

A Five-Year Study Of Upward Feedback: What Managers Do With Their Results Matters

Alan G. Walker

First Tennessee Bank

James W. Smither

La Salle University

Running head: Upward Feedback

Abstract

We present results for 252 target managers over five annual administrations of an upward feedback program (i.e., twice as long as any previous study in this area). We show that managers initially rated poor or moderate showed significant improvements in upward feedback ratings over the five-year period, and that these improvements were beyond what could be expected due to regression to the mean. We also found that (a) managers who met with direct reports to discuss their upward feedback improved more than other managers, and (b) managers improved more in years when they discussed the previous year's feedback with direct reports than in years when they did not discuss the previous year's feedback with direct reports. This is important because it is the first research evidence demonstrating that what managers do with upward feedback is related to its benefits. We use an accountability framework to discuss our results and suggest directions for future research.

A Five-Year Study of Upward Feedback: What Managers Do with Their Results Matters

In recent years, performance evaluations and feedback are increasingly gathered from non-traditional sources such as subordinates, peers, and internal or external customers (London & Smither, 1995). When performance evaluations of an employee are collected from more than one source, it is referred to as multisource feedback. Upward feedback refers to the process by which supervisors receive ratings from multiple subordinates. In the present paper, we focus on upward feedback.

Multisource and upward feedback appear to be widely used (London & Smither 1995; Romano, 1994). Although usually collected for developmental purposes, there appears to be increasing interest in using such feedback for administrative decisions such as compensation or promotion (Timmreck, 1995). Also, Graddick and Lane (1998) report that such feedback is increasingly a component of executive appraisals. Despite its growing use and importance, no study has systematically examined whether managers' responses to upward feedback (i.e., what managers do with the feedback) are related to subsequent performance improvement.

In this paper, we examine the impact of upward feedback by looking at changes in ratings (i.e., performance) over five annual administrations of an upward feedback program. We also examine whether managers' responses to feedback (in this case, holding a meeting with his or her direct reports to discuss feedback results) are related to performance improvement. Both of these issues have important practical implications. Specifically, practitioners want to know whether upward feedback is likely to promote performance improvement (e.g., as reflected by improved scores on subsequent upward feedback administrations as well as other outcome measures) and whether such improvement will be sustained over an extended period of time. Also, it is

important to identify components or characteristics of upward feedback programs that will maximize the value and impact of such feedback.

We first review previous research about upward feedback. Then we use an accountability framework to describe potential benefits that may occur when managers discuss their feedback results with others. We then present and discuss results from a five-year study that examines these issues.

Previous Research

Despite its increasing popularity and importance in organizations, much of the research on upward and multisource feedback has not examined its impact on subsequent performance. For example, research has examined issues such as between-source (e.g., self-subordinate) agreement (London & Wohlers, 1991), correlates of agreement (Hezlett, Kuncel, & Cochran, 1997), reactions to feedback (Bernardin, Dahmus, & Redmon, 1993; Smither, Wohlers, & London, 1995) and practitioner-oriented concerns such as instrument development and administration issues (e.g., Bernardin & Beatty, 1987; London, Wohlers, & Gallagher, 1990; Tornow, 1993, Van Velsor & Leslie, 1991a, 1991b).

While human resource practitioners expect that upward feedback will help employees change their behavior and improve their job performance, a comprehensive meta-analysis by Kluger and DeNisi (1996) found that feedback interventions in general (i.e., providing feedback concerning individual performance or behavior) do not always improve performance. The average performance difference (expressed in standard deviation units) between groups receiving a feedback intervention and control (no feedback) groups was .41, thereby indicating that, on average, feedback was associated with enhanced performance. However, about one-third of the effects were negative (i.e., feedback detracted from performance). The fact that feedback

interventions in general are often less than successful points to the importance of understanding whether upward feedback is effective and the conditions that enhance or detract from its effectiveness.

Several longitudinal studies have examined the extent to which performance (as measured by subsequent feedback scores) improves following upward feedback. Hegarty (1974) was the first to report that subordinates perceived performance improvement among supervisors who received upward feedback. Another study found skill increases and higher self-coworker agreement two years after receiving multisource feedback (Hazucha, Hezlett, & Schneider, 1993). However, only 48 of the original 198 managers who initially received feedback volunteered to participate in the follow-up two years later, thus raising the question of whether changes among the 48 managers were likely to be representative of changes in the larger sample that initially received feedback. In a study of multisource feedback provided to 48 assistant store managers in a retail clothing chain, Bernardin, Hagan, Ross and Kane (1995) found that subordinate ratings (i.e., upward feedback) increased over time.

Smither, London, Vasilopoulos, Reilly, Millsap, and Salvemini (1995) examined upward feedback given to 238 managers in the international operations division of a large organization at two points in time, about six months apart. Results indicated that managers whose initial feedback scores were moderate or low improved over the six-month period, and the improvement could not be attributed solely to regression to the mean. Reilly, Smither and Vasilopoulos (1996) extended the Smither et al. (1995) study by following 171 of the original 238 managers for a third administration of the upward feedback questionnaire, and 92 of the managers for a fourth administration 2.5 years later. They found that managers whose initial feedback scores were low sustained their performance improvements over the later administrations.

Atwater, Roush & Fischthal (1995) found that follower ratings of student leaders at the U.S. Naval Academy improved after upward feedback was given to leaders and that leaders receiving "negative" feedback (defined as those for whom follower ratings were substantially below self ratings) improved the most.

Johnson and Ferstl (1997) examined self-ratings and subordinate ratings from 1,903 managers in an accounting firm at two points in time (1995 and 12 to 18 months later in 1996). They found that over-raters (who initially rated themselves more favorably than they were rated by subordinates) had lower self-ratings and higher subordinate ratings at time 2 (relative to time 1). In contrast, under-raters (who initially rated themselves less favorably than they were rated by subordinates) had higher self-ratings and lower subordinate ratings at time 2 (relative to time 1). For in-agreement raters (who initially rated themselves about the same as they were rated by subordinates), self- and subordinate ratings did not change appreciably. These findings were observed regardless of the manager's initial feedback scores (and after accounting for any effects associated with regression to the mean). Among managers who did not complete self-ratings, those initially rated low by subordinates improved and those initially rated high by subordinates declined from time 1 to time 2. The cumulative effect of these changes was that gains made in performance by over-raters offset declines among under-raters; overall performance (across all managers) did not change from time 1 to time 2.

Atwater, Waldman, Atwater, and Cartier (1998) collected upward feedback ratings at two time periods (separated by 10 months) concerning two groups of supervisors from a state police agency. One group of supervisors received feedback after time-1 (based on random assignment) but the other group did not. Atwater et al. found a significant improvement in upward feedback scores over time for supervisors who received feedback at time-1 but no significant improvement

over time for the supervisors who did not receive feedback at time-1. They also found that, among supervisors who received feedback at time-1, supervisors' organizational cynicism was negatively related to improvement over time, and supervisors' acceptance of the feedback (e.g., belief that feedback was honest, valuable, and led to goal setting) was positively related to improvement over time.

Taken together, these longitudinal studies indicate that managers generally improve their performance (at least as reflected by subsequent feedback from their subordinates) after receiving upward feedback, and score improvement is greatest among managers who initially receive the most negative feedback or who initially overrate themselves (note that managers who initially receive the lowest feedback are often the same managers who initially overrate themselves). Also, several of these studies (Atwater et al., 1995; Johnson & Ferstl, 1997; Reilly et al. 1996; and Smither et al., 1995) demonstrated that performance improvements were not merely due to the effects of regression to the mean. One purpose of the present study is to examine whether similar improvements in upward feedback scores are observed and sustained over a five year period in a different setting (a regional bank).

