Are the Characteristics of Narrative Comments Related to Improvement in Multirater Feedback Ratings Over Time?

James W. Smither
La Salle University

Alan G. Walker
Bowling Green State University

Researchers have paid almost no attention to the narrative comments that typically accompany multirater feedback reports despite the fact that both anecdotal and empirical evidence suggest that feedback recipients devote considerable attention to such comments. The authors examined improvement in upward feedback ratings over a 1-year period for 176 managers as a function of (a) the number of narrative comments each manager received, (b) whether those comments were favorable (vs. unfavorable), and (c) whether the comments were behavior/task focused (vs. trait focused). The authors found that managers who received a small number of unfavorable, behavior/task-focused comments improved more than did other managers, whereas managers who received a large number of unfavorable, behavior/task-focused comments declined more than did other managers.

Recently, there has been growing interest in research examining multirater feedback (e.g., Brett & Atwater, 2001; London & Smither, 1995; Mount & Scullen, 2001). To date, all empirical studies of multirater feedback have focused exclusively on the quantitative ratings provided by raters. There has been almost no attention to the narrative comments from raters that are nearly always included in multirater feedback reports. The absence of attention to narrative comments is striking, especially in view of the fact that both anecdotal and empirical evidence indicate that feedback recipients pay a great deal of attention to raters’ narrative comments (Antonioni, 1996; Kuhnert, 2000; Rose & Farrell, 2002). In an empirical study, Ferstl and Bruskiewicz (2000) found that feedback recipients reported that they attended to narrative comments more than they did to the quantitative ratings reported in their feedback reports. Consistent with this assertion, Dalessio (1998) and Van Velsor (1998) expressed concern that feedback recipients might focus too heavily on comments that do not adequately represent their evaluation as a whole.

The decision to provide multirater feedback, including narrative comments, presumably rests on the assumption that feedback will enhance the recipient’s subsequent performance. However, in their meta-analysis, Kluger and DeNisi (1996) found that although feedback sometimes enhances performance, in over one third of the studies they reviewed, feedback had a detrimental effect on performance. They noted that people can respond to gaps between their goals and feedback in several ways, such as changing their behavior, changing their goals, rejecting the feedback, or escaping the situation (e.g., avoiding the tasks or people that led to negative feedback).

Favorable Versus Unfavorable Comments

Researchers have consistently identified the sign of feedback (positive or negative) as an important characteristic. However, Kluger and DeNisi (1996) concluded that no theory of feedback can predict a priori the effects of all the important moderators that determine how feedback sign (positive or negative) affects performance. Control theory (Carver & Scheier, 1981, 1982) argues that when people detect discrepancies between their goals and their behavior, this has affective consequences that cause them to take action to reduce the discrepancies. In general, the greater the discrepancy, the stronger the subsequent response. From this viewpoint, unfavorable feedback should motivate the feedback recipient to take steps to reduce the discrepancy. However, unfavorable feedback may not necessarily lead to improvements in behavior. Indeed, feedback rejection seems to be a more likely response to negative feedback than to positive feedback (e.g., Ilgen, Fisher, & Taylor, 1979). In addition, there are several reasons to expect that a large amount of negative feedback (e.g., a large number of unfavorable comments) will have an especially negative effect on subsequent performance. For example, Kluger and DeNisi (1996, see also Mikulincer, 1988) noted that extremely negative feedback can lead recipients to abandon their goals (e.g., abandoning the goal of performing effectively in the view of those providing the feedback). Their meta-analysis found that discouraging feedback and feedback that threatens the recipient’s self-esteem decreased the effectiveness of feedback interventions. These findings suggest that a large amount of negative feedback could cause recipients to abandon or lower their goals or direct their attention elsewhere. Consistent with this view, several studies have found that feedback recipients are more likely to accept and use favorable feedback than unfavorable feedback (Brett & Atwater, 2001; C. L. Facteau, Facteau, Curtis, Russell, & Poteet, 1998; J. D. Facteau, Facteau, McGonigle, & Fredholm, 1997; Kay & Meyer, 1965; Ladner & Greguras, 2001; Stone & Stone, 1985).
Behavior/Task-Focused Versus Trait-Focused Comments

