



**A Long-Term Study of the Impact of a 360 Feedback
Process on Self-Others' Agreement, and Performance**

by

Frank Shipper
Professor of Management & Chair
Franklin P. Perdue School of Business
Salisbury University
1101 Camden Avenue
Salisbury, MD 21801
fmshipper@salisbury.edu
Phone: 410 543-6333

Paper to be published in abbreviated format in
The Best Papers Proceedings of the
The Academy of Management Annual Meeting
August 6-10, 2010
Montreal, Canada

A Long-Term Study of the Impact of a 360 Feedback Process on Self-Others'
Agreement, and Performance

ABSTRACT

The purpose of this study is to investigate the impact of multiple iterations a 360 feedback process has on self-others' agreement, over versus under estimating, and performance. Such a study is needed because many of the prior studies on self-others' agreement have been limited to panel data and thus, interpretations regarding causality cannot be made. Both performance and self-others' agreement improved over the course of the study. How improvements in managerial skills and performance were impacted by self-others' agreement, and over versus under estimating was found to be complex. The implications for future research and practice are discussed.

A Long-Term Study of the Impact of a 360 Feedback Process on Self-Others' Agreement, and Performance

One of the basic propositions underlying the use of any 360 feedback process is that the comparison of self-evaluations of skills to those by others will help individual managers see themselves as others see them and that this in turn will motivate them to improve their skills (e.g., Smither, London, Vasiliopoulos, Reilly, Millsap, & Salvemini, 1995; Wilson & Shipper, 1992). In other words, do individual managers care enough about what others say to be motivated to improve? A corollary to this proposition is that the greater the improvement in self-other agreement, the greater the improvement in performance (Leslie & Fleenor, 1998; Smither, London, & Reilly, 2005).

Neither the proposition nor the corollary has been widely investigated. The majority of the research that has been reviewed on 360 feedback has been based on what could be described as one-shot case studies (London & Smither, 1995; Seifert, Yukl, & McDonald, 2003; Smither, London & Reilly, 2005). In addition, the majority of the research that has been reviewed on self-other agreement has been based on what could be described as panel data (Coway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Heidemeier & Moser, 2009; Mabe & West, 1982). Inferences in causality cannot be drawn from panel data. Thus, how changes in self-other agreement may have impacted changes in either skills or performance is largely unstudied, especially in studies using cross-sectional data.

Static correlations between self-other agreement and performance can readily be found in the literature (e.g. Bass & Yammarino, 1991; Atwater, Ostroff, Yammarino, & Fleenor, 1998; Atwater & Yammarino, 1992; Furnham & Stringfeld, 1994). Other research, however, casts doubt on the ability of individuals to use that information for making change (Kruger &

Dunning, 1999) or if that ability is widely held (Shipper & Dillard, 2000). Thus, studies using research designs that address these issues are needed.

Other research can also cast doubt on self-other agreement as a singular measure (Shipper & Davy, 2002; Shipper & Dillard, 2000). Much of the research on the static correlations cited in the previous paragraph is based on a single questionnaire of transformational leadership. Other research has found that questionnaire to measure a single dimension. In contrast, other researchers using a different questionnaire have found two or more dimensions of managerial skills (Shipper & Davy, 2002, Shipper, Hoffman, & Rotondo, 2007). In addition, in one study where self-other agreement was hypothesized and tested in a structural model, the researchers found support for two distinct factors – interactive skills and initiating skills (Shipper & Davy, 2002). In addition, the researchers found a complex relationship between self and others' evaluations of skills and performance. First, self evaluations of interactive skills were found to have a positive direct association with others' evaluations of interactive skills, which had a positive association with employee attitudes, and which then had a positive association with performance. In addition, researchers hypothesized and found a significant negative direct relationship between self evaluations of interactive skills and performance. Second, self evaluations of initiating skills were found to have a positive direct association with others' evaluations of initiating skills, which had a positive association with employee attitudes and which then had a positive association with performance. In addition, researchers hypothesized a negative, but found a significant positive direct relationship between self evaluations of initiating skills and performance. The researchers explained this unexpected finding in that it may be socially desirable to report strong interactive skills, but less so to report strong initiating skills.

These complex relationships from a review of the literature do not appear to have been investigated further, but further study appears to be warranted.

Another set of issues that has been raised in some of the literature on self and others' evaluations is (1) does overestimating of skills have different impact than accurate estimating, (2) does underestimating of skills have different impact than accurate estimating, and (3) does overestimating of skills have different impact than underestimating? Some of the research on self and others' ratings focus on accuracy and do not treat overestimating differently than underestimating (e.g. Atwater, Wang, Smither, & Fleenor, 2009); while others have (Shipper & Davy, 2002; Vecchio & Anderson, 2009). When overestimating and underestimating have been tested, the results suggest that their impact on performance may differ. Since all three conditions – overestimating, accurate estimation and underestimating – are possible and what research does exist on the three conditions suggests that their impact may be different, it would appear that additional research is warranted. For example, overestimating has been linked with the tendency for managers to derail in their careers (Shipper & Dillard, 2000). Narcissism has been found to be associated with overestimating on leadership skills and also to have a negative relationship on performance (Resick, Whitman, Weingarden, & Hiller, 2009). Managers who derail have been described in the literature as arrogant and abrasive (Lombardo & McCauley, 1988; McCall & Lombardo, 1983). Thus, overestimating is expected to have a negative relationship on performance. Underestimating has not been as widely studied. In studies where underestimating was investigated it was found in one study to be associated with the ability to recover from derailment, whereas the overestimating was associated with continued derailment (Shipper & Dillard, 2000); but in another study underestimating was found to be associated with lower performance (Vecchio & Anderson, 2009). The first study was based on cross-sectional data,

whereas the second study was based on panel data. In a third study where self-awareness of both interactive and controlling skills was studied using panel data, overall support for self-awareness to be associated with performance was found. The relationship was found, however, to be curvilinear in all three cultures studied (Shipper, Kincaid, Rotondo, & Hoffman, 2003). Thus, additional research is needed to understand the impact of underestimating on performance.