Limitations of Previous Research

A major limitation of all previous research concerning upward or multisource feedback is that no one has examined whether managers' responses to such feedback are associated with performance improvement. Locke and Latham (1990) have shown that feedback alone is not the cause of behavior change, instead it is the goals that people set in response to feedback. It is not enough merely to provide people with feedback, goals are also important. In the case of upward feedback programs, this suggests that performance improvements will depend on the extent to

which managers use the feedback report to set performance improvement goals and monitor their progress toward those goals.

Research showing that some managers improve more than others (e.g., those with initially low scores, especially when those scores are below their self-evaluations) is suggestive of the role that goal setting may play (i.e., these managers may be more likely than others to set goals for personal improvement because the feedback suggests a gap between their personal goal and current behavior). These studies, however, have not examined what managers actually do with the feedback.

In this context, London, Smither, and Adsit (1997) have pointed out that managers often receive upward or multisource feedback solely for developmental purposes and are not accountable for doing anything with the feedback. They argued that upward or multisource feedback is likely to have limited impact when managers are not accountable for using the feedback to guide performance improvement and that lack of accountability may be the Achilles' heel of multisource and upward feedback programs. They state that accountability involves accepting and meeting one's personal responsibilities, being or feeling obligated to someone else or oneself, or having to justify one's actions to others about whom we care. They also note that accountability may be driven by internal forces (such as the desire to gain approval, avoid embarrassment or meet an obligation) or external forces (such as financial outcomes). They summarize research indicating that those involved in multisource or upward feedback appear to want low accountability for themselves but high accountability from others. For example, raters prefer that their ratings remain anonymous (London, Wohlers, & Gallagher, 1990), whereas target managers (ratees) prefer to know the identity of individual raters (Antonioni, 1994). At the same

time, ratees prefer the feedback to be confidential (i.e., not shared with the ratee's supervisor) and not used to influence formal appraisals.

Although there has been increasing research attention devoted to the accountability of raters (e.g., Antonioni, 1994; Hauenstein, 1998; Kozlowski, Chao, & Morrison, 1998), almost no attention has been directed to the accountability of ratees to use feedback to guide performance improvement efforts.

In the most common applications of upward or multisource feedback, ratee accountability is likely to be low (London et al., 1997; London & Smither, 1995). Most often, feedback is used only for developmental purposes; it is given only to the ratee (it is not shared with the supervisor) and ratees are not required to discuss or share the feedback with others. Many (about 40 percent) programs collect and provide feedback only once (so ratees do not expect that they will be rated again by their coworkers) and many do not require ratees to participate in specific developmental or training interventions in response to the feedback (London & Smither, 1995; Timmreck, 1995).

Dalessio (1998) and London et al. (1997) describe several approaches that may increase the ratee's accountability to use upward feedback to guide performance improvement. For example, Dalessio suggests that the actual feedback report could be provided to the ratee in a confidential manner (i.e., the report would not be shared with the supervisor), however, the ratee would be required to use the report to construct a development plan that would be shared with the supervisor. The supervisor could subsequently evaluate the extent to which agreed upon developmental goals were achieved.

London et al. (1997) suggest that ratee accountability could be increased by encouraging (and training) ratees to discuss their feedback results with raters (e.g., direct reports), asking raters to clarify the feedback, and having ratees commit publicly to changing their behavior.

In the present study, we examine the extent to which holding such feedback meetings with direct reports is associated with performance improvements following upward feedback. There are several ways that holding such meetings could affect subsequent performance. For example, discussing feedback with direct reports could affect (a) depth of processing (requiring the ratee to think about the feedback), (b) goal setting (encouraging the ratee to set goals for behavior/performance improvement), and (c) goal commitment (public commitment to improvement goals and subordinates' expectations that the ratee will change make the ratee unwilling to dismiss the goals). Feedback meetings also provide an opportunity for raters to clarify their written feedback and offer specific, constructive suggestions for behavior change and performance improvement. The ratee (manager) may expect that the cooperation (e.g., effort, receptiveness to instructions) of raters (direct reports) may be enhanced if the ratee is visibly responsive to the issues discussed and suggestions offered during such feedback meetings.

Holding a meeting to discuss upward feedback results may also create a more supportive context for the manager's change efforts. Direct reports who are familiar with their manager's developmental goals may feel more empowered to provide ongoing feedback to the manager, they could be more forgiving of mistakes the manager makes as he or she tries to improve, and they may respond better to the manager's new behaviors because they understand the intent that underlies these behaviors.

The feedback seeking literature also suggests reasons why holding meetings with raters (e.g., direct reports) to discuss upward feedback results may be helpful to managers. Holding such meetings may help managers better understand the reasons for any unfavorable ratings (negative feedback) they received. Ashford and Tsui (1991) have shown that seeking negative feedback can enhance a manager's self-awareness (i.e., an accurate understanding of how the

manager is evaluated by others). They point out that this should help managers better regulate and adjust their behavior. They also found that asking for negative feedback enhanced others' opinions of the manager's overall effectiveness, whereas seeking only positive feedback decreased others' opinions.

The Present Study

We examined ratings from five annual administrations (from 1991 through 1995) of an upward feedback program at a large regional bank. Each year (except the first year), we asked raters (direct reports) whether their supervisors (ratees) had conducted a feedback session with the work group to discuss the previous year's upward feedback. The present study thereby contributes to the literature on upward feedback in two important ways. First, the five year period examined in the present study is twice as long as that examined in any previous study of upward (or multisource) feedback. Second, this is the first study to examine whether managers' responses (in this case, holding feedback meetings) to upward feedback are related to subsequent performance improvements.

Based on the review of research and theory presented above, we developed four hypotheses. The first two hypotheses are derived from (and consistent with) previous research findings.

Hypothesis 1: Managers' performance (as rated by their direct reports) will improve over the five year period.

Hypothesis 2: Managers who initially receive the poorest ratings will improve more than managers who initially receive more favorable ratings.

The third and fourth hypotheses are related to the effect that discussing upward feedback results with direct reports may have on the manager's subsequent improvement.

Hypothesis 3: Managers who meet with direct reports to discuss the previous year's feedback results will improve more than managers who do not meet with direct reports to discuss the previous year's feedback results.

If this between-subjects hypothesis is confirmed, however, we cannot rule out the possibility that a third unmeasured variable (e.g., conscientiousness, developmental orientation) may have affected both the likelihood that a manager will hold feedback meetings and the likelihood that the manager will improve. For example, managers who are high in conscientiousness may be more likely than other managers to hold feedback meetings with their subordinates. Such managers may also be more likely to set goals and engage in effective self-management (e.g., Frayne & Latham, 1987; Latham & Frayne, 1989) to improve their performance. To address this issue, we also offer a within-subjects hypothesis.

Hypothesis 4: Managers will improve more in years when they meet with direct reports to discuss the previous year's feedback results than in years when they do not meet with direct reports to discuss the previous year's feedback results.

If a manager improves more in years when direct reports indicate that he or she held a feedback meeting than in other years, this improvement cannot be attributed to between-subjects, individual differences (such as conscientiousness, developmental orientation).

Method

Overview

The organization involved in this study is competing in a very mature business in which competitors are easily able to match products, services and pricing. As such, the organization believed that it can strategically set itself apart from its competition by developing strong leaders. The data collected in this study are part of a major ongoing organizational development effort

designed to identify, measure, and develop leadership potential within the organization. The emphasis of the program is developmental. A copy of the target manager's feedback report is, however, provided to the target manager's supervisor. While this practice of sending managers' reports up one level may detract from a purely developmental framework (Antonioni, 1994), managers were instructed to use reports only to guide the target manager's development and that the information in the report should not affect decisions about compensation or promotion, etc. Further, managers were told when the program was initially implemented that the data would be used "for developmental purposes only" and the organization has continued to adhere to this philosophy. Organizational leaders believe that a developmental orientation to upward feedback has several potential benefits. First, it helps to ensure that managers' energies and efforts will be directed toward the ultimate goal of changing and improving their behavior as opposed to merely improving or "managing" their scores. Second, organizational leaders believe that a developmental philosophy may eliminate dysfunctional fallout such as managers and employees exchanging "good" scores for "good" annual reviews, or managers becoming afraid of making tough decisions for fear of employee reprisal. Target managers do not complete self-ratings as part of this process.