Narrative comments can focus the feedback recipient’s attention on specific behaviors or tasks, or they can draw attention to the feedback recipient’s traits or characteristics. Texts on developing management skills (e.g., Whetten & Cameron, 2002) argue that effective feedback is problem oriented rather than person oriented. In their feedback intervention theory, Kluger and DeNisi (1996) argued that feedback characteristics influence which of three hierarchically organized levels of control (task learning, task motivation, and self-related) will be attended to by feedback recipients. On the basis of their meta-analysis and review of the literature, they concluded that the effectiveness of feedback decreases as attention moves up the hierarchy closer to the self and away from the task. In sum, feedback that focuses on the task or specific behaviors should have a more positive effect on performance than feedback that focuses on the recipient’s broad traits, qualities, abilities, or other characteristics.

Number of Narrative Comments

It is possible that the attributes of comments (described above) may interact with the number of comments to affect the recipient’s subsequent performance. On the one hand, as noted above, a large (rather than small) number of negative comments could cause feedback recipients to abandon or lower their goals or direct their attention elsewhere. On the other hand, London and Smither (1995) argued that when a manager receives a large number of comments that are unfavorable, it may be difficult for the manager may find it difficult to dismiss the comments as reflecting merely the idiosyncratic views of one or two raters.

In sum, the purpose of this study was to examine how the characteristics of narrative comments (favorable vs. unfavorable, behavior/task-focused vs. trait-focused) and the number of comments might affect the behavior (as reflected in subsequent ratings) of feedback recipients. On the basis of the above, we expected that behavior/task-focused comments will be associated with positive changes in behavior. We also believed that the effect of the favorableness of comments on subsequent behavior may depend on the number of such comments. For example, a large number of negative comments may be associated with a decline in subsequent performance. A final purpose of this study was simply to examine the attributes of narrative comments provided to managers in a multirater feedback program. For example, is the favorableness of comments positively associated with the favorableness of ratings? Do comments tend to be behavior/task focused or trait focused?

Method

Survey Development

The upward feedback program involved in this study represents one component of the organization’s (a regional bank) overall leadership development effort. The program was developed in 1991 and although a copy of the target manager’s report has always been sent to the target manager’s supervisor (which in itself detracts from a purely developmental framework), the results are increasingly being used as one of several components in (a) identifying high-potential leaders, (b) making promotion decisions, and (c) making decisions concerning special assignment or lateral movements. As is the case with many organizational surveys, item changes occurred from year to year. However, from 1997 to 1998 (the period of study for this research) only 2 of the 29 items on the 1998 survey were modified slightly from the 1997 version. Responses ranged from 1 = strongly agree to 5 = strongly disagree. In our analyses, we reversed this scale so that higher numbers represented more effective behavior. All items were presented to raters as part of a single scale. For each year’s upward ratings, a principal axis factor analysis revealed a single underlying dimension (on the basis of examining the scree plot and the presence of only one eigenvalue greater than 1.0) that accounted for over 60% of the variance. The internal consistency (alpha) of the 29 items exceeded .95 in each year (N = 176). This single dimension has been interpreted as an overall leadership component. We therefore calculated a score for each manager in 1997 and in 1998 on the basis of the manager’s mean rating across items. In a previous study (Walker & Smither, 1999), we reported the median interrater agreement (r_{xy}) for the scale over a 5-year period to be .73. In addition, previous research showed that branch managers with more favorable upward feedback had more loyal customers (Walker & Smither, 1999).

Sample

To participate in the upward feedback program, the organization required that a manager have at least three direct reports. We examined the narrative comments provided to target managers in 1997 to determine the effects (if any) of these comments on target managers’ scores the following year (i.e., 1998). Given that the upward feedback program had been ongoing since 1991, we decided to control for managers’ previous experience with the program. Therefore, we specifically chose managers who were new to the program in 1997 and who also received 1998 results. This resulted in 79 managers being selected. To increase our sample size (and thus increase statistical power for our analyses) we also identified managers who were new to the process in 1996 and who also received 1997 and 1998 results. There were 97 such managers, thereby increasing our total sample size to 176 managers.

