

Dissertations Title: THE EFFECTS OF A PEER FEEDBACK INSTRUMENT ON TEAM MEMBER BEHAVIOR

Author(s): DOMINICK, PETER GERARD

Degree: PH.D. **Year:** 1998

Pages: 00084

Institution: STEVENS INSTITUTE OF TECHNOLOGY; 0733

Advisor: Adviser: RICHARD R. REILLY

Source: DAI, 59, no. 08B, (1998): 4518

Abstract: This study examined the effects of a peer feedback instrument on subsequent behavior using a four-dimensional model of team behavior. Previous research on peer and upward feedback had suggested that exposure to key behaviors is an important variable in behavioral improvement. This study sought to extend that line of research by identifying the aspect of exposure most likely to lead to improvement. Participants were randomly assigned to teams and teams were randomly assigned to one of five experimental conditions: feedback, anticipation of feedback, rating only, exposure only and control. In the feedback condition, participants rated themselves and each other using a 24-item behavioral observation scale (BOS) after completing the first of two decision-making tasks. Before performing the second task, they received individualized feedback reports summarizing their self- and peer ratings. Those assigned to the anticipation condition, completed the BOS expecting to receive feedback although they did not actually receive feedback until after completing the second task. Subjects in the rating condition, completed the BOS but did not anticipate or receive feedback. Those assigned to the exposure only condition used a modified version of the BOS to rate just their own behavior. Subjects in the control condition did not complete the BOS at all. All subjects were videotaped while working on both the first and second tasks and the tapes were rated by experts blinded to experimental condition. The results indicated that simply being exposed to key behaviors lead to improvement (behavior change). In addition, subjects in the feedback condition were more likely to feel that others on their team expected them to improve their behavior.

Introduction

Numerous authors have stressed the pivotal role that autonomous or semi-autonomous teams play in the success and effectiveness of modern firms (e.g. Katzenbach & Smith, 1993; Peters, 1988; Reich, 1987). Increasingly, teams have become integral parts of organizations' structures. For example, by 1990, 47% of Fortune 1000 companies reported that they used work teams, compared with 28% just three years earlier in 1987 (Lawler, Mohrman & Ledford, 1992).

Regardless of the setting, the organization of workers into teams implies an increasing emphasis on self-management by both individuals and the team as a unit (Hackman, 1987). Moreover, successful team outcomes depend heavily upon effective interaction among team members. Virtually all models of team and work group effectiveness acknowledge that interpersonal process is a crucial antecedent of team performance (Gladstein, 1984; Hackman, 1983; Nieva Fleishman & Reick, 1978; Stevens & Campion, 1994, Sundstrom, DeMeuse & Futrell, 1990). Thus, the effective transition to a team-based organization should be facilitated by organizational interventions designed to promote and reinforce the competencies and behaviors necessary for effective teamwork.

One such type of intervention is a survey-based behavioral feedback program. Survey-based feedback programs have long been a part of organization development programs and initiatives (e.g. Nadler, 1977). Such programs make it possible to summarize the opinions and perspectives of many organizational constituents quickly, clearly and confidentially.

Recent years have seen the proliferation of multi-source feedback initiatives (London & Smither, 1995). For example, Romano (1994) reported that companies spent \$152 million dollars on multi-source rating development in 1992. These programs involve using feedback surveys to collect others' perspectives on an individual's behavior and or performance. The sources of feedback typically include one or more of the following: peers, subordinates (upward), customers, self, or managers. Multi-source feedback programs are based on the assumption that there is value in providing employees with views of their performance as seen from a variety of sources, not all of which may have the same perspective or be in the same position to observe behavior. For instance, peers may have a better perspective on someone's teamwork skills, while subordinates might be in a unique position to evaluate someone's supervisory ability (London & Smither, 1995).

Studies of upward feedback have shown significant changes in behavior after the implementation of behaviorally-based feedback programs (e.g., Atwater, Roush & Fischthal, 1995; Smither, London, Vasilopoulos, Reilly, Millsap, & Salvemini, 1995; Reilly, Smither & Vasilopoulos, 1996). On the other hand, there have been few empirical investigations of the behavioral impact of peer feedback in team settings¹ (Dominick, Reilly, & McGourty, 1997). The primary purpose of this study was to examine the effects of behaviorally-based peer feedback on subsequent behavior in a team-based task. Because earlier quasi-experimental research on upward feedback has suggested that mere exposure to the desired behaviors may be a critical factor (e.g. Reilly,

¹ An exception is the study by Dominick, Reilly & McGourty (1997).

Smither & Vasilopoulos, 1996) this study examined the effects of exposure in addition to examining the effects of behavioral feedback in a team context.

LITERATURE REVIEW

Interventions in Group Process

The use of teams and the notion of teamwork have increasingly become the norm within many organizations. This is particularly true for organizations that undertake work such as new product development, concurrent engineering or systems development (Mohrman, Cohen and Mohrman, Jr., 1995). All of these activities share in common the need to integrate and coordinate diverse technical skills and expertise in a quick and cost-efficient manner. In addition to individuals' technical abilities, a key determinant of a team's performance is the quality of its process for working together (Kernaghan & Cooke, 1986). Specifically, this includes the attitudes of team members towards collaborating and their shared perceptions of what constitutes appropriate team member behavior (Kernaghan and Cooke, 1990). Bales (1950) was among the first researchers to describe taxonomy of individuals' behavior in group settings. He identified two broad behavioral categories that relate to group performance: interpersonal behaviors and task-related behaviors. Positive interpersonal behaviors include active listening, constructive confrontation and supporting and building upon others' input. Task-related behaviors include setting objectives, clarifying roles, identifying obstacles and finalizing decisions.

More recently, other researchers have put forth similar behavioral paradigms specifically with regard to work in teams. Stevens and Campion (1994) described 14 distinct competencies made more salient by the distinctive nature of work in team situations. As did Bales, they grouped these competencies into two broad categories: Interpersonal and Self-Management competencies. Subcategories relating to the

Interpersonal area include conflict resolution, collaborative problem solving and communication. The two subcategories for Self-Management are Goal Setting and Performance Management; and Planning and Task Coordination.

Gaddy and Wachtel (1992) also suggested that working in teams requires a distinct set of behavioral competencies. They refer to these competencies as generic skills. Generic skills are those that have broad applicability across various types of teams and tasks (e.g. communication, conflict management, planning). Generic skills stand in contrast to operational skills or taskwork behaviors, those related specifically to the tasks performed by individual team members. Examples of operational skills include team members interactions with equipment and the learning of technical aspects of their jobs.

Given the importance of individuals' team skills and behavior to successful group outcomes, management specialists have developed a number of process interventions over the years that help to establish and reinforce positive team behaviors (Woodman & Sherwood, 1980; Burke & Day, 1986). These process interventions are generally one of three types: diagnosis and feedback, process consultation, and task redesign (Hackman & Morris, 1975). The intervention used in this study is a type of diagnosis and feedback.

In a previous laboratory study independent observers analyzed the interpersonal and task-related behaviors of groups working on a planning task (Kernaghan and Cooke, 1990). In the middle of their work, an observer intervened and provided process feedback to the group. They found feedback on task-related behaviors had the greatest impact on group performance. These results demonstrate the potential value of feedback on group process to team members. It should be noted, however, that the basis for the feedback

were observations by independent observers as opposed to the team members themselves. In addition, the key dependent variable in this study was task performance. The study did not report whether there were changes/improvements in individual team members' behavior as the result of the feedback.

Employee Feedback Programs

The self-managing context of many teams suggests that team members themselves can play an important role in enhancing and sustaining team effectiveness by providing feedback to each other. Mechanisms which assist team members in such assessment and feedback can be crucial from both an evaluative and developmental perspective (Saavedra & Kwun, 1994; Murphy & Cleveland, 1991). Ideally, behavioral evaluation and feedback from other team members should be an important developmental tool fostering more positive team behaviors from members. In fact, several organizations have reported successfully utilizing peer feedback in team settings as the basis for both development and evaluation (e.g. Ramsay & Letho, 1994; Zigon, 1994).