To consolidate the issues discussed into testable hypotheses, five sets of hypotheses were generated. First, to test whether greater agreement across both sets of skills occurs over time between self and others' evaluations the following two hypotheses will be tested:

H1a: The difference in absolute value between self and others' evaluations of interactive skills will decrease over multiple integrations of the 360 feedback process.

H1b: The difference in absolute value between self and others' evaluations of controlling skills will decrease over multiple integrations of the 360 feedback process.

Second, to test for whether overestimating or underestimating of interactive and controlling skills impacts changes in skills, respectively, the following two hypotheses will be tested:

H2a: Changes in interactive skills observed at time $n+1$ will be associated with overestimating versus underestimating of same skills at time n .

H2b: Changes in controlling skills observed at time $n+1$ will be associated with overestimating versus underestimating of same skills at time n .

Third, to test whether the absolute value of the difference between self and others' evaluations of interactive skills impacts the changes in skills, respectively, the following two hypotheses will be tested:

H3a: Changes in interactive skills observed at time $n+1$ will be associated with the absolute difference between self and others' evaluations of the same skills at time n .

H3b: Changes in controlling skills observed at time $n+1$ will be associated with the absolute difference between self and others' evaluations of the same skills at time n .

Because the second and third set of hypotheses could be either competing or complementary hypotheses, they will be tested simultaneously.

Fourth, to test for whether overestimating or underestimating of interactive and controlling skills impacts changes in performance, the following two hypotheses will be tested:

H4a: Changes in performance observed at time $n+1$ will be associated with overestimating versus underestimating of interactive skills at time n .

H4b: Changes in performance observed at time $n+1$ are associated with overestimating versus underestimating of controlling skills at time n .

Fifth, to test whether the absolute value of the difference between self and others' evaluations of interactive skills impacts the changes in performance, the following two hypotheses will be tested:

H5a: Changes in interactive skills observed at time $n+1$ are associated with the absolute difference between self and others' evaluations of the same skills at time n .

H5b: Changes in controlling skills observed at time $n+1$ are associated with the absolute difference between self and others' evaluations of the same skills at time n .

Because the fourth and fifth set of hypotheses could be also either competing or complementary hypotheses, they will be tested simultaneously too.

METHOD

This study was derived from a middle-level management development program that was tracked for approximately 15 years and incorporated action research. In other words, when evidence to suggest that adjustments were appropriate, they were made. The changes will be described in the discussion of this study.

Design

With five observations this field study can be viewed as a series of four pretest-posttest designs (Campbell & Stanley, 1963). The data were gathered to monitor, evaluate and adjust the management development program as it occurred. Participants joined the program at different times. Early participants were considered a pilot group. Attrition of participants also occurred due to some people being promoted to senior positions, others leaving the firm, still others not continuing with the management development program, and direct reports, superiors, or managers not providing data.

Many of the research studies on 360 feedback included in the reviews referenced in the introduction of this paper have been from a single pretest-posttest design. For example, of the 14 studies included in the Seifert, Yukl, and McDonald's (2003) review, eight were of this pre-experimental design. Any such study has the potential for multiple internal sources of invalidation (Campbell & Stanley, 1963). Any one pretest-posttest study in this research also has the potential for the same internal sources of invalidation, but by incorporating four pretest-posttests occurring at different times in different places the probability is reduced that systematic contamination will occur that would call the results into question.

Sample

The sample consists of 13,661 managers of a large multi-national technology-driven firm who participated in the 360 feedback process during the time frame of this study. The average age of the participants was 40 with an average length of service of 11 years when they first engaged in this study. Demographically, 25% of the participants were female and 75% male. The participants were 71% from the US and 29% from other countries.

Interventions

The 360 feedback process as defined by Shipper, Hoffman, and Rotondo (2007) was used as the intervention in this study. In brief, the 360 feedback process consists of (1) the collection of data from peers, superiors, direct reports, and self; (2) the reporting of that information to participants via a report; (3) follow-up training and support; (4) the recollection of the data from peers, superiors, direct reports, and self; and (5) the reporting of that information to participants via a report. This process was repeated four times during the study; thus, four interventions occurred.

The 360 instrument used to gather data in this study was The Survey of Management Practices (Wilson & Wilson, 1991). It and a number of other 360 instruments were reviewed by an organizational task force and two were presented to the Chief Executive Officer (CEO) for final selection. The task force reviewed the instrument on face validity, comprehensive nature, and psychometric soundness. The CEO selected the instrument largely based on its face validity (Personal Communication, May 1990).¹

In this instrument evaluation of skills is assessed using 48 items which are broken down into multiple item scales. Direct reports of each manager rated the manager on each of these

¹ To protect the anonymity of the company the individual interviewed is not identified.

scales. Both scale and item feedback were presented to each manager receiving more than five anonymous responses.

Both research prior to the beginning of the study and during it found the instrument to be psychometrically sound (Clark, Clark, & Campbell, 1992; Leslie & Fleenor, 1998; Morrison, McCall & DeVries, 1978; Rosti & Shipper, 1998; Shipper, 1995; Shipper & Davy, 2002; Shipper & White, 1999; Van Velsor & Leslie, 1991; Wilson, 1975, 1978). These studies have examined the questionnaire and found its test/re-test reliability, internal consistency, interrater reliability, aggregation suitability, construct validity, and criterion validity to be acceptable. In addition, the questionnaire has been examined for stability across organizations and across cultures (Shipper, Kincaid, Rotondo, & Hoffman, 2003). All of the scales within the instrument have been reported in prior studies to exceed Nunnally's (1978) criteria of .70 for reliability. Thus, the instrument had both face and psychometric validity.