Of these 176 managers, 116 (66%) were women and 60 (34%) were men. Managers were located within 17 cities spanning three states. The majority of managers (86%) were concentrated within a single state in the Southeast United States. Managers represented a wide range of jobs involving most levels of the organization, including branch managers, credit officers, staff (e.g., human resources, finance), and back office operations. Managers’ salary grades ranged from 10 (senior manager) to 34 (first-line supervisor). Managers’ tenure in the organization ranged from 1.22 to 41.45 years with an average of 9.69 years of service with the organization. Although they had on average over 9 years of service in the organization, all of these managers had not previously participated in the upward feedback program. Individuals with considerable organizational tenure did not previously participate in the upward feedback program because (a) they had not previously been supervisors (i.e., they had only recently been promoted), or (b) they previously had less than three direct reports (but now had three or more direct reports).

The company’s annual performance appraisal process begins with managers and employees working together to establish areas of performance that will be evaluated over the next year. Different employees were therefore evaluated on different areas of performance. Each area of performance is weighted (on the basis of discussions with the employee’s manager). At the end of the year, the employee’s performance in each area is rated between 1 (highest rating) and 4 (lowest rating). These weighted scores are averaged to create a final appraisal rating between 1 and 4. Managers’ most recent annual appraisal ratings (prior to the administration of the 1998 survey) ranged from 1.10 to 2.30 (M = 1.72, SD = 0.25).

Procedure

Managers scheduled their employees for a group survey administration in which a human resources representative had employees complete an
opscan survey form and collected completed surveys. As a result of this procedure, nearly all direct reports completed upward feedback surveys concerning their managers. The number of direct reports (raters) ranged from 3 to 25 (M = 6.76, SD = 4.13). Managers received a feedback report that presents item means, standard deviations, range (minimum score and maximum score), and the number of valid responses. Each year managers who are new to the process are required to attend a 2-hr training session in which they receive instruction on program goals, interpreting their results, planning for and conducting feedback meetings with their raters, and developing an action plan. Included at the end of the survey form each year was a section where raters could handwrite comments to their supervisor. Raters were presented with the following instructions on the survey form: “Please print any comments that you would like to share concerning your manager’s leadership skills. All comments will be typed and then reported to your manager, word for word, as they appear in this section.” Raters were informed that “all written comments [would be] typed so that no one [would be] able to identify [their] handwriting.” Raters were also instructed “not to mention their supervisor by name” (so as to protect the manager’s confidentiality when the comments were being typed by the word-processing department).

As noted above, the purpose of this research was to investigate the relationship between the attributes of written comments and subsequent changes in the upward feedback scores received by managers. To do this, we first needed to identify the number of comments each manager received. As noted by Crano and Brewer (1986) and Macey (1996), the first step in coding written comments is determining the unit of analysis to be coded (e.g., a word, a phrase, a sentence, several sentences). To do so, we followed the advice of Crano and Brewer (1986) and Macey (1996) and identified the separate themes that were present in each rater’s comments. For example, a single sentence might contain more than one theme (e.g., “Overall a good manager, but needs to be more honest”). When this occurred, we considered the text as containing two (or more) comments (e.g., “Overall a good manager” and “needs to be more honest”). In some instances, several sentences reflected only a single theme (e.g., “I appreciate the way we are able to communicate. I talk, you listen, and vice versa. When I make suggestions, you never just shut them out, you help me to make the right one. Communication is very important.”). In such instances, we considered the text to be a single comment. In other instances, each sentence addressed a separate theme (e.g., “My manager is a great leader for whom I have the utmost respect. Although I feel that at times he needs to tell me what his feelings are about my job performance. He is very understanding about personal issues and family matters.”). In such instances, we treated each sentence as a separate comment. We used an iterative process in which one author would identify the number of comments (themes) provided by each rater (i.e., where a comment started and ended for our coding purposes). The other author would review these decisions and when necessary the authors would discuss a rater’s comments and reach consensus about the number of comments that were present. Following the procedure described above, 157 (88%) of the 176 target managers received 1,096 written comments in 1997. The number of raters providing comments per target manager ranged from 1 to 12 (M = 3.10, SD = 2.21). The number of comments per target manager ranged from 1 to 35 (M = 7.35, SD = 6.14).