Several studies have examined the effects of feedback on subsequent behavior. Hegarty (1974) found that upward feedback leads to subordinates perceiving positive changes in the boss's subsequent behavior. However, this study only included 28 managers and the same employees were questioned about their managers' performance before and after feedback. In another study, Bernardin, Hagan and Kane (1995) found improvement in subordinate and peer ratings but no changes in supervisor or customer ratings after managers received 360 feedback.

Hazucha, Gentile, and Schneider (1993) reported increases in the skill levels of managers two years after they received 360 degree feedback. In this study, 198 managers received 360 feedback based upon a popular 360 instrument, the Management Skills Profile (MSP). Of the original 198 managers who received feedback, 48 participated in the same feedback process two years later. It was the MSP ratings of these managers at Time-one and Time-two that were compared. They report improvement effect sizes of .29 for others' ratings of managers and .40 according to self ratings. They also found that self-other ratings were more consistent at Time-two. One important limitation to this study is that the absence of a control condition makes it difficult to determine the cause of the improvement. In addition, there is a problem with substantial subject loss (from 198 to 48). The researchers did however, test to see if the smaller sample was representative of the larger group. They found no significant differences in the Time-one ratings of the sample and the larger group. On the other hand, it is still possible that those who participated at Time-two were managers who were more motivated to improve.

Atwater, Roush & Fischthal (1995) found that follower ratings of student leaders improved after feedback was given to leaders and that leaders receiving "negative" feedback (defined as self-ratings that were considerably higher than follower ratings) improved the most. Subjects in this study were 978 leaders (juniors) at the U.S. Naval Academy and 1,232 followers (freshman). The Leaders received confidential upward feedback from the freshman followers they had in their squads. The leaders' feedback was aggregated information based upon two rounds of followers' ratings taken six weeks apart. The dependent measure was a post feedback rating provided some eighteen weeks

after the second round of feedback. These post feedback ratings were actually provided by a different group of followers than those who had provided the original feedback. In addition to finding improvements in performance based upon follower ratings, they also found that self-ratings became more consistent with follower ratings after receiving feedback. One of the limitations to this study is that there was no control group. Another possible limitation is that the unique military setting in which the study was conducted may limit its generalizability to civilian settings.

Johnson and Ferstl (1997) reported on findings from two waves of an upward feedback process in a large accounting firm. The study included 2,171 managers who received individualized feedback reports. Their investigation focused on whether or not the level of self-other agreement impacted improvement from the first wave to the second. Specifically, researchers found that over-raters at all levels of initial performance tended to improve (in terms of both self and subordinate ratings), not just over-raters at a low level of initial performance. On the other hand, managers whose self-ratings were about the same as their subordinates ratings of them tended to have no changes from the first wave to the second. In addition, managers who underrated themselves at Time 1 tended to have higher self-ratings and lower subordinate ratings at Time 2. Because the large sample size ensured good power, their findings provide strong evidence for the impact of self-other discrepancies on people's reactions to multi-source feedback.

Another interesting finding from their study was that improvement for managers who did not complete self-ratings was related to initial levels of performance. In other words, among managers without self-ratings, those whose initial performance was low were

more likely to improve and the performance of those whose initial level was high tended to decline. The researchers suggest that even though these managers did not complete self-ratings they may still have made implicit self-evaluations when reviewing their feedback or perhaps when completing the feedback instrument to rate others. Although it will be explained in more detail later, it is worth noting that one of the questions asked in the present study is a reverse of Johnston & Frestl's finding. Whereas they found that people can improve without doing a written self-evaluation, the present study asked if people can improve their performance after *only completing* a self-evaluation and *not receiving* feedback based upon others evaluations.

Limitations to the Johnston and Ferstl (1997) study include the fact that there was not a control group. They also point out that all managers were required to share their feedback with their immediate supervisors. This requirement may have impacted their subsequent self-ratings and may also have helped to promote improvement in ways that the feedback alone would not have.

Walker (1997) reported on the effects of an upward feedback program over five years for 252 managers in a large regional bank. All managers with three or more direct reports received feedback annually based upon ratings from their subordinates. He reported that with the exception of the first two years ('91 and '92), manager's ratings by their subordinates incrementally changed in a positive direction over time. Specifically, post-hoc analyses revealed a significant difference between managers' ratings during the first two years of the program and their ratings during the last two years ('94 and '95). He also found (after controlling for regression to the mean) that managers' who initially

scored the lowest were the ones who improved the most over the five year period.

Another interesting finding from his study was that managers who conducted formal post-feedback discussions with employees were more likely to improve than those who did not conduct these kinds of meetings. As was the case for some managers in Johnston and Frestl's (1997) study, it does not appear that any managers in Walker's study completed self-ratings. They may however, have made implicit assessments of their own behavior and thereby identified discrepancies between their own and subordinate ratings.

In an earlier study, Smither, London, Vasilopoulos, Reilly, Millsap and Salvemini (1995) reported that self-other discrepancies impacted subsequent improvement. These researchers examined the effects of an upward feedback program in which 238 managers in a Fortune 100 company received feedback at two points in time roughly six months apart. They found a significant but small improvement in subordinate ratings across the 238 managers participating in the program. Using time as an independent variable, these researchers detected a small ($d=.16$) but significant effect size, indicating that managers' performance had improved from the first wave of feedback to the second. The researchers also found that improvement was greatest for managers who initially received the most negative upward feedback (e.g., lowest ratings) and also rated themselves higher than their subordinates had. Their analysis did take into account regression to the mean.

In their study, Smither et al. (1995) had a subgroup of managers who completed the feedback instrument but did not receive individualized feedback ($n=54$) - because they had ratings from fewer than three subordinates and providing feedback would have violated the raters' confidentiality. They reported that these managers were just as likely

to improve their performance as were managers who received individualized feedback. On the other hand, as a quasi-experimental study conducted in a large organization, it is possible that managers who did not receive individualized feedback reports may have gotten important feedback cues from other sources in the organization.

Reilly, Smither, and Vasilopoulos (1996) tracked 92 managers from the original group in the Smither et al. (1995) study through two additional administrations of the feedback survey. This represented a total time period of 2.5 years and four survey administrations. Using ratings from the fourth administration of the survey as the dependent variable, this follow-up study found that most of the performance improvements occurred between the first and second administrations of the survey but that these initial improvements were sustained over time.

Another aspect of their study shed further light on the role of feedback relative to exposure to the feedback instrument. Among the 92 managers in this study, 31 received feedback during the first three administrations, 27 received feedback two out of three times, 18 received feedback one of the first three times and 16 did not receive feedback during any of the administrations of the survey. The researchers used hierarchical regression to determine whether or not improvement was related to the number of times managers received feedback. Their analysis found no such relationship. The authors do point however, that this aspect of their study had lower power than what may have been needed to detect a difference. They also indicated that managers who had not received any feedback (n=16) were initially rated higher than managers who received feedback at least once. Consequently they may have had less room to improve.

Nonetheless, this study (Reilly et al., 1996) when combined with the findings from Smither et al. (1995) suggests that feedback itself may not be the critical variable in producing change. In these studies, improvement for individuals who did not receive feedback but were exposed to the feedback instrument (by completing self-ratings, and ratings for their boss) was as great as those who actually got feedback reports.

Along similar lines, Smither, Wohlers, and London (1995) used control theory (Carver & Scheier, 1981) and goal setting theory (Locke & Latham, 1990) as the rationale for hypothesizing that team leaders who received individualized upward feedback would view the feedback as more useful when compared to the reactions of team leaders who only received normative upward feedback (average of team leaders ratings). In this study, subjects were team leaders responsible for coordinating recruitment efforts for a large telecommunications corporation. One group of 42 team leaders received both individualized feedback and normative feedback. A second group of 41 team leaders only received normative feedback. Six months after receiving the feedback, the team leaders were surveyed about their reactions to the feedback process.