Theoretically, the 360 intervention used in this study was drawn from Wilson's Task Cycle Theory, which is based on Tolman's (1932) learning theory and Bolles's (1972) updating of it (Wilson, O'Hare, & Shipper, 1990). Prior research on the Wilson model has validated the skills and demonstrated their impact on managerial performance (Shipper & Wilson, 1992; Shipper, 2004; Shipper, Kincaid, Rotondo, & Hoffman, 2003; Shipper, Hoffman & Rotondo, 2007). One advantage of this model is that it presents the skills such as goal setting, planning and problem solving, coaching, conducting a performance appraisal, exercising control, and reinforcing performance in an integrated model.

Managers were given skills and performance feedback in the form of a report summarizing the data from superior, subordinates, self and peers. In the early years of the study this feedback was part of a five-day management retreat with colleagues from multiple work

sites from around the world. In the half-day dedicated to the report, the focus was on how to interpret the report. After some research into the effectiveness of the program the five-day management retreat was reduced to a one-day in situ management workshop with the majority of managers from the same work site. In this workshop, the participants received a report and instructions on how to interpret it early in the day. During the middle of the day what resources were available to improve a specific skill, how to develop each skill, and what skill to focus on for the biggest payoff were discussed. By the end of the day the participants were encouraged to have developed a preliminary learning agenda. To further develop the learning agendas, each participant was encouraged to share the results with his/her superior, peers, and subordinates, and based on that input, to finalize his/her learning agenda. As part of the learning contract, participants were expected to experiment with different skills and techniques, to practice them, and to receive feedback from their subordinates and peers. The organization provided also additional skill training support from a variety of sources. The explicit use of learning contracts resulted from the use of action research during this study. Early in the program, some participants were observed to have made significant improvement and the converse was also true. Informal interviews were conducted to learn why this was occurring. From the interviews, it was thought that those who had developed learning contracts, even informally, were making progress, whereas those who did not may or may not have been. A survey was conducted to confirm the results of the interviews. The results were that 100% of the managers who had developed learning contracts had improved, whereas the same was not true for those who did not develop such (Personal communication, May 17, 1994). Given that learning contracts are considered to be one of the most powerful tools for helping people to reprogram the neuro-pathways of even stroke victims, these findings should not be surprising (Law, 2004).

To build trusting relationships, groups of managers from the same facility were organized into intact teams, asked to share problems and develop action plans to solve them. The use of intact teams of managers was initiated at some facilities. Intact teams consisted of all the managers in a facility participating in the 360 process together. As part of the action research model, statistical analysis was conducted on improvements made by participants in intact teams versus those not in such teams. The results suggested that those in intact teams were making statistically significantly ($p \leq .01$) greater improvements than those not in such teams. This prompted the adoption of the one-day workshop described earlier as the normal way to deliver the feedback.

Members of the team who scored high on their skill profiles were asked to serve as mentors to others. These development activities are similar to those used in two of the four multilevel, pre-post studies identified by Seifert, Yukl, and McDonald (2003) as having shown significant improvements. Those two studies were reported in Wilson, O'Hare, and Shipper (1990). Additional description of the training activities can be found in other studies (Shipper, Hoffman & Rotondo, 2007; Wilson & Shipper, 1992).

After a period of time, the participant's skills and performance were assessed again to determine the level and impact of learning that had occurred. Each iteration of the 360 feedback process lasted approximately 18 months. The time lag on prior 360 studies has varied from three to 48 months, with an average lag of 13.4 months (Seifert, Yukl, & McDonald, 2003). Because learning can be characterized as a permanent change in behavior, it may take significant time to develop improved managerial skills and for them to have observable impact on the manager's performance. Thus, a longer time period between assessments appears to be appropriate rather

than a shorter one. In addition, Taylor, Russ-Eft, and Chan (2005) have called for studies that assess training effects for periods greater than one year out, something this study satisfies.

This process parallels the Boyatzis self-directed learning model (Boyatzis, 1999; Goleman, Boyatzis, & McKee, 2002). This model as it pertains to the 360 feedback process is portrayed in Figure 1. The 360 feedback process as delineated is dynamic and iterative, providing for continuous feedback on skill use to allow for future changes in behaviors.

Insert Figure 1 about here

Measures

As previously discussed, The Survey of Management Practices (Form LB) was used to collect data (Wilson & Wilson, 1991). Reasons for its selection were also previously discussed.

Specific skills used in this study—interactive and controlling—were taken from prior research. These measures were developed through cluster analysis and have been found to be stable (Shipper & Wilson, 1992; Shipper, 2004; Shipper, Kincaid, Rotondo, & Hoffman, 2003). Cronbach's alphas for both self and others' evaluations of interactive skills exceeded .90. Cronbach's alphas for both self and others' evaluations of controlling skills exceeded only .60. The latter two reliability scores are considered low, but acceptable for exploratory research. Self-others' evaluations are measured in two ways. First, self-others' evaluation is measured as the absolute difference between the evaluations of the direct reports for the measures of interactive and controlling skills and the self-reported evaluations of the same skills. Second, self- other's evaluation was measured by using an indicator variable – zero for underestimating and one for overestimating. Managers who evaluated their skills to be less than what their direct reports evaluated them to be were classified as under-estimators. Conversely, managers who evaluated

their skills to be greater than what their direct reports evaluated them to be were classified as over-estimators. For statistical testing of Hypotheses 2 and 3, change scores were used for the dependent variable. The change score is simply the difference between the two scores as measured in the pre and post administration of the Survey of Management Practices for each of the four interventions and overall, t_1 to t_5 .

To assess managerial performance, four items were asked of the superiors when completing the Survey of Management Practices. A typical item asked was if the manager's group "works well." Analysis of the scale found Cronbach's Alpha equal to .90. This measure was chosen as the criterion variable for Hypotheses 4 and 5 because it is representative of the fourth and ultimate level in Kirkpatrick's (1959) evaluation model. Kirkpatrick's model is recognized as one of the most influential assessments of development interventions (Smither, Houston, & McIntire, 1996). For Hypotheses 4 and 5, the change score was used as the dependent variable for each of the four interventions.