We then developed a form to guide our ratings concerning each comment’s attributes. We rated the favorableness of each comment using a 1-to-5 scale in which 1 = unfavorable, 2 = slightly unfavorable, 3 = neutral, 4 = slightly favorable, 5 = favorable. We also rated the trait versus behavior/task focus of each comment using a 1 to 5 scale in which 1 = comment that focuses entirely on traits or characteristics of the manager, 2 = comment that focuses mostly on traits or characteristics of the manager, but to a lesser extent also refers to a specific behavior or task, 3 = comment that focuses equally on traits or characteristics of the manager and specific behaviors or tasks, 4 = comment that focuses mostly on specific behaviors or tasks but to a lesser extent also refers to traits or characteristics of the manager, and 5 = comment that focuses entirely on specific behaviors or tasks. Both authors independently rated the attributes of each comment for each target manager. At the time we completed these ratings, we were unaware of the manager’s upward feedback scores for 1997 or 1998. To illustrate the attributes of comments, in Table 1 we present four written comments and the attribute ratings for each comment. We calculated interrater reliability as the intraclass correlation (ICC) where ICC = (MSbetween − MSwithin)/MSbetween (Shrout & Fleiss, 1979). The interrater reliability estimates concerning the attributes of the comments were .93 for favorableness and .78 for behavior/task-focused. For each comment, we calculated the mean of the two coauthors’ ratings for each attribute. For each target manager, we then calculated the mean attribute rating across all comments.

We also asked two doctoral students to code a randomly selected subset of 95 comments. We calculated interrater reliability for these 95 comments using all four raters (i.e., the two doctoral students and the authors). The interrater reliability for favorableness was .95 (vs. .89 for the two authors alone). The interrater reliability for behavior/task-focused was .83 (vs. .78 for the two authors alone). Also, in no instance would removing one of the four raters from the preceding analyses have increased interrater reliability. These supplemental analyses indicate that interrater reliability was not an artifact of having the narrative comments coded by the two coauthors but could also be established (and to some extent enhanced) by using other raters who were not part of the research.

Results

We examined whether managers who first received feedback in 1996 differed from managers who first received feedback in 1997 in terms of their 1997 and 1998 upward feedback scores or in terms of the attributes (favorableness, trait vs. behavior/task fo-

Table 1
Examples of Attribute Ratings for Written Comments

<table>
<thead>
<tr>
<th>Comment</th>
<th>Favorable (5) vs. unfavorable (1)</th>
<th>Focus on behaviors/tasks (5) vs. traits (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>My manager is a great leader for whom I have the utmost respect</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>I would like her to delegate more responsibility to me</td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>Needs to be more honest</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>You try your best to keep us informed and that we get the same attention as anyone else in the department. Ex.: New equipment, software upgrades, computer training</td>
<td>4.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Note. Table entries are the mean rating of each comment’s attributes provided by the two authors.
These two groups for all future analyses. Manager’s characteristics (gender [coded 1 = female and 2 = male] and organizational tenure) were unrelated to the attributes of the narrative comments they received.