The researchers found that team leaders who had received individualized feedback were likely to see the feedback as useful, were more likely to discuss their feedback with others and were also more satisfied with the feedback process. On the other hand, they also found that the expressed intentions of team leaders to change their behavior were the same regardless of whether the leaders received

individualized upward feedback or only normative feedback. It is important to note however, that there were some differences in the backgrounds of subjects in the two groups. The researchers report that recruiting was more likely to be the full-time job of those who received individualized feedback. Subjects in this condition also tended to be more experienced. These facts may be confounds that also help to explain why subjects in the individualized feedback condition had more positive reactions. In any event, the two groups were similar when it came to intentions to change behavior and this fact is of more immediate relevance to the present study (which focuses on behavior change). While their findings with regard to intentions are interesting, they did not investigate any actual differences in post feedback improvement between the two groups.

In summary, the studies discussed above suggest that multi-source feedback systems can lead to changes in behavior and in attitudes. In addition, most of the studies also found that improvement tends to be greatest for managers whose initial ratings are low. Several of these studies also suggest that exposure to the feedback instrument plays an important role in leading to behavior change. However, because of their quasi-experimental nature, it is difficult to determine whether exposure or feedback is the critical variable in producing change. In addition, in virtually all of these studies measurements of behavioral change were obtained directly from study participants.

Explanations for the Effects of Feedback Programs

Both Control Theory (Carver & Scheier, 1981) and Goal Setting Theory (Locke & Latham, 1990) provide a useful framework for explaining how feedback programs effect the behavior of participants. First, both theories view behavior as goal directed. Second, both theories suggest that in order to achieve goals people use feedback to evaluate their performance relative to their goals (Kluger & DeNisi, 1996). The two theories share in common several other key activities, namely: self-monitoring of behavior; followed by self-evaluation of behavior by comparison to a standard; and then self-reaction in the form of satisfaction or dissatisfaction, which in turn leads to further adjustment of behavior and or the modification of goals (Kanfer, 1990). In essence, the theories describe self-regulatory loops in which feedback becomes the basis for evaluation and subsequent adjustment (see figure xx).

For example, from the perspectives of Control theory (Carver & Scheier, 1981) receipt of feedback should be critical. Control theory suggests that specific feedback is the basis for identifying goal-feedback discrepancies, which in turn direct one's attention towards change and improvement. According to Carver & Scheier (1981) two things must happen in order for people to change their behavior. First, they must focus attention on their behavior so that they can obtain input for comparisons purposes. Second, they must self-regulate their behavior based upon whether or not they perceive a discrepancy between their observed behavior and their standard/goal.

A peer based feedback program should provide the opportunity for both prerequisites to occur. Attention is directed toward one's own behavior at both the time a person completes the feedback instrument and at the time one receives feedback. The receipt of feedback also satisfies the second prerequisite by providing the opportunity for individuals to identify discrepancies between how they perceive their behavior and how constituent others perceive it.

Goal Setting Theory (Locke & Latham, 1990) predicts similar outcomes, however, from this perspective, people are not motivated by the need to reduce the discrepancy but rather by the desire to achieve the goal (DeNisi & Kluger, 1996). In goal setting language individuals either work to achieve the goal, change the goal, reject the feedback or abandon commitment to the goal. However, they are most likely to continue working towards goal achievement when their initial performance is below the goal standard (Locke & Latham, 1990). As is the case with Control Theory, receipt of feedback becomes valuable because it is the means by which people know whether or not their performance is above or below the standard.

It is clear that feedback plays a valuable role in the self-regulating mechanisms associated with both theories. On the other hand, given the previously discussed findings of Smither, et al. (1995) and Reilly, et al. (1996) an important question is whether or not these self-regulating mechanisms can be activated without actually receiving feedback.

Within the context of a peer feedback program there are several reasons to believe that behavioral improvements should occur simply through exposure to and completion of the feedback instrument. Research on frame of reference training indicates that information impacts raters' schemas of effective behavior (e.g. Cardy & Keefe, 1994).

This perspective has been applied by many researchers to the study of performance appraisal (e.g. DeNisi, Cafferty & Meglino, 1984; Feldman, 1981). Information about a behavioral construct (e.g. leadership) can impact a person's prototype or schema for that construct. That schema then helps to shape how people will, on the one hand, evaluate others' leadership behavior, and on the other hand can also act as a script that guides their own behavior (London & Smither, 1995).

Similarly, just introducing a feedback instrument provides people with referencing information that may influence their notions of what constitutes effectiveness (London & Smither, 1995). As Van Velsor & Leslie (1991) note, a feedback instrument consisting of behavioral items is by nature, prescriptive and provides people with a learning opportunity. By reviewing the items, individuals are exposed to examples of effective behavior and are given a framework for understanding/self-evaluating their behavior in a particular context (e.g. teamwork).

Second, Locke and Latham (1990) have argued that the just implementing a formal feedback system into a work group may sometimes cause spontaneous goal

setting. In other words, introducing the system send a strong message that performance should be improved in those areas that are being measured. For example, understanding that effective decision-making includes “anticipating problems and developing contingency plans,” may increase individuals’ efforts to do so in the future. Third, the process of completing self and peer-ratings provides individuals with an opportunity to reflect upon their own behavior and to establish normative standards and personal improvement objectives. Consistent with Control Theory, individuals can then use these standards and objectives as a basis for self regulation. Specifically, individuals who perceive a discrepancy between their behavior and the standard behavior presented, might be most motivated to change their behavior (Atwater, et al., 1995; Smither, et al., 1995, Reilly, et. al, 1996).

In a recent laboratory study, Dominick, Reilly and McGourty (1997) examined the behavioral impact of feedback relative to exposure to the feedback instrument. Instead of focusing on upward feedback however, their study examined peer feedback. Subjects were randomly assigned to teams with four or five members, and teams were randomly assigned to one of three experimental conditions: feedback, exposure, or control. Each team performed two group decision making tasks. In the feedback condition, team members rated themselves and each other using a 24-item behavioral observation scale after completing the first task. The scale was based upon a four dimensional model of team behavior (McGourty, DeMeuse & Dominick, 1994). Prior to performing the second task (one week later), they received individualized feedback reports that summarized how

they rated themselves and how their peers had rated them. In the exposure condition, team members completed the same 24-item scale after completing the first task but did not receive individualized feedback until after completion of the second task. Team members in the control condition did not complete the instrument or receive a feedback report. The second task was videotaped and rated by experts blind to experimental condition. Results showed significantly higher means on all four dimensions for participants in the feedback and exposure conditions compared to the control group; there were no differences between subjects in the feedback and exposure conditions.

The results of this study were consistent with the quasi experimental results reported by Smither et al. (1995) and Reilly et al. (1996) for upward feedback. At the same time, these results provide a more definitive picture of feedback's role relative to exposure. Rather than relying on participants' perceptions of behavioral change, this study measured behavioral differences using external raters blind to condition, thereby showing that it is actual behavior that is different (not merely perceptions of program participants). In addition, unlike the previous studies, this study included a sample not exposed to the feedback instrument and as a result more clearly identifies exposure as the critical mechanism in creating behavior change.

On the other hand, there are a few theoretical and methodological limitations to the Dominick, et al. study. First, the study only evaluated subjects' behavior after the second task. Although one can assume, on the basis of random assignment, that subjects' behavior during the first task was equal across conditions, this cannot be stated unequivocally. In addition, one could argue that the observed differences were the result

of a decline in the effectiveness of the control condition rather than improvements amongst those in the exposure and feedback conditions. These questions would have been more clearly addressed if subjects' behavior after the first task were also evaluated and then used as a covariate in the final analysis of results.