Data Collection

Data for both the skills and the criterion variable were initially collected with paper instruments. As the study progressed, the data were collected electronically via the Internet. Procedures were taken to protect the anonymity of the direct reports who responded by using a third party to collect data via mail in the early stages and by a secure server in the later stages of the study. The data and the individualized reports were considered to be the property of the participants and not the organization. The researcher was involved early in the process and was granted permission to publish aggregated results. The response rates of the managers (self, superiors and subordinates) were 99.7%, 84% and 63%, respectively.

Analysis

To report statistics on the data used in this study, Table 1 was constructed reporting the means, standard deviations and bivariate correlations for the variables. To test the overall effectiveness of the 360 feedback process, one-way analysis of variance was also performed using Duncan's multiple range tests to conduct multiple comparisons over performance evaluations reported for times 1 through 5. Without evidence of changes in performance, the results of the study could be called into question. To test Hypotheses 1 and 2, one-way analyses of variance were also performed using Duncan's multiple range tests to conduct multiple comparisons over the five observations of interactive and controlling skills.

Insert Table 1 about here

The correlations in Table 1 were used to test for multicollinearity among the variables. No two variables had a variance inflation factor (VIF) greater than 1.33. Thus, since variables with VIF's less than 5 are considered to be unsusceptible to multicollinearity, regression analyses were used for testing Hypotheses 3 through 5 (O'Brian, 2007). Since it can be shown mathematically to be easier to improve lower skill or performance levels than higher, the skill or performance level at the beginning of the interventions was entered as a covariate as appropriate.

Results

The results of one-way analyses of variance were also performed using Duncan's multiple range tests to conduct multiple comparisons. They are reported in Table 2. The results of the one-way analysis of variance for the superior reported performance and direct report observations of interactive and controlling skills were all highly statistically significant ($p < .001$). In addition, performance steadily increased from t_1 to t_5 , whereas the absolute value of difference between self versus others' evaluations of both the managers' interactive and controlling skills

steadily declined from t_1 to t_5 . Thus, support was found that the 360 feedback process significantly impacted the managers' performance and for H1a and H1b.

Insert Table 1 about here

For ease of interpretation and discussion, the mean of each observation was standardized for the overall sample. Those standardized means were then transformed into percents and reported as a line graph in Figure 2.

Insert Figure 2 about here

Duncan's multiple range tests to conduct multiple comparisons on (1) changes in effectiveness, (2) the absolute values of self and others' evaluations of interactive skills, and (3) the absolute values of self and others' evaluations of controlling skills indicate with the exception of the absolute values of self and others' evaluations of interactive skills significantly different changes did not occur after a single intervention. These tests also showed that significantly different changes did occur after two or three interventions for all three criterion variables.

Hypotheses 2 and 3 are tested using regression. In the regression equations, the dependent variable is either change in interactive or controlling skills and the covariate, the initial level of interactive or controlling skill, respectively. The absolute difference in the self versus others' evaluations of skills and the over versus under estimating of skills are the independent variables. The tests are reported for all four interventions and overall. The results are in Table 2.

As can be seen in Table 3, all 10 equations are highly significant and the covariate in each equation is highly significant with a negative coefficient. The covariates explain the majority of the explained variance in each equation. Only in one of the 10 equations is the absolute difference in the self versus others' evaluations of skills significant. Only in one of the five equations on interactive skills is the over versus under indicator variable significant and then only marginally ($p \leq .10$). In four of the five equations on controlling skills the over versus under indicator variable is significant ($p \leq .05$).

Insert Table 3 about here

As can be seen in Table 4, nine equations are significant overall and in each of those equations the covariate is highly significant with eight having negative coefficients. The covariates explain the majority of the explained variance in each significant equation. Only in two of the ten equations is the over versus under indicator variable significant and then only marginally ($p \leq .10$) – one on interactive skills and one on controlling skills. Again, only in one of the 10 equations is the absolute difference in the self versus others' evaluations of skills significant. In the 10 equations, seven of the changes in skills were significantly different – four interactive and three controlling. Two of the three significant coefficients for controlling skills were only at the marginal level ($p \leq .10$). The betas for the first three pairs of equations were higher for interactive than controlling skills and only in the fourth pairing was controlling higher than interactive skills. In the fifth pairing, neither change in skills coefficient was significant so they are not compared.

Insert Table 4 about here

Discussion

There can be little doubt that the effectiveness of the managers increase over the four iterations of the 360 feedback process in this study. This finding in the one-way analysis of variance on performance is consistent with findings on the use of the 360 feedback process in other studies (Rosti & Shipper, 1998; Shipper, 2009; Shipper & Neck 1990). Some reviews cited earlier have not been supportive of the effectiveness of “360 feedback,” but the studies in those reviews did not contain steps three through five of the “360 feedback process” outlined earlier. Whether or not those are included in the management development program may explain the difference in findings. Also, the changes appear to be meaningful in that managers who went through the program four times improved their effectiveness 10%.

The Duncan’s multiple range tests to conduct multiple comparisons did indicate that changes did occur on all three criterion variables, but not after a single intervention with one exception. These findings could be attributed to skepticism on the part of the early participants and lack of effective follow-up. Initially, the 360 process was inserted into a week-long training program as described earlier. At that point, the participants may have viewed it as a one-shot program and not part of an ongoing process that was going to be examined, supported and held accountable by a succession of CEO’s. As discussed earlier, both the way the training was delivered and the follow-up provided changed over the course of this study. Given these changes were based on observations and statistical testing, the follow-up activities may have been more important than the feedback. Other researchers have found also in control studies that follow-up interventions have greater impact than simply feedback (Rosti & Shipper, 1998).