Correlations among the key variables are presented in Table 2. Consistent with previous research (e.g., Johnson & Ferstl, 1999; Reilly, Smither, & Vasilopoulos, 1996; Walker & Smither, 1999), 1997 and 1998 upward feedback scores were positively correlated ($r = .59$). Narrative comments tended to be favorable ($M = 3.88$ on the 5-point scale) and the favorableness of comments was positively related to upward feedback scores in 1997 ($r = .67$) and 1998 ($r = .50$) but unrelated to the behavior/task focus of the comments. The mean behavior/task focus of the comments (2.83) was near the center of the 5-point scale (1 = trait focused, 5 = behavior/task focused). The number of comments was unrelated to upward feedback scores and the attributes of the comments, but not surprisingly it was positively related ($r = .52$) to the number of direct reports (raters). Raters who received more favorable comments from their direct reports also received a more favorable annual review from their supervisors ($r = -.23$, $p < .05$; the correlation is negative because more favorable annual reviews are associated with lower numbers).

To examine the research question, we conducted a hierarchical regression analysis with 1998 upward feedback scores as the criterion. On the first step we entered 1997 upward feedback scores (Cronbach & Furby, 1970). On the second step we entered the attributes of the comments (favorableness, behavior/task-focus, and number of comments). On the third step we examined two-way interactions between the attributes of comments by entering the relevant cross-product terms. On the fourth step, we entered a cross-product term representing the three-way interaction among favorableness, behavior/task-focus, and number of comments. Results are presented in Table 3.

It can be seen that the main effect for favorableness of comments reached significance ($p = .05$). The positive beta coefficient indicates that feedback recipients tended to improve more after receiving favorable (rather than unfavorable) comments. No other main effect or two-way interaction reached statistical significance.

The three-way interaction was significant (increase in $R^2 = .02$, $t(144) = 2.13$, $p = .04$). To describe this interaction, we first computed a new score (hereafter referred to as change) for each manager that represented a measure of change in upward feedback ratings from 1997 to 1998 (i.e., the manager’s actual 1998 upward feedback score minus the manager’s predicted 1998 upward feedback score based on the manager’s 1997 upward feedback score). We then split managers into two groups (high vs. low) on each of the three predictors using a median split and examined the mean change score for each of the resulting eight cells. The results are presented in Table 4.

It can be seen that the joint effect of the number of comments and the behavior/task focus of comments on subsequent ratings depended on whether the comments were favorable or unfavorable. When managers received a small number of unfavorable, behavior/task-focused comments, their subsequent performance improved more than that of managers in other conditions. However, when managers received a large number of unfavorable, behavior/task-focused comments, their subsequent performance declined more than that of managers in other conditions. The effect size (standardized mean difference) associated with the difference between receiving a small versus a large number of unfavorable, behavior/task-focused comments was large ($d = 1.20$; according to Cohen, 1988, small, medium, and large values of $d$ are about .20, .50, and .80, respectively).

We conducted several additional analyses to determine whether the three-way interaction described above was an artifact of the number of comments being moderately correlated ($r = .52$) with the number of direct reports. We also explored whether number of comments might be related to other variables (e.g., the manager’s salary grade) in which case the results associated with number of comments might be an artifact of such a relationship. First, number

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 1997 upward feedback score</td>
<td>4.06</td>
<td>.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>2. 1998 upward feedback score</td>
<td>4.08</td>
<td>.45</td>
<td>.59*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>3. Favorableness of comments</td>
<td>3.88</td>
<td>.99</td>
<td>.67**</td>
<td>.50**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>4. Behavior/task vs. trait focus of comments 1997</td>
<td>2.83</td>
<td>.79</td>
<td>-.15</td>
<td>-.13</td>
<td>-.11</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>5. No. of comments 1997</td>
<td>7.35</td>
<td>6.14</td>
<td>-.08</td>
<td>-.10</td>
<td>-.13</td>
<td>.14</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>6. No. of direct reports (raters) 1997</td>
<td>6.76</td>
<td>4.13</td>
<td>-.17*</td>
<td>-.16*</td>
<td>-.13</td>
<td>-.01</td>
<td>.52**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>7. Ratio of comments to direct reports 1997</td>
<td>1.09</td>
<td>.73</td>
<td>.08</td>
<td>.03</td>
<td>-.03</td>
<td>.18*</td>
<td>.72**</td>
<td>-.08</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>8. Most recent annual review rating (pre-1998 upward feedback)</td>
<td>1.72</td>
<td>.25</td>
<td>-.15</td>
<td>-.08</td>
<td>-.23**</td>
<td>-.02</td>
<td>.06</td>
<td>.00</td>
<td>.03</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>9. Organizational tenure (years)</td>
<td>9.69</td>
<td>7.99</td>
<td>.01</td>
<td>.00</td>
<td>.07</td>
<td>.00</td>
<td>-.07</td>
<td>.01</td>
<td>-.09</td>
<td>-.13</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>10. Gender (1 = female, 2 = male)</td>
<td>1.34</td>
<td>.48</td>
<td>-.08</td>
<td>-.23**</td>
<td>-.04</td>
<td>.11</td>
<td>.07</td>
<td>.03</td>
<td>-.04</td>
<td>-.16*</td>
<td>-.27*</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. Cell ns ranged from 131 to 176. *$p < .05$. **$p < .01$. 