Although an exposure condition was included in the experiment, there are really several elements of exposure that need to be disentangled: 1) exposure to the items themselves; 2) rating of self and others' behavior, thereby requiring some depth of processing concerning the items; and 3) anticipation of feedback from others. Exposure, as operationalized in Dominick et al. (1997) is akin to the third of these three. That is, subjects completed the 24 - item scale for themselves and their team members and anticipated feedback but did not expect to receive (nor did they actually receive) feedback prior to working on the second task. As a result, more than mere exposure may have been at work. It is conceivable that the obtained results were due to the cognitive processing of the information in the survey and or the anticipation of feedback from peers. Both of these factors may heighten individuals' awareness of their own behavior and encourage changes or improvement efforts. A more complete understanding of how a peer feedback instrument affects behavior could be obtained by isolating these variations of exposure. This could be done by expanding the number of conditions in the study from three to the following five groups:

Group 1: *No exposure, no rating, no anticipation, no feedback.* This is the control condition. subjects would not see the items, complete the ratings or receive feedback.

Group 2: *Exposure, no rating, no anticipation, no feedback.* Subjects would be exposed to the items at the end of the first task but would not complete peer ratings, anticipate or receive feedback.

Group 3: *Exposure, rating, no anticipation, no feedback.* Subjects would complete self and peer ratings at the end of the first task but would not anticipate or receive feedback.

Group 4: *Exposure, rating, anticipation, no feedback.* Subjects would complete the self and peer ratings at the end of the first task, anticipate feedback but do not receive it prior to working on the second task.

Group 5: *Exposure, rating, anticipation, feedback.* Subjects would complete the self and peer ratings at the end of the first task, anticipate feedback and receive it prior to working on the second task.

The research previously discussed suggests that individuals need not *receive* peer feedback in order to improve their behavior. At the same time, because we know so little about what aspect of exposure has the greatest impact, it is still reasonable to suggest that individuals will be more likely to modify their behavior as their level of cognitive interaction with the feedback instrument increases. For instance, just asking people to rate themselves on specific behaviors can help to shape their understanding of those behaviors and may lead them to establish standards for their own performance. These standards may then become the basis for self-assessment and subsequent improvement efforts. Requiring people to not only rate themselves but to also repeatedly rate others can further increase the time spent reflecting upon the behaviors and as a result, may lead subjects to develop an even stronger schema of effective team behavior. The rating of self and others should

also increase the likelihood that people will identify self-other discrepancies which in turn may lead to goal setting and behavioral improvement.

It is also possible that completing the instrument in anticipation of feedback will increase the salience of the information being reviewed, leading people to pay even closer attention to the items in the instrument. As a result, they should be more motivated to set improvement goals or regulate their future behavior. Specifically, the following hypotheses are proposed for investigation:

H1: The main behavioral effects of a peer feedback program come from being exposed to the feedback items, rating self and others and anticipating feedback.

*H1A: There **will be no** significant behavioral differences between subjects who complete a peer feedback instrument and anticipate feedback (yet do not receive it), and subjects who receive feedback. (Group 4 = Group 5)*

*H1B: There **will be** significant differences between subjects who complete a peer feedback instrument and anticipate feedback (yet do not receive it) in comparison with subjects who complete the instrument but do not anticipate feedback (Groups 4 & 5 > Group 3).*

*H1C: The behavior of subjects who complete the feedback instrument but do not anticipate or receive feedback **will be** significantly more effective than that of subjects who are merely exposed to the feedback items (Group 3, 4, 5 > Group 2)*

*H1D: The behavior of subjects who are exposed to the feedback items **will be** significantly more effective than that of subjects who are not exposed to the items (Group 2, 3, 4, 5 > Group 1).*

METHOD

Sample and Procedure

Participants were 128 undergraduate psychology students from several colleges and universities in northern New Jersey. Subjects were randomly organized into teams of three to five. Thirty teams were randomly assigned to one of the five conditions described above (feedback, control or one of the three exposure variations). The results section contains a more complete description of sample characteristics. Each team was videotaped performing two counterbalanced group decision-making tasks, one at each of two sessions on successive weeks. Appendix A contains complete experimenter instructions.

Tasks

All teams completed the same two group decision-making tasks, Black Bear (Glaser, 1993) and Outback (Glaser & Glaser, 1993) in counterbalanced order. Both tasks require team members to first read the case and make individual judgments regarding selection of a strategy alternative and the prioritized usefulness of several resources. Team members must work together and reach consensus on the most appropriate strategies and resource prioritization. Completion of these tasks requires significant interaction and discussion among participants. During both tasks, team members are required to work interdependently as a self-managed unit. As in many actual workplace teams, members must establish procedures for working together, jointly diagnose problems and alternatives and collaborate to develop solutions (Van de Ven, Delbecq, & Koenig, 1976). Appendix B contains copies of both tasks.

Feedback Instrument

The feedback form was based on a behavioral model of team performance that specifies four dimensions of team performance: Collaboration, Communication, Decision-making and Self-management (McGourty, DeMeuse & Dominick, 1994). A copy of the form is in Appendix C. It contains 24 behavioral items, six for each of the four dimensions. Subjects used a five point observation scale (1=never observed and 5=always observed) to rate themselves and all of their team members on the extent to which they displayed each of the behaviors listed on the form. Procedures for completing the team tasks according to condition are described below.

Exposure, rating, anticipation, feedback (Feedback condition)

Participants in the Feedback condition completed ratings on themselves and their teammates following completion of the first task. They were told that the next time they met with the experimenter, they would receive feedback reports based upon their ratings. Prior to beginning the second task, they were instructed to have a planning session during which they could discuss ways that they might wish to improve how they worked together. Before the planning session, they were given their feedback reports (Appendix C) and given approximately 5-10 minutes to review its contents. These individualized feedback reports included average dimension scores on each participants' self and peer ratings and a listing of self ratings and average peer ratings for each of the 24 items. Upon finishing the second task they completed another rating form based upon their interactions during that task as well as a post-treatment questionnaire (see Appendix D).

The post-treatment questionnaire served as a manipulation check and helped to determine whether or not individuals attempted to improve their behavior. The first three

questions investigated whether or not individuals' understanding of their behavior in a team setting was increased during the experiment (There are two issues here: 1) Were their schemas of team behavior impacted by the feedback instrument and process?; 2) Was their understanding of their own behavior enhance? This is important because as Ilgen, Fischer and Taylor (1979) suggested, the extent to which feedback is useful depends upon whether or not it increases someone's understanding of their behavior beyond that which they already possessed. The fourth and fifth questions asked people if they made deliberate efforts to change their behavior (e.g. set goals) during the second task. The sixth question asked people if they believed they actually changed their behavior. The seventh question asked individuals if they believed other team members expected them to change their behavior. An eighth question was included only for individuals in the Feedback condition and asked subjects if the peer feedback helped them improve their behavior.

Exposure, rating, anticipation, no feedback (Anticipation Condition)

Participants in this condition also completed ratings on themselves and their teammates following completion of the first task. Subjects were told that the next time they met with the experimenter, they would receive feedback reports based upon their ratings. They were not however, actually given feedback until the completion of the second task. Prior to beginning the second task, they were instructed to have a planning session during which they could discuss any ways that they might wish to improve how they worked together. Upon finishing the second task, they completed another rating form based upon their interactions during that task as well as the post-treatment questionnaire.

Exposure, rating, no anticipation, no feedback (Rating Condition)

Upon completion of the first task, subjects in this condition were asked to complete the rating form. However, they were told that the information was being collected for research purposes. They did not expect to receive feedback, nor did they receive any. Subjects in this condition, were also instructed to have a planning session prior to beginning the second task. Upon finishing the second task, they completed another rating form based upon their interactions during that task and also completed the post-treatment questionnaire.

Exposure, no peer-rating, no anticipation, no feedback (Exposure Condition)

Upon completion of the first task subjects in this condition were given a list of the items in the rating form. They were asked to review it and to rate themselves on each of the items listed (see Appendix C). The purpose of the self-rating was to ensure that they took the time to review and process the behavioral items listed on the form. Subjects in this condition were also instructed to have a planning session prior to beginning the second task. Upon finishing the second task, they completed the full (peer and self) rating form based upon their interactions during that task as well as the post-treatment questionnaire.