Survey Development

The survey was constructed based on a series of employee and management focus groups designed to elicit behaviors believed to be associated with effective leadership, productivity, and implementation of strategic business objectives. The instrument consists of 29 behaviorally-based items (see appendix) and utilizes a five-point Likert-type format with responses ranging from 1 (Strongly Agree) to 5 (Strongly Disagree). Note that for this survey, as is the case of measurements of many types for financial institutions, the lower the score the better. A sixth

response point "Too New to Rate" was also added to the scale due to the fact that managers were fearful that many new employees may utilize a 3-rating (Sometimes Agree, Sometimes Disagree) when making their responses. In our analyses, this sixth response point was treated as missing data. Yearly principal components analyses with a varimax rotation consistently revealed a single component (using a decision rule of eigenvalues greater than one, item loadings .40 or higher on only one component and scree plot examinations) accounting for over 60% of the variance (results available from the first author). This single component is interpreted as an overall leadership component.

Because the organization sought to improve the quality of the survey every year, focus groups and evaluation surveys were conducted after each administration of the survey. Based on these evaluations, some items on the survey were either modified or replaced with other, more relevant items. Also, at the conclusion of each administrative year, meetings were held with senior executives in an effort to continuously ensure the survey instrument captured data relevant to the attainment of corporate goals. For example, during 1994, a major training initiative was implemented to help equip managers with empowerment leadership principles and practices. Therefore, senior management wanted to know if subordinates perceived their managers as implementing the behaviors taught during training. As a result of item changes from year to year, only 12 items are consistent or stable for the five years. This 12-item measure will subsequently be referred to as the 'core-item' measure or 'core item' total score.

Sample

The survey is administered within a major division of the organization employing approximately 5,000 employees, mostly within one Southeast state. All managers and supervisors with at least three direct reports are involved with the program. Two hundred fifty-two managers

participated in the program in all five years. One hundred twenty-two managers (48.4%) were female while 130 (51.6%) were male. The 252 managers had, on average, 19.74 years experience within the organization. Twenty-six managers (10.3%) were African American, while 226 managers (89.7%) were white non-Hispanic.

Procedure

The survey is administered annually. All managers and supervisors with at least three direct reports are contacted by phone to schedule their employees for a group survey administration that is facilitated by a representative of the human resources department. Work units located away from major centers or in rural areas are visited by personnel representatives who administer the survey in person. While this type of administration is quite time-consuming, it does have the advantage of substantially increasing overall participation and response rates. In fact, due to administrative thoroughness, the present organization virtually guarantees near 100% participation rates for managers and response rates of over 90% for subordinates for each of the five waves.

Surveys are returned to the human resources department for tabulation and report generation. All managers involved with the program receive a 5-page report which summarizes their individual data and presents comparisons to larger groups within the organization. Data are presented to managers in terms of item means, standard deviations, range (minimum score and maximum score) and valid N aggregated across subordinates. A leadership "total score" is also computed (based on the mean across the means on the individual items) and presented to managers.

After the first administration of the survey in 1991, all managers were required to attend a 2-hour training program. During this training, managers were provided an overview of the

process, information about the history of the program, and the program's intended purpose.

Managers were also provided with instruction in interpreting their data, facilitating a feedback meeting with their employees, and developing subsequent action plans. Managers were not held formally accountable for holding feedback meetings or developing action plans, although they were strongly encouraged to do so in initial training and all subsequent communications.

Tracking Feedback Meetings

Included in the survey since 1992 (i.e., in years 2, 3, 4, and 5) is the item "My manager has held a feedback session concerning last year's Leadership Survey with our work unit." This item allowed us to investigate score differences based on the extent to which raters agreed or disagreed that their manager held a meeting to discuss the prior year's feedback results (i.e., to test Hypotheses 3 and 4).

The feedback item was measured on the same five-point (1 = strongly agree to 5 = strongly disagree) Likert scale used for all other items. No action was ever taken by the organization based on results on this item and the results of this item were not used for scoring.

Before using this item to investigate any hypotheses, we wanted to be certain that responses to this item did not appear to merely reflect a generalized response bias (i.e., subordinates were not merely responding to this item based on some overall impression of the manager that may have affected their responses to other items in the upward feedback survey). To do so, we examined two aspects of subordinates' responses to this item: the item-total correlation and the distribution of responses to the item across the five-point rating scale.

First, we compared the item-total correlation for the feedback meeting item to the item-total correlations for all other items. Note that, if the item-total correlation for this item was equal to or higher than the item-total correlations for other items on the survey, it would suggest

that responses to the item reflected merely the rater's overall impression of the manager rather than measuring whether a feedback meeting had taken place. In every year, the item-total correlation for the feedback meeting item was lower than the item-total correlations for all other items, suggesting that this item was indeed measuring something different than other items (i.e., responses to this item did not merely reflect a generalized response bias).

Next, we examined whether neutral scores (i.e. "sometimes agree/sometimes disagree") were less prevalent and extreme scores (e.g., strongly disagree) were more prevalent for the feedback meeting item than for all other items. Results showed that in each year the feedback meeting item had a lower percentage of neutral responses than any other item; while the use of extreme scores (e.g., strongly disagree) was generally higher for the feedback meeting item than for other items. A chi square test was computed by comparing the response distribution of the feedback meeting item versus the response distribution for all other items combined for the 1995 data. Results of this analysis ($X^2 = 598$; $df = 4$; $p < .01$) revealed that there were significant differences in the response distributions for the feedback meeting item than for all other items. Consistent with the previous description, neutral scores (i.e. sometimes agree/sometimes disagree) were much less prevalent for the feedback meeting item while extreme scores (in particular, strongly disagree) were much more prevalent.

Taken together, these analyses indicate that subordinate responses to this item did not merely reflect a generalized response bias concerning the manager's leadership effectiveness.

Finally, in all our analyses, we treated the feedback meeting item as a continuous variable. For 1992 (year-2), 1993 (year-3), 1994 (year-4), and 1995 (year-5), each manager received a score that reflected the average subordinate rating given to the manager on this item. Note that disagreement among subordinates on this item could be expected for several reasons. First, a

subordinate may have been absent the day the feedback meeting occurred and therefore scored his or her manager a "4" (disagree) or "5" (strongly disagree) on the feedback item although, in fact, the meeting occurred. Second, some managers may have "sandwiched" discussion about their feedback into the agenda for a meeting that addressed many other topics. This could make the feedback session less salient to some subordinates and create disagreement about whether a feedback session had occurred. Third, some subordinates may have realized that their manager was attempting to conduct a meeting about his or her feedback, but was either unsuccessful or incompetent in conducting it thereby leaving them uncomfortable with "agreeing" that a feedback meeting occurred. By treating the feedback meeting item as a continuous variable, we avoided the problem of taking a naturally occurring continuous variable and arbitrarily deciding whether a feedback meeting occurred or did not occur.