Table 2

Intercorrelations Among Key Variables
of comments and number of direct reports were unrelated to the manager’s salary grade, gender, organizational tenure, or position tenure, thereby indicating that number of comments does not appear to be a surrogate for the manager’s organizational status. Second, we repeated the hierarchical regression analysis described above, but this time we also included number of direct reports (along with 1997 upward feedback scores) on the first step of the analysis. The results were virtually identical to those described above. Finally, we examined the ratio of comments to direct reports. We created a new variable by dividing the number of comments by the number of direct reports (hereafter referred to as Comments/DR) and we conducted a hierarchical regression analysis like the one first described above except that we included Comments/DR instead of number of comments. The results closely paralleled the results presented in Tables 3 and 4.\(^1\) In sum, these analyses found nothing that would change the conclusions drawn from Tables 3 and 4.

Table 4
Three-Way Interaction Among Favorableness of Comments, Behavior/Task Focus of Comments, and Number of Comments

<table>
<thead>
<tr>
<th>No. of comments</th>
<th>Favorable comments</th>
<th>Unfavorable comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trait-focused</td>
<td>Behavior/task-focused</td>
</tr>
<tr>
<td>High M change</td>
<td>-.001</td>
<td>.048</td>
</tr>
<tr>
<td>SD</td>
<td>.29</td>
<td>.40</td>
</tr>
<tr>
<td>n</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>Low M change</td>
<td>.107</td>
<td>-.020</td>
</tr>
<tr>
<td>SD</td>
<td>.31</td>
<td>.37</td>
</tr>
<tr>
<td>n</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>d</td>
<td>-.036</td>
<td>.17</td>
</tr>
</tbody>
</table>

Note. Mean change = the manager’s actual 1998 upward feedback score minus the manager’s predicted 1998 upward feedback score based on the manager’s 1997 upward feedback score. Positive scores reflect improvement, whereas negative scores reflect a decline in performance. \(d = \) the mean standardized difference between managers who received a high number of comments versus managers who received a low number of comments.

Discussion

We found that narrative comments, like ratings, were generally favorable. Seventy percent of the comments were coded as 3.5 or higher (1 = unfavorable, 5 = favorable). It was also encouraging to find that the favorableness of comments was positively correlated with the favorableness of ratings. The narrative comments in this study represented the entire spectrum from trait focused to behavior/task focused; 38% of the comments were coded 2 or less and 37% were coded 4 or higher (1 = trait-focused, 5 = behavior/task focused). It is noteworthy that raters were simply told, “Please print any comments that you would like to share concerning your leadership skills.” It would be helpful for future research to examine whether more precise instructions (e.g., urging raters to describe specific behaviors or to offer concrete suggestions for improvement) would affect the characteristics of narrative comments and thereby affect subsequent performance.