No exposure, no rating, no anticipation, no feedback (Control Condition)

Upon completion of the first task, participants in this control condition completed a placebo instrument on the task content (See Appendix E). They did not complete self or peer ratings, receive feedback, or receive exposure to the behavioral items. Subjects in

this condition were also instructed to have a planning session prior to beginning the second task. Upon finishing the second task, they completed the full (peer and self) rating form based upon their interactions during that task. They also completed the post-treatment questionnaire.

Dependent Variables

The teams were videotaped working on both tasks and experienced assessors rated the behavior of subjects working on the two tasks. A total of four assessors were used all of whom had previous experience doing behavioral assessment. Two of the assessors had evaluated team member behavior on similar tapes during the study by Dominick et al., (1997). All assessors were given standard instructions on observing and rating the subjects on the same 24-item rating form completed by participants in the feedback and exposure conditions (See Appendix F). The two assessors who had not done ratings during the Dominick et al. (1997) study also completed one practice tape and discussed their ratings with another assessor before beginning to rate the tapes used in this study. The videotapes of the teams were randomly distributed to the raters who were blind to condition. Each videotape was reviewed and independently rated by two assessors with the averages of the two ratings used as the dependent variables. In some instances, however, the same assessor reviewed tapes of the same group at Time-one at at Time-two. When this did occur, the tapes were never reviewed in succession and the experimenter asked the assessors if they were able to determine from the tape whether or not they were reviewing a first or second meeting. In many cases, the assessors did not

even remember having seen the individuals before and in no instance were they were they able to indicate whether they were reviewing a first or second meeting.

RESULTS

Descriptive statistics

For each subject (n=128), average assessor ratings for the items measuring Collaboration, Communication, Decision-making and Self-management were calculated for both Time-one and Time-two. These averages were derived by first summing the twelve ratings (six per dimension X two raters) for each dimension. The dimension averages for each subject were then combined to produce overall behavioral ratings at Time-one and Time-two for each subject. Table 1 lists means and standard deviations at Time-one and Time-two by condition and overall.

There were a total of thirty teams, each consisting of three to five members. Five teams were in the control condition. The exposure, rating and anticipation conditions each had six teams and the feedback condition had seven teams.

Cronbach's alpha was used to calculate the level of interrater agreement between expert raters. at Time-one (.82) and at Time-two (.85). The correlation between average expert ratings at Time-one and average expert ratings at Time-two was $r = .61$, $p = .001$.

Since all subjects completed the feedback instrument after Time-two, Cronbach's alpha was also used to obtain an estimate of their interrater agreement (.33). Since all subjects were rated by at least two others this estimate was based upon two sets of ratings per subject. Based upon the Spearman-Brown formula, reliability for three raters would be (.60) and for four raters reliability would be (.67)

Forty subjects were males and eighty-eight subjects were females. The average age of the subjects was 23.77 with a range from 18 to 49. One-way ANOVA tests revealed no

significant differences across conditions by age $F(4, 122) = .417, p = .796$ or by gender $F(1, 127) = 1.01, p = .32$.

Overall Results

A one-way ANCOVA, using expert ratings at Time-two as the dependent variable and expert ratings at Time-one as the covariate, found a significant main effect for condition $F(4,122) = 7.61, p = .000$. A 2X5 ANOVA (time by condition) was consistent with the results of the ANCOVA. This analysis found a significant main effect for time $F(1,123) = 8.77, p = .004$, a significant main effect for condition $F(4,123) = 6.17, p = .000$, and a significant interaction between condition and time $F(4,123) = 3.75, p = .007$. A third analysis also yielded similar results. This was a one-way ANOVA of subjects' Time-one -Time-two residual scores by condition. This test also found a significant main effect for condition $F(4, 123) = 7.55, p=.000$.

Analysis of Main Hypotheses

Stepwise hierarchical regression was used to test each of the planned comparisons relating to the hypotheses. Table 2 lists the contrasts codes set up for each planned comparison. A reduced versus full regression model (Overall & Spiegel, 1969) was used because the contrast codes were not perfectly orthogonal due to unequal within cell samples. Assessor ratings at Time-two were used as the dependent variable. Time-one ratings were used as a covariate and entered on the first step of the regression. The contrast codes representing each of the experimental conditions were entered on the next step. The overall $R^2 = .50$. Residual means and standard deviations are listed in Table 3. Results for each of the planned comparisons are listed below.

1) Results for No Exposure versus Exposure (Control condition versus all others)

The effect of exposure (self-rating only) to the feedback instrument versus no exposure was statistically significant $F(1,122) = 20.14, p = .000, R^2 \blacktriangle = .080$.

2) Results for Exposure versus Rating (Exposure condition versus all rating conditions)

The effect of using the feedback instrument to rate both self and others versus using it only to rate oneself was not statistically significant $F(1, 122) = .157, p = .6922, R^2 \blacktriangle = .0006$.

3) Results for Rating versus Anticipation of Feedback (Rating Condition versus Anticipation and Feedback Conditions)

The effect of using the feedback instrument to rate both self and others but without anticipating feedback versus completing the instrument in anticipation of feedback was not statistically significant $F(1, 122) = .228, p = .6337, R^2 \blacktriangle = .0009$.

4) Results for Anticipation of Feedback versus Actual Receipt of Feedback (Anticipation Condition versus Feedback Condition)

The effect of using the feedback instrument in anticipation of feedback but without actually receiving it versus actually receiving feedback was statistically significant $F(1,122) = 9.47, p = .0026, R^2 \blacktriangle = .0378$.

Tukey Comparisons

A post-hoc Tukey test conducted at the level of $p=.05$, compared the mean residual scores of each group to that of all others. This test found that the exposure, rating and feedback groups differed significantly from the control condition but not from one

another. The test also found a significant difference between the feedback and anticipation conditions.

A second post-hoc Tukey test was also conducted at the level of $p=.10$. This test did not detect any additional significant differences other than those found at $p=.05$. Table 4 lists the pairwise comparisons of residual means used for the Tukey analyses.

Power Analyses

In order to better assess the robustness of the findings, post-hoc power analyses were done for each of the planned comparisons and for the tukey comparisons at $p=.05$ and $p=.10$. Separate deltas were calculated for small (.20), medium (.50) and large (.80) effect sizes as defined by Cohen (1988).

For the first planned comparison (control condition versus all others), the power for a large effect size was .93, for a medium effect size power was .56 and for a small effect size power was $<.17$. For the second planned comparison (exposure condition versus all rating conditions), power for a large effect size was .94 , power for a medium effect size was .60 and power for a small effect size was $<.17$. For the third planned comparison (rating condition versus feedback and anticipation condition), power for a large effect size was .91, power for a medium effect size was .56 and power for a small effect size was $<.17$. For the fourth planned comparison (feedback condition versus anticipation condition), power for a large effect size was .83, for a medium effect size power was .44 and for a small effect size power was $<.17$.

For the Tukey comparisons at $p=.05$, power for a large effect size ranged from .80 to .83, for a medium effect size power was .44 and for a small effect size power was $<.17$.

For the Tukey comparisons at $p=.10$, power for a large effect size ranged from .88 to .9, for a medium effect size power was .56 and for a small effect size power was $<.26$.

Additional Analyses

The items comprising the post-treatment questionnaire (See Appendix A) provided some insight into the cognitive processes underlying individual reactions to the feedback instrument. Table 5 provides the mean responses and standard deviations overall and by condition for each item on the questionnaire.

Tukey Tests were used to test for differences by condition for each item. With an alpha level of .05, the tests revealed no significant differences between conditions for items one through six. However, for item 7, (“Other people on this team expected me to change my behavior from the first task to the second, “ a significant difference was detected between the feedback condition and the exposure only condition $F(4, 122) = 2.97, p = .022$.