The one exception in which change was found to occur after a single intervention was in the absolute values of self and others’ evaluations of interactive skills. This change may have

occurred because interactive skills are seen to be socially desirable, where controlling skills are not. In addition, it may take more time for changes in skills to impact changes in performance as evaluated by the managers' superiors than the 18 months between interventions. By the end of the study all three criterion variables in these analyses had made, however, significant changes in the expected directions. Expectations that a single intervention of the 360 feedback process will change established levels of performance or differences in evaluations of self and others' evaluations of skills may be unrealistic (Shipper, Hoffman, & Rotondo, 2007; Yukl & Lepsinger, 1995). Studies in the other fields such as neuropsychology (Law, 2004) and learning disabilities (e.g., Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich, 1996; Vickery, Reynolds, & Cochran, 1987) would suggest that changing in-grained patterns of behavior and reasoning requires intensive training. Management scholars who incorporate neuropsychology into their understanding of management development would also agree that intensive training is needed to changed long-term patterns of managerial behavior (Boyatzis, 1999; Goleman, Boyatzis, & McKee, 2002). These findings support those from reviews of 360 feedback that found little significant effects after a single intervention that simply provided feedback to participants (London & Smither, 1995; Smither, London & Reilly, 2005). If performance were so easily enhanced through such basic and short-term programs of management development, then becoming a high performing manager would not be so highly valued.

The results in Table 3 suggest that those who begin with low levels of interactive or controlling skills are more likely to make improvements than those who begin at high levels. In other words, the best predictor of future behavior is past behavior, and those with low initial levels of skills are more likely to make significant changes than those who begin with high

levels. Given that it can be shown mathematically that greater change has to occur to make the same percentage increase at low and moderate initial levels, the results are not surprising. The results also suggest that overestimating controlling skills initially is a predictor of increases in controlling skills, whereas neither overestimating nor underestimating interactive skills is a predictor of increase in interactive skills. These findings may be so because those managers who are willing to admit to the use with controlling skills are more willing to improve them than those who do not admit. The social stigma associated to controlling skills may further explain these findings.

The results from Table 4 suggest that those managers who begin with low levels of performance are more likely to make improvements than those who begin at high levels. In other words, similar to predicting behavior, the best predictor of future performance is past performance, and those with low initial levels of performance are more likely to make significant changes than those who begin with high levels. The findings in Table 4 also suggest that to improve performance managers should first work on their interactive skills and only after improving those skills should they work on improving controlling skills. This finding is in congruence with advice derived from Task Cycle Theory (Wilson, 2003), and with a study of skills of effective managers at different stages of their careers (Shipper & Dillard, 2000). The last pair of equations in Table 4 suggests that in the long-term, improvement in managerial performance is possible regardless of the value of the self versus others' evaluation of skills early in the process and that both over and under estimators will improve equally in the long run.

Conclusions and Implications for Practice

Although this study focused on examining the relationships among self versus others' evaluation of skills, under versus overestimating, and improvement in skills and performance,

the findings supported the use of a sustained 360 feedback process for effective management development (Shipper, 2009). This conclusion and implication are supported by the findings in other research (Seifert, Yukl, & McDonald, 2003). Overall, the results suggest that managers do care about how others evaluate their skills and that they do adjust their self-evaluations to reduce the differences while at the same time working to improve their skills. In addition, the findings suggest that managers who are evaluated to have low skills will improve their skills even more than people with moderate or high skills. This may be true only under conditions, as in this study, where the participating managers are given follow-up support.

These findings are in agreement with studies on what it takes to make meaningful changes in other fields of human behavior (e.g., Law, 2004; Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich, 1996; Vickery, Reynolds, & Cochran, 1987). To expect meaningful changes to occur simply from the process of feedback without providing additional assistance and a supportive environment is probably naïve at best.

There are also implications for executives wanting to improve the effectiveness of their managers. First, the chief executive officer's and other key executives' participation and active support are needed throughout the process. This cannot be demonstrated statistically, but it was one of the conditions that existed in the organization where this study occurred. Second, it is probably not going to happen based on one-shot participation in 360 feedback. In this application, the 360 feedback highlighted only what skills a manager needed to target for improvement based on the Managerial Task Cycle. The follow-up organizational support and efforts of the individual managers probably led to significant changes in effectiveness occurring. Third, creating a supportive environment and activities at the local level will be more effective than centralized activities. This principle was demonstrated when the organization switched from

management retreats to intact workshops and when managers were urged to work with their immediate subordinates to assist in improving. Fourth, meaningful change yields a competitive advantage. Improvements in the effectiveness of managers can yield improvements in productivity. Previously, this finding has been demonstrated monetarily (Shipper & Neck, 1994). Thus, a sustained 360 process can be a valuable business practice that leads to sustainable improvements in managerial effectiveness.

Limitations and Future Research

The results of this study should be considered in light of its limitations. There is the obvious limitation that this study was conducted in only one organization. The findings relative to the effectiveness of the 360 feedback process are supported in other pre-post studies (Rosti & Shipper, 1998; Shipper & Neck, 1990), but the findings on self-other agreement and over versus underestimating have not been studied widely if at all using pre-post data. Approximately 30% of the participants in this study are from outside the United States. Since cultures in other countries have been shown to have different impacts self and others' evaluations (Atwater, Wang, Smither, & Fleenor, 2009), the findings in this study may underestimate the impact that difference in self and others' evaluations of skills and changes in these skills may have on performance in the United States. Also, other demographics such as age and gender were not accounted for in this study. Thus, the impact that difference in self and others' evaluations of skills and changes in these skills may have on performance could be further underestimated (Vecchio & Anderson, 2009). Therefore, in future research on self-other agreement and over versus under estimation, culture and other demographic characteristics should be included in the study.