Feedback Training

As mentioned above, in 1991, managers were required to attend a two-hour training session during which they received their individualized feedback reports. A major part of this training program was a module devoted to providing instruction in conducting feedback meetings with subordinates. This module had three major components. During the first component, managers were instructed to identify their three poorest-scored items (or the 2-3 item scores that concerned them the most). Managers then received instruction in using a "Feedback Planning Worksheet" designed to help them prepare for, elicit, and receive feedback on these three items. The second component presented the Johari Window (Luft, 1961) as a model to convey the overall goal or purpose of the process. Specifically, managers were instructed that the program, in general, and the feedback meeting in particular, can lead to the identification and discussion of possible "blind spots" in their leadership behaviors. Using this information managers can then

develop, with the help of their subordinates, action plans to address these weaker areas. Likewise, managers were instructed that the feedback meeting provides an opportunity for them to communicate areas of the "facade," that is, things that the manager is aware of, but that subordinates may be unaware of. The last component of the feedback training module presented managers with a list of "do's and don'ts" for conducting their feedback meetings. For example, on the "do" side, managers were encouraged first to discuss 1 or 2 positively scored items (thereby starting on a positive note) and to encourage everyone to participate. On the "don't" side, managers were warned about overreacting to negative scores and getting into fruitless debates about the wording of specific items. Finally, managers viewed a video tape depicting a very negative and unsuccessful feedback meeting and then a very positive and successful meeting. Managers were then led through a facilitated discussion that compared and contrasted the two feedback meetings and what made each successful or unsuccessful.

Results

Table 1 presents the internal consistency of the ratings and the number of managers and employees (raters) involved by year. As can be seen, internal consistency is quite high across all years.

For each year, we calculated a total score for each manager by averaging responses to all items (except, as noted above, the feedback meeting item) across all subordinates. Table 2 presents correlations between managers' scores by year for both the all-item and core-item total scores. As can be seen, correlations were fairly stable over a one-year period, but declined over time. This was observed for both the total score composed of all items as well as the core-item total score. Of course, the decline in correlations over time may reflect the fact that some

managers changed their leadership behaviors over time. Table 3 presents the mean (and standard deviation) of the total scores in each year.

Insert Tables 1, 2, and 3 about here

In order to determine the appropriateness of the composite measures used in the present study, within-group interrater agreement was assessed using the r_{wg} formula developed by James, Demaree and Wolf (1993; James, Demaree & Wolf, 1984, & Kozlowski & Hattrup, 1992). Yearly descriptive statistics for the total score revealed moderate to large negative skewness (.64 in 1995 to .84 in 1991) as is often found for surveys of this type. Using the formula for estimating s_e^2 for five-point scales put forth by James, Demaree and Wolf (1984) for samples with moderate to large negative skew, r_{wg} was found to be acceptable for each year (.77 in 1991, .73 in 1992, .52 in 1993, .76 in 1994, and .55 in 1995).

As mentioned above, only 12 items on the survey remained stable over the five-year period. To determine if the use of different total scores (i.e., 'core item' total score vs. all-item total score) could result in different findings, we conducted separate analyses treating each total score as the dependent variable. Results were the same regardless of whether the 'core-item' total score or the all-item total score was used as the dependent variable. Further, as shown in Tables 1 and 3, internal consistency, means, and standard deviations were found to be highly similar for both measures across all five years. All subsequent results are therefore presented only for the all-item total score (hereafter referred to as "score").

Hypothesis 1: Managers' performance (as rated by their direct reports) will improve over the five year period.

To examine this hypothesis, we used procedures described by Cohen and Cohen (1983) for repeated measures regression analysis (pp. 428 - 451). Each manager had an upward feedback score representing direct reports' ratings of the manager's performance (hereafter referred to as 'score') for each of five years (years 1, 2, 3, 4, and 5).

The variance of 'score' where $N = 1260$ (five scores for each of 252 managers) is .199. This is the total variance that we want to partition and explain in examining the hypothesis. We first determined the proportion of 'score' variance that is between-subjects variance. Note that we did not seek to explain this between-subjects variance in our present analysis, but we needed to determine the between-subjects variance so that we could calculate the relevant error terms (per Cohen and Cohen, page 430, 1983). We determined the between-subjects variance in 'score' by calculating the average 'score' for each manager across the five years. We then calculated the variance of this average across managers. This between-subjects variance for 'score' was .125. The ratio of between-subjects variance to total variance ($.125/.199$) was .63. That is, 63% of the variance in 'score' is between-subjects variance.

The next step was to determine the amount of within-subjects variance in 'score' explained by the linear trend of year (1, 2, 3, 4, 5). On the first step of a hierarchical regression equation, we regressed 'score' onto year. The R^2 was .02 ($F = 16.26$, $df = 4,1004$, $p < .01$). That is, the linear trend of year explained significant variance in 'score,' with upward feedback scores improving over time. This finding supports hypothesis 1.

Hypothesis 2: Managers who initially receive the poorest ratings will improve more than managers who initially receive more favorable ratings.

On the next step of the regression analysis, we added managers' initial upward feedback scores (from year 1) to the equation. (Note that any variance in 'score' explained by managers'

initial score is between-subjects variance and is not of direct interest for examining this hypothesis.) On the next step of the regression analysis, we entered the interaction between year and managers' initial scores. The increment in R^2 was .04 ($F = 34.96$, $df = 4$, 1004 , $p < .01$). The presence of this significant interaction indicates that the relationship between year and 'score' varies depending on the manager's initial upward feedback score.

To illustrate this interaction, we first divided managers into three groups (low, middle, and high) based on their initial (year-1) upward feedback scores and then separately conducted the analysis described above to examine hypothesis 1 for each the three groups.

For managers initially receiving the least favorable feedback, the variance in 'score' explained by the linear trend in year was $R^2 = .13$ (standardized beta = $-.36$, $F = 26.13$, $df = 4$, 332 , $p < .01$). The negative beta coefficient indicates that as year increased (1, 2, 3, 4, 5) upward feedback scores became more favorable (remember that low ratings represented more favorable feedback).

For managers initially in the middle group, the variance in 'score' explained by the linear trend in year was $R^2 = .02$ (standardized beta = $-.15$, $F = 3.79$, $df = 4$, 340 , $p < .01$). The negative beta coefficient indicates that as year increased (1, 2, 3, 4, 5) upward feedback scores became more favorable (again, note that low ratings represented more favorable feedback).

For managers initially receiving the most favorable feedback, the variance in 'score' explained by the linear trend in year was $R^2 = .01$ (standardized beta = $+.10$, $F = 1.48$, $df = 4$, 324 , $p > .05$). The positive beta coefficient indicates that as year increased (1, 2, 3, 4, 5) upward feedback scores became less favorable (i.e., high ratings represented less favorable feedback), however this trend was not statistically significant.

Of course, some improvement in scores would be expected due to regression to the mean for managers initially rated poorly. And regression to the mean would also be expected to lead to slightly less favorable ratings for managers initially rated most favorably. We therefore estimated the amount of score improvement or deterioration that could be expected solely due to regression to the mean (using the procedures described by Smither et al., 1995). First, we used the average year-to-year correlation across the five year period (i.e., .69, .61, .63, .46) as a conservative estimate of reliability ($r = .60$). We then used this estimate of reliability in a bivariate regression equation to predict mean year-5 scores in each of the three groups based on the mean year-1 scores for each group.

For those managers who initially received the least favorable ratings in year-1, regression to the mean would predict an improvement in scores from 2.59 to 2.39. Actual score improvement for these managers was 2.59 to 2.14 which represents a mean gain of .25 over the value for predicted performance based on regression to the mean. Further, the effect size (i.e., mean score gain divided by the pooled within-group standard deviation) was moderate ($d = .64$). Thus, only a small part of the improvement in scores could be attributed to regression to the mean for managers who initially received unfavorable ratings.

For those managers who initially received moderate feedback (i.e., the middle third of managers), regression to the mean would predict a slight *deterioration* in scores from 2.05 to 2.07. These managers however actually *improved* their scores over the five-year period from 2.05 to 1.93 resulting in a mean score improvement gain of .14 over the value for predicted performance based on regression to the mean. The effect size was moderate ($d = .67$). That is, none of the score improvement for managers who initially received moderate ratings can be attributed to regression to the mean.