Correlations between questionnaire items 4-7 and change in behavior as measured by residual scores were also determined. Correlations were obtained for all conditions combined and by condition (See Table 5). For item 7, significant correlations were found for all conditions combined, $r = .21, p = .008$; the anticipation condition, $r = .32, p = .053$; and for the feedback condition, $r = .36, p = .032$. The correlations were not significant for the other conditions (See Table 6).

Because all subjects completed the feedback instrument after Time-two, it was possible to calculate the level of agreement between peer ratings and expert ratings. For the entire sample overall the correlation was ($r = .03, p = .35$). By condition, the

The Effects of Peer Feedback Instrument on Team Member Behavior

correlations were as follows: control - ($r = -.29$), $p = .096$; exposure only - ($r = 0$), $p = .49$;
rating - ($r = .12$), $p = .28$; anticipation - ($r = .06$), $p = .38$; feedback - ($r = .31$), $p = .057$.

DISCUSSION

Theoretical and Practical Implications

The results of the analyses for the main hypotheses in this study suggest that: simply being exposed to behaviors (self-ratings) leads to improvement (behavior change). The detection for an effect from exposure is consistent with previous research on multi-source feedback e.g. (Smither et al., 1995; Reilly et al., 1996; Dominick et. al, 1997). On the other hand, these findings may suggest that simply completing self-ratings is a sufficient level of exposure to cause behavior change.

Earlier studies did not distinctly isolate this particular aspect of exposure (completing self-ratings). For instance, in the studies by Smither et al. (1995) and Reilly et al. (1996) exposure involved more than just *completing* self-ratings. The managers also received summary reports that organized their self-ratings by behavioral dimension and in some instances completed a feedback survey in order to provide ratings of their manager.

Another way in which the results from this study are different from those in prior studies is that the greatest rate of improvement was obtained for people who actually received feedback. It is harder however, to draw definitive conclusions about this finding. After all, the planned comparison used to test for the effect of feedback contrasted the feedback group with the anticipation group and subjects in this later group actually got worse at Time-two. Furthermore, post-hoc Tukey tests did not find a statistically significant difference between the feedback condition and the rating and exposure conditions. Nonetheless, the mean residual change score was greatest for people in the feedback condition. In addition, post-hoc analyses suggest that there may not have been

enough power in this experiment to detect a significant difference between the feedback conditions and the other exposure conditions.

Other data also suggest that subjects in the feedback condition responded differently than did those in the other conditions. First, the mean responses to six of the seven items in the post-treatment questionnaire were highest for subjects in the feedback condition. For example, they were more likely to report that they made deliberate and specific efforts to change their behavior. Although post-hoc analyses only detected a significant difference in the case of item #7, it is worth noting that mean responses tended to increase as levels of exposure to the feedback instrument increased.

The correlations between expert ratings and peer ratings at time-two provide some additional evidence for how the feedback condition differed from other conditions. Overall, there was virtually no correlation between expert and peer ratings. It was only in the feedback condition that a significant correlation was obtained. This finding may suggest that receiving feedback puts individuals in a better position to make subsequent judgments about one another's behavior. Theoretically, when subjects received feedback they were reviewing the behaviors within a self-referencing context. For instance, for each behavior they also received information about themselves. Research on the self-referencing effect suggests that people are more likely to recall information when it is directly related to them (Matlin, 1994, pp. 82-85). In this case, the self-referencing nature of the feedback may have increased individuals' awareness of the behaviors listed in their feedback reports. This increased awareness/ability to recall the behaviors also makes it

more likely that individuals would engage in the discrepancy identification and goal setting that control and goal-setting theories specify as prerequisites for improvement.

At the same time, the fact that overall, correlations between expert ratings and peer ratings were so dramatically low, calls into question the accuracy of the peer-based behavioral ratings. Even for those individuals in the feedback condition it is hard to argue that any improved performance is the result of the accuracy of the feedback, given the relatively low correlation. On the one hand, the main hypotheses for this study are based upon the psychometric assumption that there is one objective true score. From this perspective, the obtained peer ratings are not very helpful. It is possible that the “accuracy” of the peer feedback could have been improved by providing more detailed instruction to team members on what the behaviors meant and how to observe them. Such an additional step might be particularly important when introducing a feedback instrument to teams that have little or no knowledge of process issues in teams or groups.

From an another point of view however, it is also plausible to argue that for any feedback recipient, there are as many realities as there are observers/constituents and that there is value in understanding these multiple perspectives (Tornow, 1993). When looking at multi-source feedback from this point of view, feedback becomes less about getting the behavior “right” and more about establishing mutual expectations and standards amongst constituents to the feedback process. In this case, however, interrater agreement amongst peers was also quite low and this may indicate that these mutual expectations may not have been created.

In spite of questions regarding the accuracy of peer ratings, the data in this study suggest that cognitive interaction with a feedback instrument increases the likelihood that people will improve their behavior during subsequent interaction. Another finding from this study suggests that social influence processes may play a part as well. Across all conditions, a significant correlation ($r = .21$, $p = .008$) was found between behavior change (as measured by residual scores) and individuals' tendency to indicate that they felt *others* on their team *expected* them to change their behavior (Item #7 on the post treatment questionnaire). This correlation was highest for subjects in the anticipation condition ($r = .32$, $p = .053$) and the feedback condition ($r = .36$, $p = .053$). This may suggest that normative/social expectations may play an important role in encouraging people to change their behavior. The fact that these correlations increased for people who anticipated and received feedback might further imply that the feedback process helps to establish and reinforce social norms regarding behavior. One way of defining an organization's culture is as patterns of behaviors that are expected, accepted and supported by co-workers and facilitate fitting into the organization (e.g. Cooke and Rousseau, 1988). Therefore, because of their impact on behavioral expectations, feedback initiatives can play an important role in efforts to change an organization's culture.

Limitations

Before drawing more definitive conclusions, however, a number of limitations to this study should be noted. First, this study may have lacked sufficient power to detect

differences between the various types of exposure to the feedback instrument (self-rating; rating self and others; receiving feedback). Even at $p=.10$, power in this study is relatively weak for anything but a large effect size.

It also important to note that subjects in the anticipation condition (those who expected but did not receive feedback) present an exception to the general conclusions discussed so far. The behavior of those in this condition, actually decreased during Time-two. One explanation for this result may be that these subjects experienced a degree of disappointment or frustration when they did not actually receive feedback as anticipated. As a result, their motivation to set goals or to regulate their behavior was diminished. This explanation is consistent with the general principles of organizational justice and fairness theories. For example, equity theory (Adams, 1963) proposes that perceptions of inequity between one's efforts and one's rewards (outcomes) generate negative tension which in turn motivates cognitive and behavioral responses geared towards reducing this tension (Kanfer, p. 99, 1990). From this perspective, the experimenter's promise to provide feedback could be construed as the promise of a reward for one's effort. The fact that subjects did not actually receive this "reward" might then lead them to reduce their behavioral efforts. Assuming this explanation is correct, one important practical implication is that organizations run the risk of antagonizing members if they fail to follow through on feedback initiatives.

On the other hand, in contrast to the results of this investigation, the previous study conducted by Dominick et. al (1997) found no difference between subjects in the anticipation and feedback conditions. The difference between the findings of the two

studies may be the result of a subtle but important distinction between the instructions given to subjects in the anticipation conditions. In the current study, subjects in the anticipation condition were told that they would receive feedback the next time they met with the experimenter. In the previous study, they were told that they would be receiving feedback at some point in the future. The latter is a more general instruction and may have been less likely to create the expectation that feedback would actually be provided prior to working on the subsequent task. Consequently, individuals assigned to the exposure condition in the previous study may have been less likely to perceive any inequities between their efforts and a reward.