The attributes of comments and the number of comments explained about 4% of the improvement in upward feedback scores (see Steps 2–4 of the hierarchical regression analysis). Managers who received favorable comments tended to improve more than managers who received unfavorable comments. Contrary to our expectations, behavior/task-focused comments did not lead to greater improvement than did trait-focused comments. We also observed a three-way interaction. Managers who received a large number of unfavorable, behavior/task-focused comments declined more than did other managers. This finding is consistent with Kluger and DeNisi’s (1996) finding that extremely negative and discouraging feedback can decrease the effectiveness of feedback interventions. Perhaps feedback recipients who receive a large number of unfavorable, behavior/task-focused comments become discouraged or overwhelmed and conclude that successful behavior change is not feasible. In contrast, managers who received a small number of unfavorable, behavior/task-focused comments improved more than did other managers. Perhaps a small number of negative but behavior/task-focused comments point to specific areas in which feedback recipients can change their behavior without discouraging the feedback recipient from thinking that

\(^1\) Detailed results are available from James W. Smither.
such changes are feasible. Finally, we remind the reader that although the direction of this three-way interaction is interesting, it explained only 2% of the variance in improvement.

There are several areas in which additional research about narrative comments would seem especially desirable. First, it is important that future research examine whether the direction of the effects observed in the current study are replicated in other settings. Second, other characteristics of raters’ comments may also be important (e.g., whether the rater’s intentions appear to be helpful vs. mean spirited). Third, it would also be helpful to determine whether the characteristics of narrative comments are related to the likelihood that managers will remember the comments or discuss their feedback with coworkers (e.g., to clarify the feedback or seek specific suggestions for improvement; see Walker & Smither, 1999). Fourth, if researchers can answer some of the questions posed above, it would be desirable to determine whether raters can be taught to provide narrative comments that are memorable, motivating, and helpful in guiding behavior change.

The limitations of this study should be considered when interpreting our results. For example, our sample consisted of 176 managers who were receiving upward feedback for the first or second time in this organization. We do not know the effect of narrative comments on managers who receive multirater feedback repeatedly over many years. Also, the comments in this study were provided by the manager’s direct reports. It would be desirable to understand whether managers are more or less affected by comments provided by their supervisor or peers. It would also be useful to identify additional criteria (other than subsequent upward feedback ratings) that could help us understand the impact of comments on the performance of feedback recipients. Also, the generalizability of our results may be limited by the fact that upward feedback in this organization could be used to influence administrative decisions. Research suggests that raters may distort their ratings when feedback is gathered for administrative rather than for developmental purposes (Waldman, Atwater, & Antonioni, 1998). Finally, although we obtained interrater reliability estimates concerning the attributes of comments (i.e., favorableness and behavior/task focus), it would be desirable for future researchers to obtain reliability estimates concerning the number of comments (as described above, we derived the number of comments using an iterative process between the two authors). Also, we encourage researchers to monitor developments in computer-based text analysis (e.g., Mosholder, Settoon, Armenakis, & Harris, 2000; Pennbaker & Graybeal, 2001; Perrin, 2001). These developments may allow future researchers to more easily analyze narrative comments in a systematic manner.

We hope that researchers will increasingly turn their attention to the role of narrative comments in multirater feedback programs. To do so, researchers will need to grapple with the challenges of coding narrative comments. As we noted above, this is not always a straightforward process (e.g., sometimes a single sentence contains more than one theme whereas sometimes several sentences contain only a single theme). Moreover, researchers who study narrative comments in multirater feedback are likely to benefit from understanding the methods that have been used to study comments in other types of organizational surveys (Macey, 1996) as well as the tools and challenges associated with content analysis of qualitative data in social research (e.g., Crano & Brewer, 1986).

Finally, because narrative comments are likely to share some similarities with verbal, face-to-face feedback (as well as with written feedback often contained in formal performance appraisal processes), research in this area may enhance our understanding of how people can provide more effective feedback in a variety of situations.

References


Received November 20, 2002
Revision received June 6, 2003
Accepted June 23, 2003

---

**Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted**

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- **To be selected as a reviewer, you must have published articles in peer-reviewed journals.** The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.

- **To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review.** Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.

- **To select the appropriate reviewers for each manuscript, the editor needs detailed information.** Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.

- **Reviewing a manuscript takes time (1–4 hours per manuscript reviewed).** If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.