Finally, for managers who initially scored most favorably, regression to the mean would predict a deterioration in scores from 1.65 to 1.85. These managers' scores actually deteriorated from 1.65 to 1.78 over the five-year period which represents a .07 mean score improvement over the value for predicted performance based on regression to the mean. The effect size was small ($d = .22$). Thus, regression to the mean can explain all of the apparent score deterioration for those managers who initially received the most favorable ratings.

Taken as a whole, these regression to the mean analyses mirror the results reported by Smither et al. (1995) and Reilly et al. (1996). That is, only a small part of the score improvement for those managers who initially received the most unfavorable ratings can be attributed to regression to the mean. For those managers who initially received moderate ratings, none of the score improvement can be attributed to regression to the mean. For those managers who initially scored most favorably, all of their apparent score deterioration can be attributed solely to regression to the mean. Further, the effect size associated with real improvement (i.e., beyond regression to the mean) was found to be moderate in size for those managers who initially received unfavorable or moderate ratings. Taken together, these results offer strong support for hypothesis 2.

Hypothesis 3: Managers who meet with direct reports to discuss the previous year's feedback results will improve more than managers who do not meet with direct reports to discuss the previous year's feedback results.

To examine this hypothesis, we used a traditional approach to change analysis where a time-2 score serves as a criterion measure, a time-1 score serves as a covariate, and the independent variable of interest (in this case, holding a meeting with direct reports to discuss the previous year's feedback) is entered on the second step of the regression equation.

We conducted four hierarchical regression analyses. For the first analysis, we first treated each manager's year-2 score as the dependent variable in a hierarchical regression analysis. On the first step of the regression analysis, we entered year-1 scores. On the second step of the analysis we entered 'feedback' (that is, direct reports' ratings concerning whether the manager had conducted a meeting to discuss the previous year's feedback). On the third step of the analysis, we entered the interaction between year-1 scores and 'feedback' (thereby examining whether improvement in score associated with discussing the previous year's feedback may have differed depending on the manager's score in the previous year). We conducted parallel analyses treating scores from year-3, year-4, and year-5 as dependent variables. For each analysis we entered the score from the previous year on the first step of the analysis and on the second step we entered the 'feedback' score indicating whether the manager had discussed the previous year's feedback with direct reports. The results of all these analyses are presented in Table 4.

Insert Table 4 about here

It can be seen that, in each year, improvement in 'score' from the previous year was greater for managers who had discussed the previous year's feedback with direct reports than for managers who had not discussed the previous year's feedback with direct reports (all F values were significant, $p < .01$). These findings support hypothesis 3.

In each analysis, the interaction term was not significant thereby indicating that the association between discussing the previous year's feedback and improvement was similar regardless of the manager's score in the previous year.

Hypothesis 4: Managers will improve more in years when they meet with direct reports to discuss the previous year's feedback results than in years when they do not meet with direct reports to discuss the previous year's feedback results.

To examine the within-subjects effect of holding feedback meetings, we included only those managers who had a feedback meeting score for each of the four years. That is, each manager had a score (hereafter referred to as 'feedback meeting') for years 2, 3, 4, and 5 that reflected direct reports' ratings concerning whether the manager had conducted a meeting to discuss the previous year's feedback. Thus, this analysis was based on 224 managers.

Each manager also had four scores (hereafter referred to as 'score') representing direct reports' ratings of the manager's performance in each year (2, 3, 4, and 5). (Note that we did not include managers' year-1 scores in these analyses because managers could not hold a feedback meeting prior to receiving their year-1 feedback. That is, the effect of holding a feedback meeting to discuss the previous year's feedback could be evident only for managers' scores in year-2, year-3, year-4, and year-5.)

To examine this hypothesis, our analyses paralleled procedures described by Cohen and Cohen (1983) for repeated measures regression analysis (pp. 428 - 451). The variance of 'score' where $N = 896$ (four scores for each of 224 managers) was .196. This is the total variance that we wanted to partition and explain in examining this hypothesis. We first determined the proportion of 'score' variance that is between-subjects variance. Note that we did not seek to explain this between-subjects variance in our present analyses, but we needed to determine the between-subjects variance so that we could calculate the relevant error terms (per Cohen and Cohen, page 430, 1983). We determined the between-subjects variance in 'score' by calculating the average 'score' for each subject across the four years. We then calculated the variance of this

average across subjects. This between-subjects variance for 'score' was .129. The ratio of between-subjects variance to total variance ($.129/.196$) = .66. That is, 66% of the variance in 'score' is between-subjects variance.

The next step was to determine the amount of variance in 'score' explained by holding a feedback meeting in the previous year. There were between-manager differences in the average feedback meeting score (across the four years). We therefore first removed all between-subjects variance in 'feedback meeting.' This is important because the test of this hypothesis requires that we focus only on within-subjects covariance. Stated differently, to test this hypothesis, we do not want to consider any between-subjects covariance between holding a feedback meeting and 'score.' We therefore removed the between-subjects variance in feedback score by first calculating each manager's mean feedback score across the four years. For each manager, we then subtracted the manager's mean feedback score (across the four years) from the manager's feedback score in each year. The resulting score ('adjusted feedback meeting score') retains all within-subjects variance (i.e., showing the years in which each manager was more or less likely to hold a feedback meeting) while removing between subjects variance in holding a feedback meeting. Note that the within-subjects correlation between 'feedback meeting' and 'adjusted feedback meeting score' equals 1.00. Note also that each manager has the same mean 'adjusted feedback meeting score' (i.e., 0) across the four years (i.e., there is no between-subjects variance in 'adjusted feedback meeting score'). As a result, any covariance between 'adjusted feedback meeting score' and 'score' will reflect only within-subjects covariance. If we did not make the adjustment described above, the R^2 associated with feedback meetings would reflect both within-subjects covariance (between holding a feedback meeting and score across the four years) and between-subjects covariance (between holding a feedback meeting and score across managers),

and would thus provide an inflated estimate of the within-subjects covariance that is relevant for testing this hypothesis.

To examine hypothesis 4, we used a hierarchical regression model where 'score' was the dependent variable. The variables entered on each step and the increase in R^2 at each step are presented in Table 5.

Insert Table 5 about here

The first step removes any within-subjects variance in 'score' explained by the linear trend in year (i.e., the tendency to improve over time). The second step removes between-subjects variance in 'score' due to each manager's initial (year 1) score. The third step removes within-subjects variance in 'score' due to the interaction between year and initial score. Note that these analyses parallel those presented earlier to examine hypotheses 1 and 2. By including these variables in the present analyses, we ensure that any within-subjects variance in 'score' explained by 'adjusted feedback meeting score' will be unique variance (i.e., not explained by previously examined variables).

On the fourth step, we remove any variance in 'score' explained by the manager's previous year's 'score'. This is similar to traditional change analysis models where a time-2 score serves as a criterion measure, a time-1 score serves as a covariate, and the independent variable of interest is entered on a subsequent step of the regression equation. This analysis (treating the previous year's score as a covariate so that each subject has a covariate for each of four dependent variables) directly parallels what Keppel (1982, pp. 511-512) refers to as within subjects designs where "a covariate score is obtained before each of the treatments a subject

receives.” As Keppel states, “the analysis of covariance can be used to increase the sensitivity of within-subjects designs” via “the use of a covariate to adjust within-subject sources ... this requires a different covariate score ... before the administration of each of the treatments given to a subject.”