This study did not test for curvilinear relationships, although two studies have found such relationships between self-other agreement and over versus under estimation, and performance (Shipper, Kincaid, Rotondo, & Hoffman, 2003; Vecchio & Anderson, 2009). If the relationships are curvilinear, using linear regression as done in this study would cause the effect sizes to be underestimated. Thus, future research may want to control for demographic characteristics as well as test for curvilinear relationships when investigating the issues raised in this research.

Furthermore, this study did not test simultaneously the impact of self-other agreement and over versus under estimation of interactive and controlling skills on performance. The only study found in which this was done used structural equation modeling (Shipper & Davy, 2002). Future research on the issues addressed in this study may want to use structural equation modeling using cross sectional data to provide additional evidence of the hypothesized cause and effect relationships.

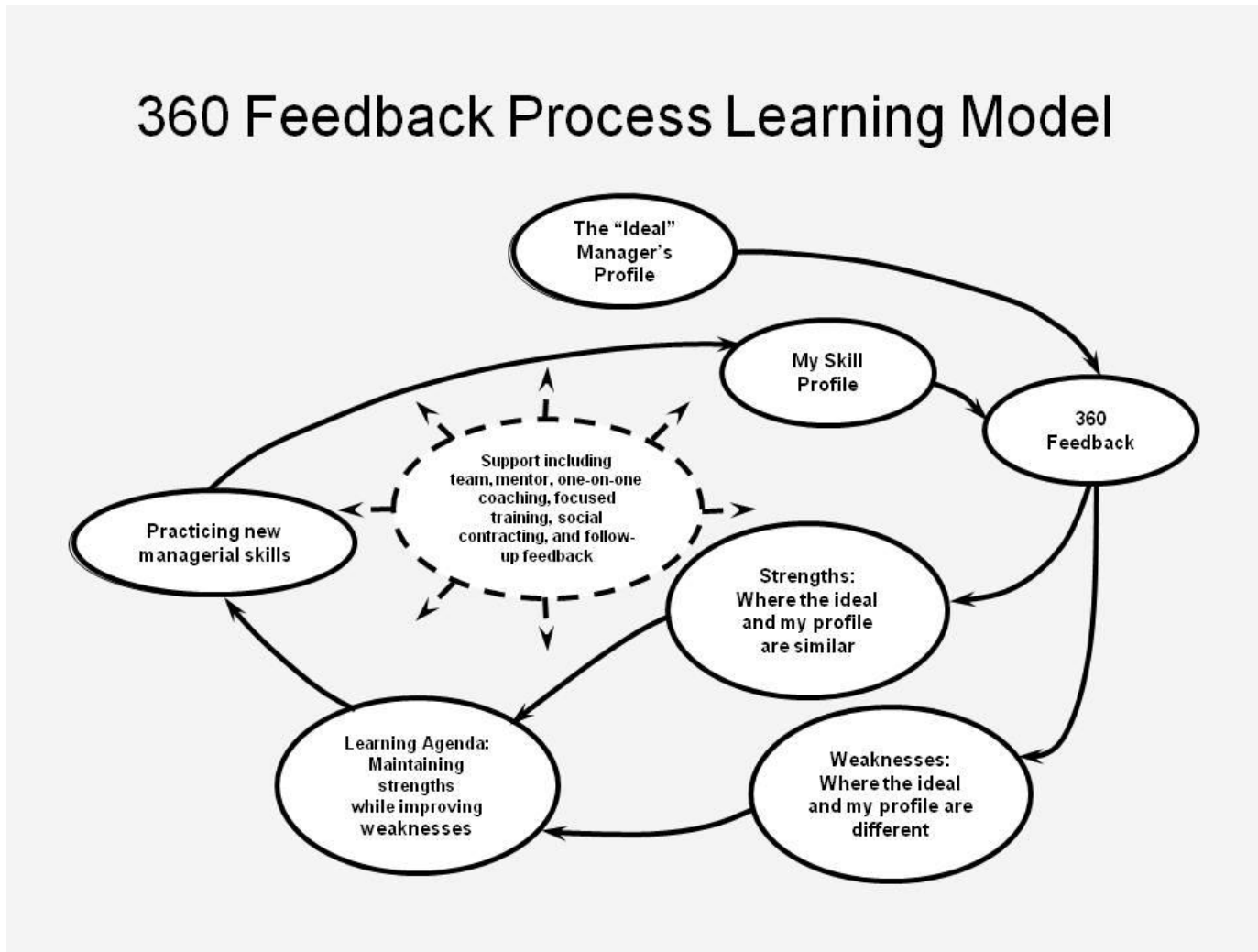
One limitation of this field study is that the assignment to groups was not randomized as is expected in true experimental designs (Campbell & Stanley, 1963). This condition is seldom met in long-term field studies because operating objectives of the organization could be disrupted by randomized assignment to experimental and control groups.

Another limitation of this study is that the experimental conditions changed overtime. As previously discussed, this research was carried out based on the action research model. Changes were made in the follow-up interventions based on observation and testing, as also previously discussed.

Another limitation that also needs to be recognized is that what occurred in this study may not be repeatable if the same instrument and supporting activities are not provided. First, the 360 instrument selected by the organization has been validated in multiple studies (e.g., Rosti &

Shipper, 1998; Shipper, 1995; Shipper & Davy, 2002; Shipper & White, 1999) and recognized as one of the better 360 instruments (Leslie & Fleenor, 1998; Morrison, McCall & DeVries, 1978; Van Velsor & Leslie, 1991). The validity and usefulness of feedback from some 360 instruments have to be questioned. In addition, a succession of corporate leaders was willing to invest their time and corporate resources into both the feedback process and the follow-up training over an extended period of time. Finally, the corporate culture may also partially account for the success of this 360 process in this study. It is one where openness, honesty, and constructive confrontation are explicitly recognized and reinforced. Future research that focuses on organizational variables such as leadership, and culture; and program characteristics such as instrumentation and follow-up support may lead to explaining why some 360 feedback processes seem to consistently yield results and others do not.

