, 90,* 322-351.

Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin, 86,* 549-572.

Shrauger, J. S., & Terbovic, M. L. (1976). Self-evaluation and assessments of performance by self and others. *Journal of Consulting and Clinical Psychology, 44,* 564-572.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Thorton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263-271.

Tucker, M. F., Cline, V. B., & Schmitt, J. R. (1967). Prediction of creativity and other performance measures from biographical information among pharmaceutical scientists. *Journal of Applied Psychology, 51,* 131-138.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465-490.

Williams, W. E., & Seiler, D. A. (1973). Relationship between measures of effort and performance. *Journal of Applied Psychology, 57,* 49-54.

Kenneth M. Nowack is a consultant with Organizational Performance Dimensions, Woodland Hills, California.