On the fifth step, we enter the adjusted feedback meeting score. This examines the remaining within-subjects variance in ‘score’ associated with holding a meeting to discuss the previous year’s feedback. The incremental R^2 associated with this step (i.e., .058) represents the within-subjects covariance of interest for testing hypothesis 4. Note that adjusted feedback meeting score explains about 6% of the variance in ‘score’ after accounting for any variance in ‘score’ explained by the linear trend of year, the interaction of year and the subject’s initial score, and the subject’s previous year’s score. This increment in R^2 is significant ($F = 52.36$, $df = 3$, 669 , $P < .01$) thereby supporting hypothesis 4.

On the final step, we enter the interaction between adjusted feedback meeting score and the subject’s initial score. This examines whether the within-subjects covariance between ‘score’ and adjusted feedback meeting score varies as a function of subjects’ initial scores. The incremental R^2 associated with this step (.001) was very small, thereby indicating that holding a feedback meeting was associated with improvements in score regardless of the manager’s initial score.

Relationship of Manager’s Scores, Gender, Level, and Work Unit Size with Likelihood to Hold Feedback Meetings

Given the relationship between holding feedback meetings and score improvement, we examined whether several variables were related to the likelihood that a manager would hold a feedback meeting.

Managers' scores in each year were correlated with the likelihood that they would hold a feedback meeting to discuss the feedback results. For example, the correlation between year-1 scores and holding a feedback meeting to discuss the year-1 feedback results was .31, ($p < .01$). The corresponding correlations for year-2, year-3, and year-4 scores and subsequently meeting with direct reports to discuss the feedback results were .37, .41, and .24 (all $p < .01$). Also, managers' initial scores (in year-1) were significantly related to the likelihood that they would hold a feedback meeting to discuss their year-1, year-2, year-3, and year-4 feedback results. That is, managers who received favorable feedback were more likely than other managers to hold a meeting with direct reports to discuss the feedback results.

In the analyses that follow, the effects of managers' initial scores were partialled out (given the relationships presented in the previous paragraph). Thus, if significant relationships were detected between any of these demographic variables and number of feedback meetings held, the results could not be attributed merely to managers' initial scores.

Managers' gender was unrelated to the likelihood of holding a feedback meeting in any year (all correlations between $-.06$ and $+.05$, all $p > .30$). Managers' organizational level was unrelated to the likelihood of holding a feedback meeting to discuss year-1, year-2, or year-3 feedback results. There was a small but significant correlation between level and holding a meeting to discuss year-4 feedback results ($r = .14$, $p < .05$) such that lower-level managers were more likely than other managers to meet with their direct reports to discuss their year-4 feedback.

The relationship between work unit size and holding a feedback meeting was also examined. In this analysis, managers' average number of respondents per year were averaged across the five survey administrations. This average group size was used as a proxy for work unit size. Work unit size was unrelated to the likelihood of holding a feedback meeting in each year.

In sum, demographic and work group characteristics were generally unrelated to managers' likelihood of holding a meeting with direct reports to discuss the feedback results.

Discussion

The results presented above support each of the four hypotheses. First, managers' performance (as reflected in ratings from their direct reports) improved over the five year period. Second, managers who initially received poor or moderate ratings improved more than managers who initially received more favorable ratings. The pattern of results is consistent with Smither et al. (1995) and Reilly et al. (1996) but in a different setting and over a longer period of time, thereby providing evidence of the generalizability of the positive benefits of upward feedback, especially for those who initially receive poor or moderate ratings.

Third, managers who held feedback sessions to discuss their upward feedback with direct reports improved more than other managers. Fourth, within-subjects analyses showed that managers improved more in years when they held feedback meetings with direct reports than in year when they did not hold such meetings. This is important because it is the first research evidence demonstrating that what managers do with upward feedback is related to its benefits.

The within-subjects finding (i.e., managers improve more in years when they hold feedback meetings than in years when they do not) is especially important. As we noted above, our confirmation of hypothesis 3 (i.e., managers who hold feedback meetings improve more than managers who do not hold such meetings) does not allow us to rule out the possibility that a third unmeasured variable (e.g., conscientiousness, developmental orientation) may have affected both the likelihood that a manager will hold feedback meetings and the likelihood that the manager will improve. However, the finding that a manager improves more in years when direct reports indicate that he or she held a feedback meeting than in other years suggests that such

improvement cannot be attributed to between-subjects, individual differences (e.g., conscientiousness, developmental orientation). This strengthens our conclusion that meeting with direct reports to discuss upward feedback results enhances a manager's likelihood of improving over time.

The present results have practical implications because they suggest that upward feedback programs are likely to lead to improvements in feedback scores that can be sustained over many years. Also, it appears that the value of upward feedback can be enhanced when managers hold meetings with their direct reports to discuss the feedback.

These results thereby suggest the benefits that may accrue to organizations that actively consider ways to increase the accountability of ratees to use upward or multisource feedback. The approach described here is only one of many possible approaches to increase accountability. London et al. (1997) described other interventions that may increase ratee accountability such as discussing the results with a skilled facilitator, creating a development plan, publicly committing to behavior change, participating in training linked to feedback results, and rewards for continuous improvement.

By devoting developmental resources to support upward feedback and by increasing the accountability of ratees (e.g., using the above approaches), organizations create a climate where ratees know they are expected to use the feedback to improve performance. However, a survey by London and Smither (1995) found that only 40% of respondents reported that multisource feedback was linked to a specific development program. A process put into place by one company provides a good example of how multisource or upward feedback can be linked to a development program. In this organization, managers were rated by their direct reports on a few key organizational competencies prior to attending a week-long development program. On each

morning of the program, the participants received upward feedback relevant to the competency that would be developed during the remainder of the day. For example, if the day's training focused on conflict resolution skills, then participants would start the day by receiving upward feedback about their conflict resolution skills. At the end of day (after completing the training module), the participant would create goals and action plans related to conflict resolution.

Yukl and Lepsinger (1995) also suggest that the benefits of multisource feedback are more likely to be achieved if other follow-up activities also support the process. These activities include not only formal training sessions, but also on-the-job learning opportunities for competencies identified as developmental needs. Also participants should be provided with the opportunity to obtain on-the-job coaching from supervisors or peers on identified strengths and developmental needs.

Another interesting feature of these results is that there was no improvement in the mean score between year-1 and year-2 (see Table 3). Unlike previous research that has reported significant score improvement on the second administration after relatively short periods of time (e.g. six months in the Smither et al., 1995 study and 32 weeks in the Atwater et al., 1995 study), score improvement for this sample of managers was more gradual. This suggests that that organizations may need to be patient and persistent in their use of upward feedback.

There are a number of limitations to the present study that point to the need for additional research. For example, this was not an experimental study (i.e., managers were not randomly assigned to hold or not hold feedback meetings with their direct reports). Also, although the content of the survey was based on focus groups designed to elicit behaviors believed to be associated with effective leadership and the organization's strategic objectives, the validity of the instrument was unknown. Also, due to practical constraints, we were not able to measure the

extent to which ratees felt accountable to change their behavior and improve their performance. Moreover, we do not know if holding feedback meetings increased ratees' sense of accountability. An alternative explanation could be that ratees who held feedback meetings in a given year did so because they already felt a greater sense of accountability to make use of the upward feedback. Thus, it is not clear that forcing managers to hold feedback meetings will necessarily be effective. Instead, we believe the approach used here (encouragement and training) is more likely to be accepted by ratees. This approach seems especially reasonable in view of the fact that the majority of managers in the present study held three or four feedback meetings.

Holding feedback meetings could have enhanced the impact of upward feedback by increasing accountability (a motivational effect), providing ratees with additional information (clarification of feedback, constructive suggestions for behavior change) or both. Further, by conducting feedback meetings, managers may have created a more supportive context or environment for positive behavior change to occur. That is, the feedback meeting may provide a framework in which employees (raters) work with their managers in creating a developmental plan. This may in turn lead to employees being more supportive of the changes the manager is trying to make. For example, they may feel more empowered to give ongoing feedback, be more forgiving of mistakes as the manager tries to improve, or they may respond better to new behaviors since they have become aware of and better understand the manager's intent. Future research should try to disentangle the independent contributions of additional information and increased accountability that may result from holding feedback meetings with raters.