Figure 1



360 Feedback Process Learning model. Adapted from Boyatzis R., 1999. Self-Directed Change and Learning as a Necessary Meta-Competency for Success and Effectiveness in the 21st Century. In R. Sims & J. G. Veres, III (Eds.), *Keys to Employee Success in the Coming Decades*.

Figure 2

Improvement in Managerial Effectiveness Across Time

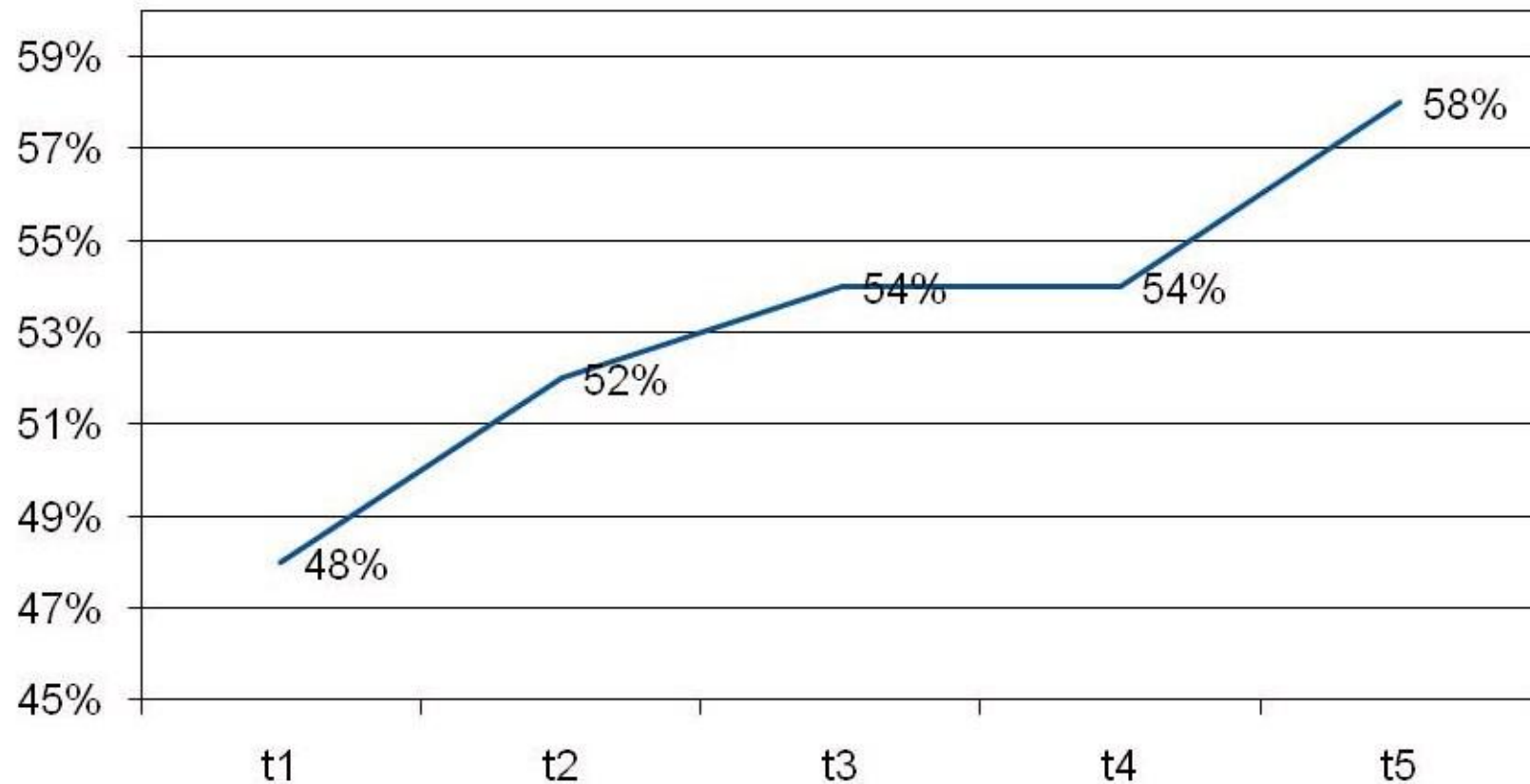


Table 1

Means, Standard Deviations, Cronbach's Alphas, and Correlations Among Linear Variables Used in the Study

Variables			Self Evaluation of Interactive Skills	Self Evaluation of Controlling Skills	Others' Evaluation of Interactive Skills	Others' Evaluation of Controlling Skills	Absolute Difference between Self and Others' Evaluation of Interactive Skills	Absolute Difference between Self and Others' Evaluation of Controlling Skills	Superior's Evaluation of Performance
Self Evaluation of Interactive Skills	Mean = -.0037 Std. Dev. = .81306	Pearson Correlation Sig. (2-tailed) N	$\alpha = .89$						
Self Evaluation of Controlling Skills	Mean = -.0002 Std. Dev. = .97290	Pearson Correlation Sig. (2-tailed) N	.484** .000 31536	$\alpha = .61$					
Others' Evaluation of Interactive Skills	Mean = .0464 Std. Dev. = .91205	Pearson Correlation Sig. (2-tailed) N	.190** .000 26541	.057** .00 0 26576	$\alpha = .94$				
Others' Evaluation of Controlling Skills	Mean = -.0129 Std. Dev. = .98797	Pearson Correlation Sig. (2-tailed) N	.129** .000 26583	.394** .000 26618	.444** .000 26549	$\alpha = .60$			
Absolute Difference between Self and Others' Evaluation of Interactive Skills	Mean = .8653 Std. Dev. = .67836	Pearson Correlation Sig. (2-tailed) N	-.041** .000 26541	.016** .009 26539	-.059** .000 26541	.053** .000 26509	N/A		
Absolute Difference between Self and Others' Evaluation of Controlling Skills	Mean = .8477 Std. Dev. = .66837	Pearson Correlation Sig. (2-tailed) N	.024** .000 26581	.052** .000 26618	-.010 .090 26544	.042** .000 26618	.342** .000 26507	N/A	
Superior's Evaluation of Performance	Mean = 5.3059 Std. Dev. = .81317	Pearson Correlation Sig. (2-tailed) N	.091** .000 22689	-.011 .090 22740	.175** .000 19045	.015* .044 19058	-.031** .000 19023	-.018* .015 19055	$\alpha = .90$

Notes: ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed); N/A Not Applicable.