Another limitation of the present study has to do with whether or not the teams in this study should actually be defined as teams or groups. Although previous literature is not consistent on the definition of teams, Salas, Dickinson, Converse & Tannenbaum (1992) define teams as including the following characteristics: 1) Two or more people who interact dynamically, interdependently toward a common and valued goal, objective or mission.; 2) Have a limited life-span of membership. These same characteristics have been ascribed to groups (e.g. Johnson & Johnson, 1994, pp. 12-13) . In any event the “teams/groups” in this study possessed these characteristics. Although the teams met only twice for approximately one hour, the interactive tasks they performed were designed specifically to teach people about teamwork and what some researchers (e.g. Gaddy & Wachtel, 1992) refer to as “generic” team skills (behaviors that are beneficial to team members regardless of the work setting). Nonetheless, an increase in the number and length of team meetings would have provided increased opportunities to observe

behavior prior to the intervention. Additional meetings would also have allowed more opportunities for subjects to modify behavior in response to their level of exposure to the feedback instrument.

A third limitation has to do with the simulated environment of the study. Field replication of these results would strengthen conclusions. Organizational factors such as feedback from other sources and reward structures could confound and weaken the effects of feedback/exposure reported in this study.

As indicated in the Results Section, overall inter-rater reliability was .82 for Time-one and .85 for Time-two. These levels of reliability are better than field ratings of performance by supervisors. Rothstein (1990), for example, found the highest reliabilities for supervisor ratings of employees in field settings to be .55. Nonetheless, efforts should also be made to improve the inter-rater reliability of the observed behaviors. This could be done by obtaining more ratings, providing more training to raters and by requiring raters to review the tapes at least twice.

Future Research

Change resulting from feedback initiatives could probably be strengthened by supporting these initiatives with broader and more concerted change efforts. For example, we might expect even greater changes in behavior when information obtained through peer feedback instruments is strategically combined with additional interventions, such as the establishment of specific performance goals or by linking improvements to pay and other rewards. Along these lines, Hazucha et. al (1993) found that organizational support

systems such as career development enhance the likelihood that people will improve after receiving 360 feedback. Therefore, the impact of peer-feedback in combination with other interventions should be investigated as well.

One recent study, suggests that another important component of a multi-source feedback process is whether or not recipients conduct post-feedback discussions with feedback providers (Walker, 1997). Specifically, this study found that the consistency with which managers held post-feedback discussions with subordinates also positively impacted improvement. Although subjects in the present study were encouraged to discuss their behavior prior to beginning the second task they had relatively little time to do so. When they did take advantage of this opportunity, the discussions were more often focused on overall group process rather than on any individual team member. Furthermore, it is possible that the dynamics of such feedback discussions would be different in team/peer setting than they would be in manager-subordinate context. Therefore, future research could explore how post feedback discussions impact behavioral expectations and change in a team setting.

Another area of investigation could focus on the implications of varying the type of feedback provided (e.g. reporting self-results v. not including them). For example, Johnson and Ferstel (1997) suggest that it may be preferable to tailor the nature of feedback an individual receives depending upon their level of self-other agreement. They suggest that for in-agreement raters and under-raters, the self-other discrepancy should be de-emphasized and more emphasis placed on the absolute level of the ratings or a comparison to a norm-group. In the present study, it would have been interesting to ask

feedback recipients about what, if any particular aspect of their feedback most impacted their improvement efforts (e.g. absolute numbers or self-other discrepancies). The implications of providing team members with normative data, in addition to, self and peer ratings could also be explored. Finally, attempts should be made to identify relationships between behavioral changes due to peer-feedback and overall team task performance.

REFERENCES

- Adams, J.S. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67, 422-436.
- Atwater LE, Roush P, Fischthal A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology*, 48, 35-59.
- Bales, RF (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, 15, 257-263.
- Bernardin HJ, Hagan C, Kane JS. (1995, May). The effects of a 360 degree appraisal system on managerial performance: No matter how cynical I get, I can't keep up. Paper presented at the 10th Annual Meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Burke M, Day R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, 47, 232-245.
- Carver CS, Scheier MF. (1981). Attention and self-regulation: A control theory approach to human behavior. New York: Springer-Verlag.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd.ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooke RA, Rousseau DM (1988). Behavioral norms and expectations: A quantitative approach to the assessment of organizational culture. *Group and Organization Studies*, 13, 245-273.

- Cardy RL, Keefe TJ. (1994). Observational purpose and evaluative articulation in frame-of-reference training: The effects of alternative processing modes on rating accuracy. Organizational Behavior and Human Decision Processes, 57, 338-357.
- DeNisi AS, Cafferty TP, Meglino BM. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.
- Dominick PG, Reilly, RR., McGourty J. (1997). The effects of peer feedback on team member behavior. Group and Organization Management, 22, 508-520.
- Feldman JM. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Gaddy CD, Wachtel JA. (1992). Team skills training in nuclear power plants. In RW. Swezey & E. Salas (Eds.), Teams: Their training and performance (pp. 379--396). Norwood, NJ: Ablex Publishing.
- Gladstein D. (1984). Groups in context: A model of task group effectiveness. Administrative Science Quarterly, 29(4), 499-517.
- Glaser B. (1993). Black bear: A team adventure. King of Prussia, PA: Organization Design & Development.
- Glaser R., Glaser C. (1993). Outback: A team adventure. King of Prussia, PA: Organization Design & Development.
- Hackman JR. (1983). A normative model of work team effectiveness. (Technical Report: No. 2, Research Program on Group Effectiveness). New Haven, CT. Yale School of Organization and Management.

- Hackman JR. (1987). The design of work teams. In J.W. Lorsch (Ed.), Handbook of organizational behavior (pp.315-342). Englewood Cliffs, NJ: Prentice Hall.
- Hackman JR., Morris, CG. (1975). Group tasks, group interaction process and group performance effectiveness: a review and proposed integration. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, 8, (pp.45-99). New York: Academic.
- Hazucha JF, Gentile SA, Schneider RJ. (1993). The impact of 360-degree feedback on management skills development. Human Resources Management, 32, 325-352.
- Hegarty WH. (1974). Using subordinate ratings to elicit behavioral changes in supervisors. Journal of Applied Psychology, 59(6), 764-766.
- Ilgen D, Fischer C, Taylor M. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64(4), 349-371.
- Johnson DW, Johnson, FP. (1994). Joining together: Group theory and group skills. Boston, MA: Allyn & Bacon.
- Johnson, JW., Ferstl, KL. (1997, April). The effects of interrater and self-other agreement on performance improvement following upward-feedback. Paper presented at the Twelfth Annual Meeting for the Society for Industrial and Organizational Psychology, St. Louis, MO.
- Katzenbach J., Smith D. (1993). The wisdom of teams. Cambridge, MA: Harvard Business School Press.
- Kanfer R. (1990). Motivation theory and industrial and organizational psychology. In MD Dunnette and LM hough (Eds.), Handbook of industrial/organizational

psychology, 2nd edition, (pp.75-170). Palo Alto, CA: Consulting Psychologists Inc. Press.

Kernaghan JA, Cooke, RA (1986). The contribution of the group process to successful project planning in R&D settings. IEEE Transactions on Engineering Management, 33, 134-140.

Kernaghan JA., Cooke RA. (1990). Teamwork in planning innovative projects: Improving group performance by rational and interpersonal interventions in group process. IEEE Transactions on Engineering Management, 37, 109-116

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

Lawler EE, Mohrman SA, Ledford GE, Jr. (1992). Employee involvement and total quality management: Practices & results in fortune 1000 companies. San Francisco: Josey-Bass.

Locke EA., Latham GP. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice Hall.

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

Lord R , Hanges P. (1987). A control system model of organizational motivation. Behavioral Science, 32, 161-178.

Matlin, MW. (1994). Cognition. Fort Worth, FL: Harcourt Brace Publishers.