One could also argue that managers who hold feedback meetings subsequently receive more favorable ratings not because they improve their performance but instead because they use the meetings as part of an impression management strategy designed to convince subordinates

that the manager is concerned with their input and willing to change. On the other hand, this could be a very dangerous strategy. If a manager makes commitments and thereby increases subordinates' expectations that he or she will improve, then failure to follow-through on those commitments is likely to create distrust and disappointment among subordinates and lead to less favorable ratings of the manager.

We also suggest caution against concluding from our results (and previous results reported by Smither et al., 1995, and Reilly et al., 1996) that managers who initially receive very favorable feedback do not subsequently improve. The finding that such managers do not improve their scores over time may be due to ceiling effects. For example, in the present study, the average score received in year-1 by those managers initially rated most favorably was 1.65. This leaves very little room on the 1 (most favorable) to 5 (least favorable) rating scale for such managers to improve in subsequent years. To address this issue, future research could ask direct reports to rate the extent to which each manager has improved since receiving the previous feedback report.

There are several factors that may affect the generalizability of the present results. For example, managers were given considerable guidance about how to conduct feedback meetings. It is not clear whether similar benefits would be observed where managers receive little guidance (e.g., would some managers conduct ineffective meetings in which goals are not formulated or direct reports feel threatened?). Also, each manager's supervisor also received a copy of the manager's feedback report (although it was emphasized that the feedback was to be used only for developmental purposes). This may create a climate where managers assume that their supervisors expect behavior/performance improvement. Perhaps managers' meetings with direct

reports would have less effect where managers are not also concerned about the expectations of their supervisors.

All of this points to the need for additional research. Perhaps most important is the need for researchers to better understand what ratees do after they receive upward or multisource feedback. It would also be helpful to examine other approaches to increase ratee accountability such as those outlined above. It would be especially valuable to use an experimental (or quasi-experimental) design where some managers are assigned to conduct feedback meetings with raters while other managers do not hold such meetings. Another approach could involve developing a questionnaire that measures the extent to which ratees feel accountable to different sources (e.g., supervisor, peers, subordinates, customers, the organization, oneself). It would be useful to examine whether there are differences between the impact of accountability to others (meeting coworkers' expectations or receiving rewards for performance improvement) vs. the impact of self-imposed accountability. Finally, research on multisource feedback (rather than upward feedback alone) is needed. Managers may be better able to improve their performance in response to upward feedback than in response to multisource feedback (from peers, customers, direct reports and the supervisor). For example, upward feedback provides a single focus for the manager's improvement efforts (as opposed to addressing the potentially complex, perhaps even conflicting, expectations of several rating sources). Also, it may sometimes be easier for managers to change their behavior toward their subordinates (e.g., holding team meetings, making decisions more quickly, improving day-to-day communication) than it is to change their behavior toward others such as peers or customers where they have less control over the timing and nature of interactions.

From a theoretical perspective, we think the results suggest there is value in using an accountability framework such as that proposed by London et al. (1997) to develop upward feedback program components and characteristics that are likely to enhance the impact of such feedback. Dominick, Reilly, and McGourty (1997) have recently suggested that it is not so much feedback from others that drives behavior change but rather exposure to and completion of the feedback instrument. But the present results indicate that actually receiving feedback is valuable to the extent that feedback results are subsequently discussed with coworkers. In doing so, these results affirm the importance of focusing on what managers do with the feedback they receive and the steps that organizations can take to encourage managers to use feedback to guide behavior change.

References

- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. Personnel Psychology, 47, 349-356.
- Ashford SJ, Tsui AS. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. Academy of Management Journal, 34(2), 251-280.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self-and follower ratings of leadership. Personnel Psychology, 48, 34-59.
- Atwater, L., Waldman, D., Atwater, D., & Cartier, P. (1998). The effects of follow-up and cynicism: An upward feedback experiment. Arizona State University West Working Paper.
- Bernardin, J.H., Beatty, R.W. (1987). Can subordinate appraisals enhance managerial productivity? Sloan Management Review, 28(4), 63-73.
- Bernardin, H.J., Dahmus, S.A. & Redmon, G. (1993). Attitudes of firstline supervisors toward subordinate appraisals. Human Resource Management, 32, 315-324.
- Bernardin, H.J., Hagan, C., Ross, S. & Kane, J.S. (1995, May). The effects of a 360-degree appraisal system on managerial performance. Paper presented in a symposium titled "Upward feedback: The ups and downs of it" (Chair: W.W. Tornow) at the Tenth Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences, second edition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Dalessio, A.T. (1998). Using multi-source feedback for employee development and personnel decisions. In J.W. Smither (Ed.), Performance Appraisal: State-of-the-Art in Practice. San Francisco: Jossey-Bass.

Dominick, P., Reilly, R.R. & McGourty, J. (1997). The effects of peer feedback on team member behavior. Group and Organization Management, 22, 508-520.

Frayne, C.A. & Latham, G.P. (1987). The application of social learning theory to employee self-management of attendance. Journal of Applied Psychology, 72, 387-392.

Graddick, M.M., & Lane, P. (1998). Evaluating executive performance. In J.W. Smither (Ed.), Performance Appraisal: State-of-the-Art in Practice. San Francisco: Jossey-Bass.

Hauenstein, N.M.A. (1998). Training raters to increase the accuracy of appraisals and the usefulness of feedback. In J.W. Smither (Ed.), Performance Appraisal: State-of-the-Art in Practice. San Francisco: Jossey-Bass.

Hazucha, J.F., Hezlett, S.A. & Schneider, R.J. (1993). The impact of 360 degree feedback on management skills development. Human Resource Management, 32, 325-351.

Hegarty, W.H. (1974). Using subordinate ratings to elicit behavioral changes in managers. Journal of Applied Psychology, 59, 764-766.

Hezlett, S.A., Kuncel, N.R., & Cochran, C.C. (1997) Individual difference predictors of self-other discrepancies in performance ratings. Paper presented in a symposium titled “Antecedents and Consequences of Interrater Agreement in Multirater Feedback” (Chair: Jeff Johnson) at the Twelfth Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, Missouri.

James, L. R., Demaree, R. G. and Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. Journal of Applied Psychology, 69, 85-98.

James, L. R., Demaree, R. G., & Wolf, G. (1993). r_{wg} : An assessment of within-group interrater agreement. Journal of Applied Psychology, 78, 306-309.

Johnson, J.W., & Ferstl, K.L. (1997, April). The Effects of Inter-Rater and Self-Other Agreement on Performance Improvement Following Upward Feedback. Paper presented in a symposium titled “Antecedents and Consequences of Interrater Agreement in Multi-Rater Feedback” (Chair: Jeff Johnson) at the Twelfth Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, Missouri.

Keppel, G. (1982). *Design and analysis: A researcher’s handbook*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, *119*, 254-284.

Kozlowski, S.W.J. and Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. Journal of Applied Psychology, *77*, 161-167.

Kozlowski, S.W.J., Chao, G.T., & Morrison, R.F. (1998). Games raters play: Politics, strategies, and impression management in performance appraisal. In J.W. Smither (Ed.), Performance Appraisal: State-of-the-Art in Practice. San Francisco: Jossey-Bass.

Latham, G.P. & Frayne, C.A. (1989). Self-management training for increasing job attendance: A follow-up and replication. Journal of Applied Psychology, *72*, 411-416.

Locke, E.A. & Latham, G.P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice Hall.

London, M. & Smither, J.W. (1995). Can multisource feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. Personnel Psychology, *48*, 803-839.

London, M., Smither, J.W., & Adsit, D.J. (1997). Accountability: The Achilles heel of multisource feedback. Group and Organization Management, 22, 162-184.

London, M. & Wohlers, A.J. (1991). Agreement between subordinate and self-ratings in upward feedback. Personnel Psychology, 44, 375-390.

London, M., Wohlers, A.J., & Gallagher, P. (1990). 360 degree feedback surveys: A source of feedback to guide management development. Journal of Management Development, 9, 17-31.

Luft, J. (1961). The johari window. Human Relations Training News, 5, 6-7.

Reilly, R. R., Smither, J. W. & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. Personnel Psychology, 49, 599-612.

Romano, C. (1994). Conquering the fear of feedback. HR Focus, 71(3), 9-19.

Smither, J.W., Wohlers, A.J., & London, M. (1995). A field study of reactions to normative vs. individualized upward feedback. Group and Organization Management, 20, 61-89.

Smither, J. W., London, M., Vasilopoulos, N. L., Reilly, R. R., Millsap, R. E., Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. Personnel Psychology, 48, 1-34.

Timmreck, C.W. (1995, May). Upward feedback in the trenches: Challenges and realities. Paper presented in a symposium titled "Upward feedback: The ups and downs of it" (W.W. Tornow: Chair) at the Tenth Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, Florida.

Tornow, W.W. (1993a). Editor's note: Introduction to special issue on 360-degree feedback. Human Resource Management, 32, 211-219.

Van Velsor E., Leslie, J.B. (1991a). Feedback to managers, Volume 1: A guide to evaluating multirater feedback instruments. Greensboro, NC: Center for Creative Leadership.

Van Velsor E., Leslie, J.B. (1991b). Feedback to managers, Volume 2: A review and comparison of sixteen multirater feedback instruments. Greensboro, NC: Center for Creative Leadership.

Yukl, G. & Lepsinger, R. (1995). How to get the most out of 360-degree feedback, Training, 32, 45-50.

Table 1

Number of Participants and Internal Consistency by Year

	1991	1992	1993	1994	1995
Number of Employees (Raters)	3,397	3,499	3,817	4,031	4,256
Number of Managers	473	512	560	591	636
Coefficient Alpha (Number of Items)	.97 (22)	.97 (23)	.96 (26)	.97 (26)	.98 (27)
Coefficient Alpha for Core-Items	.95	.93	.94	.95	.95

Note: Entries represent survey administration dates, number of employees involved per year, number of managers involved per year, coefficient alpha for all-item total score, coefficient alpha for core-item total score.

Table 2

Correlations Between Managers' Scores by Year

	1991	1992	1993	1994	1995
1991, Year-1	-	.69 (.67)	.56 (.56)	.45 (.42)	.35 (.35)
1992, Year-2		-	.61 (.62)	.48 (.47)	.36 (.35)
1993, Year-3			-	.63 (.62)	.48 (.48)
1994, Year-4				-	.46 (.44)
1995, Year-5					-

Note: Entries are correlations for the all-item total score (core-item total score correlations are in parentheses).

Table 3

Managers' Mean Total Scores by Year

Year	Mean	SD	Minimum	Maximum
1991, Year-1	2.10 (2.11)	.44 (.44)	1.20 (1.17)	3.86 (3.96)
1992, Year-2	2.10 (2.12)	.46 (.46)	1.13 (1.19)	3.43 (3.43)
1993, Year-3	2.00 (2.00)	.45 (.45)	1.03 (1.01)	3.54 (3.54)
1994, Year 4	1.96 (1.94)	.43 (.42)	1.13 (1.17)	3.35 (3.27)
1995, Year-5	1.95 (1.95)	.42 (.43)	1.12 (1.11)	3.28 (3.35)

Note: Lower scores represent higher levels of performance. Entries represent means, standard deviations, minimum and maximum scores for both the all-item total score and the core-item total score (in parentheses) for the 252 managers involved in the program over the entire 5-year period.

Table 4

Hierarchical Regression Analysis for Hypothesis 3

Dependent Variable	Variable Entered on Step 1 (Previous Year's Score)	R ² after Step 1	Variable Entered on Step 2 (Discussed Previous Year's Feedback with Direct Reports?)	R ² after Step 2	Change in R ²	F	R ² after Entering Interaction Term	Change in R ²
Year-2 Score	Year-1 Score	.52	Discussed year-1 feedback?	.58	.06	31.41 *	.58	.00
Year-3 Score	Year-2 Score	.47	Discussed year-2 feedback?	.57	.10	50.42 *	.57	.00
Year-4 Score	Year-3 Score	.47	Discussed year-3 feedback?	.56	.09	47.61 *	.56	.00
Year-5 Score	Year-4 Score	.23	Discussed year-4 feedback?	.40	.17	63.71	.41	.01

Note: For all F tests in table, $df = 1, 221$. No interaction terms reached significance.

* $p < .01$.

Table 5

Hierarchical Regression Analyses for Hypothesis 4

Dependent Variable is 'Score'

Step	Variable	R^2	Increase in R^2 from previous step
1	Year (2, 3, 4, or 5)	.015	.015
2	Subject's Initial Score (in Year 1)	.331	.316
3	Interaction of Year and Subject's Initial Score	.350	.019
4	Subject's Score in the Preceding Year	.450	.100
5	Adjusted Feedback Score i.e., Did the subject meet with his or her direct reports to discuss the previous year's feedback?	.508	.058
6	Interaction of Adjusted Feedback Score with Initial Score	.509	.001

Appendix: 1995 Leadership Survey Items

(1) = Strongly Agree (2) = Agree (3) = Sometimes Agree/Sometimes Disagree (4) = Disagree (5) = Strongly Disagree

- * 1. My manager shares with me the information I need to do my job.
- * 2. When I need it, my manager provides information about how I'm performing my job.
- 3. My manager has helped define the boundaries of empowerment for my position.
- 4. My manager promotes teamwork within our work unit.
- 5. My manager promotes teamwork between people in our work unit and people in other work units including those companywide.
- * 6. My manager listens to my suggestions.
- * 7. My manager keeps me informed on what the company is trying to accomplish.
- * 8. My manager keeps me informed on what our work unit is trying to accomplish.
- 9. My manager involves our work unit in continuously improving the way we service our customers.
- * 10. My manager encourages me to develop myself.
- * 11. My manager makes sure I am trained to do my job.
- 12. My manager treats me with respect.
- * 13. My manager lets me know when I've done a good job.
- * 14. My manager presents a positive attitude toward the company and company policy.
- 15. My manager shares how Empowerment, Family Matters, and Continuous Improvement initiatives can continue to improve our work unit's customer loyalty.
- 16. My manager works with me to ensure I understand the standards/goals on which my performance review will be based.
- 17. My manager is accessible for discussions.
- * 18. My manager and I have discussed the knowledge, skills, and abilities that could affect my progress at this organization.
- * 19. I have confidence in the fairness of my manager.
- * 20. My manager makes sure that I present my views on my performance reviews.
- 21. My manager promotes a positive atmosphere.
- 22. My manager makes sure that I get the recognition for my performance.
- 23. Our work unit has plans in place for resolving customer problems.
- 24. My manager helps me understand our work unit's service approach.
- 25. If I thought I needed to go out on a limb to deliver excellent service, I am confident my manager would support me.
- 26. My manager works with me to help resolve conflicts between work and family/personal issues.
- 27. My manager coaches me on how I can continuously improve the service I provide to my customers.
- 28. My manager has held a feedback session concerning last year's Leadership Survey with our work unit.
- 29. Over the next 12 months, which one of the above items (1-28) should your manager work on the most?

Note: Items marked with an asterisk are core items that were retained throughout all five years of the study.