- McGourty JW, DeMeuse KP, Dominick PG. (1994, August). Self-managed work teams: A conceptual model for practice & research. Paper presented at the Academy of Management Annual Meeting, Dallas, TX.
- Mohrman SA Cohen SG, Mohrman AM (1995). Designing team-based organizations. San Francisco: Josey-Bass.
- Murphy KR., Cleveland JN. (1991). Performance appraisal: An organization perspective. Needham Heights, MA: Allyn & Bacon.
- Nadler DA. (1977). Feedback and organization development: Using data based methods. Reading, MA: Addison-Wesley.
- Nieva VF, Fleishman EA, Reick A. (1978). Team dimensions: Their identity, their measurement and their relationships. (Technical Report: No. DAHC19-78-C-0001. Washington, D.C.: Advanced Research Resources Organizations.
- Peters T. (1988). Thriving on chaos. New York: Harper & Row.
- Ramsay ML, Letho H. (1994). The power of peer review. Training & Development, July, 38-41.
- Reich RB (1987). Entrepreneurship reconsidered: The team as hero. Harvard Business Review, 65(3), 77-83.
- Reilly RR, Smither JW, Vasilopoulos NL. (1996). A longitudinal study of upward feedback. Personnel Psychology, 49, 599-612.
- Romano C. (1994). Conquering the fear of feedback. HR Focus, 71, 9-19.

- Rothstein HR. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology; 75, 322-327.
- Saavedra R, Kwun SK. (1993). Peer evaluation in self-managing work groups. Journal of Applied Psychology, 78, 450-462.
- Salas E, Dickinson TL, Converse SA, Tannenbaum SI.. (1992). Toward an understanding of team performance and training. In RW. Swezey & E. Salas (Eds.), Teams: Their training and performance (pp. 3-29). Norwood, NJ: Ablex Publishing.
- Smither JW, London M, Vasilopoulos NL, Reilly RR, Millsap RE, Salvemini N. (1995). An examination of the effects of an upward feedback program over time. Personnel Psychology, 48(1), 1-32.
- Smither JW, Wohlers AJ, London, M. (1995). A field study of reactions to normative versus individualized upward feedback. Group and Organization Management, 20, 61-89.
- Stevens MJ, Campion MA (1994). The knowledge, skill and ability requirements for teamwork: Implications for human resources management. Journal of Management, 20, 503-530.
- Sundstrom E, DeMeuse KP, Futrell D. (1990). Work teams: Applications and effectiveness. American Psychologist, 45, 120-133.
- Tabachnik BG, Fidell, LS (1989). Using multivariate statistics, 2nd edition. New York: Harper-Collins.

Tornow WW. (1993). Perspective or Reality: Is multi-perspective measurement a means or an end? Human Resources Management, 32, 221-229.

Van de Ven AH, DelBecq AL, Koenig R. (1976). Determinants of coordination modes within organizations. American Sociological Review, 41, 322 -338.

Van Velsor E, Leslie JB. (1991). Feedback to managers, Volume II: A review and comparison of sixteen multi-rater feedback instruments. Greensboro, NC: Center for Creative Leadership.

Walker AG. (1997, April). Upward feedback: Incremental improvement in managers' performance over five years. Paper presented at the 12th Annual Meeting for the Society for Industrial and Organizational Psychology, St. Louis, MO.

Woodman RW, Sherwood JJ. (1980). The role of team development in organizational effectiveness: A critical review. Psychological Bulletin, 88, 166-186.

Zigon J. (1994). Making performance appraisal work for teams. Training, June, 58-63.

Table 1
Means and standard deviations for overall team behavior ratings

Condition	Time-1		Time-2	
	Mean	SD	Mean	SD
1) Control (n = 22)	2.64	.95	2.53	.75
2) Exposure (n = 26)	3.14	.52	3.38	.52
3) Rating (n = 26)	3.07	.71	3.33	.72
4) Anticipation (n = 26)	3.15	.50	3.11	.37
5) Feedback (n = 28)	3.09	.50	3.52	.64
Total (n = 128)	3.03	.66	3.20	.69

Table 2
Contrast codes for regression analysis and F values

- A) Control v. All Other Conditions (Groups 2, 3, 4, 5 > Group 1)
- B) Exposure v. All Rating Conditions (Groups 3, 4, 5 > Group 2)
- C) Ratings Only v. Anticipation and Feedback Conditions (Groups 4, 5 > Group 3)
- D) Anticipation v. Feedback Condition (Group 5 > Group 4)

Condition	A	B	C	D
1) Control (n = 22)	-1	0	0	0
2) Exposure (n = 26)	¼	-1	0	0
3) Rating (n = 26)	¼	1/3	-1	0
4) Anticipation (n = 26)	¼	1/3	½	-1
5) Feedback (n = 28)	¼	1/3	½	1
Result - F Value	20.14*	.157 ns	.228 ns	9.47*

* p < .05

Table 3
Residual scores and standard deviations

Condition	Residual Score	
	Mean	SD
1) Control (n = 22)	-.42	.51
2) Exposure (n = 26)	.11	.44
3) Rating (n = 26)	.11	.59
4) Anticipation (n = 26)	-.15	.33
5) Feedback (n = 28)	.28	.55
Total (n = 128)	.00	.54

Table 4
Pairwise comparisons of mean residual scores (ordered by size)

Means	1	4	2	3	5
1) -.42 (Control)	-	.27	.53*	.53*	.70*
4) -.15 (Anticipation)	-	-	.26	.26	.44*
2) .11 (Exposure)	-	-	-	0	.18
3) .11 (Rating)	-	-	-	-	.17
5) .28 (Feedback)	-	-	-	-	-

*p < .05

Range statistic at .05 level = 3.92; at .10 level = 3.52

Table 5
Response means and standard deviations for post-treatment questionnaire

Question*	Control		Exposure		Rating		Anticipation		Feedback		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1) Understanding of team behavior	3.95	.72	4.04	.72	4.16	.8	4.19	.8	4.36	.48	4.14	.71
2) Understanding of own behavior	4.09	.61	4.27	.78	4.08	.76	4.12	.91	4.07	.81	4.13	.78
3) Understanding of ways to improve	3.95	.65	3.81	.98	3.96	1.06	4.0	.89	4.14	.76	3.98	.88
4) Made deliberate efforts to change behavior	3.05	.84	3.04	1.34	3.6	1.41	3.35	1.13	3.75	1.04	3.37	1.19
5) Set specific goals	2.72	.88	3.0	1.06	2.96	1.24	2.96	1.11	3.39	1.03	3.02	1.08
6) Successfully changed behavior	2.73	.77	2.96	1.34	3.00	1.22	2.92	.89	3.32	.98	3.0	1.07
7) Others expected me to change my behavior	2.18	.91	2.04	.92	2.20	1.19	2.27	.96	2.89	1.03	2.33	1.04
8) Peer Feedback helped me improve**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	3.78	.89	n/a	n/a

* See Appendix D for complete questions.

** Question 8 only applies to Feedback Condition.

Table 6
Correlations between responses to post-treatment questionnaire and residual scores

Question*	Control	Exposure	Rating	Anticipation	Feedback	Overall
All items combined (1-7)	.21 p = .170	- .014 p = .472	- .09 p = .326	.10 p = .31	.27 p = .079	.16 p = .038
4) Made deliberate efforts to change behavior	.43 p = .021	- .15 p = .235	- .07 p = .358	- .16 p = .216	.21 p = .152	.09 p = .149
5) Set specific goals	.02 p = .463	- .01 p = .478	- .25 p = .115	.16 p = .223	.27 p = .082	.09 p = .149
6) Successfully changed behavior	.34 p = .061	- .01 p = .475	- .07 p = .37	.01 p = .482	- .0021 p = .496	.09 p = .154
7) Others expected me to change my behavior	- .14 p = .271	.20 p = .171	.10 p = .32	.32 p = .053	.36 p = .032	.21 p = .008
	n=22	n=26	n=26	n=26	n=28	n=128

* See Appendix D for complete questions.

APPENDIX A

Experimenter Instructions

(Not Included)

APPENDIX B

Team Tasks

(Not Included)

APPENDIX C

Rating Forms and Feedback Report

(Not Included)

APPENDIX D
Post-Treatment Questionnaires
(Not Included)

APPENDIX E
Placebo Questionnaire for Control Condition
(Not Included)

APPENDIX F
Assessor Instructions

(Not Included)