Table 2

Overall & Duncan's One-Way Analysis of Variance across Time for Performance, and the Absolute Values of Self vs. Others' Evaluations of Interactive and Controlling Skills

Data Collection Points	Performance				Absolute Value of Self vs. Other Evaluations of Interactive Skills				Absolute Value of Self vs. Other Evaluations of Controlling Skills			
	n	Subset for alpha \leq 0.05			n	Subset for alpha \leq 0.05			n	Subset for alpha \leq 0.05		F
		1	2	3		1	2	3		1	2	
t ₁	13113	5.2677			15559			.8997	15629		.8700	
t ₂	5976	5.3373	5.3373		6894		.8384		6898	.8315	.8315	
t ₃	2434		5.3785		2730	.7924	.7924		2733	.7926		
t ₄	843		5.3838		929	.7580			929	.7869		
t ₅	270			5.4667	294	.7572			294	.7681		
	Sig.	.069	.254	1.000		.279	.130	1.000		.052	.199	
	Sum of Squares	df	Mean Square	F Sig.	Sum of Squares	df	Mean Square	F Sig.	Sum of Squares	df	Mean Square	F Sig.
Between	52.080	4	13.020	28.376	52.080	4	13.020	28.376	23.219	4	5.805	13.025
Within	12113.668	26401	.459	.000	12113.668	26401	.459	.000	11800.496	26478	.446	.000
Total	12165.748	26405			12165.748	26405			11823.715	26482		

Note: The group sizes are unequal. The harmonic mean of the group sizes is used. Harmonic means of the group sizes used are 901.813, 989.430, and 989.582, respectively.

Table 3

Changes in Skills as Associated with Initial Self-Other Evaluations of Skills Controlling for Over & Under Evaluation and Initial Level of Skill

		Beta Over vs. Under at t_n	Beta Skill Level at t_n	Beta of Absolute Value of Self-Other Evaluation at t_n	Overall R^2	Degrees of Freedom	Significance
Change t_1-t_2	Interactive $_{n+1}$.023*	-.615****	.001	.394	3,5369	.000
	Controlling $_{n+1}$.110****	-.531****	.017	.342	3,5406	.000
Change t_2-t_3	Interactive $_{n+1}$.017	-.602****	-.016	.372	3,2204	.000
	Controlling $_{n+1}$.095****	-.486****	.010	.284	3,2209	.000
Change t_3-t_4	Interactive $_{n+1}$.026	-.573****	-.017	.340	3,773	.000
	Controlling $_{n+1}$.111***	-.438****	-.015	.247	3,773	.000
Change t_4-t_5	Interactive $_{n+1}$.053	-.526****	.166***	.355	3,242	.000
	Controlling $_{n+1}$.047	-.515****	-.012	.288	3,242	.000
Change t_1-t_5	Interactive $_{n+1}$.075	-.676****	.045	.516	3,226	.000
	Controlling $_{n+1}$.145**	-.561****	.016	.392	3,227	.000

Note: * $p \leq .10$; ** $p \leq .05$; *** $p \leq .01$; **** $p \leq .001$

Table 4

Changes in Performance as Associated with Changes in Self-Other Evaluations of Interactive & Controlling Skills Controlling for Over & Under Evaluations and Initial Level of Performance

Performance	Skill	Beta Over vs. Under at t_n	Beta Skill Level at t_n	Beta of Absolute Value of Self-Other Evaluation at t_n	Beta of Change in Skill from t_n to t_{n+1}	Overall R^2	Degrees of Freedom	Significance
Change t_1 - t_2	Interactive $_{n+1}$	-.074*	-.620****	-.008	.115****	.384	4,2524	.000
	Controlling $_{n+1}$	-.018	-.622****	-.033**	.029*	.623	4,2534	.000
Change t_2 - t_3	Interactive $_{n+1}$	-.050*	-.595****	.010	.103****	.366	4,1416	.000
	Controlling $_{n+1}$	-.009	-.598****	-.014	.043*	.599	4,1418	.000
Change t_3 - t_4	Interactive $_{n+1}$	-.031	-.623****	.053	.071**	.400	4,558	.000
	Controlling $_{n+1}$.055	-.622****	.048	-.011	.397	4,558	.000
Change t_4 - t_5	Interactive $_{n+1}$	-.072	.071	-.057	.217**	.045	4,119	NS
	Controlling $_{n+1}$	-.144	.084****	.016	.274***	.080	4,119	.041
Change t_1 - t_5	Interactive $_{n+1}$.039	-.658****	-.032	-.002	.436	3,56	.000
	Controlling $_{n+1}$.053	-.660****	.041	-.033	.437	4,55	.000

Note: * $p \leq .10$; ** $p \leq .05$; *** $p \leq .01$; **** $p \leq .001$; NS Non-Significant

References

- Atwater, L., Ostroff, C., Yammarino, F., & Fleenor, J. 1998. Self-other agreement: Does it matter? *Personnel Psychology*, 51, 577–598.
- Atwater, L., Wang, M., Smither, J. W., & Fleenor, J. 2009. Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology*, 94(4): 876–886
- Atwater, L. E., & Yammarino, F. J. 1992. Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45(1): 141–164.
- Bass, B., & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review*, 40, 437–454.
- Bolles, R. C. 1972. Reinforcement, expectancy, and learning. *Psychological Review*, 79(5): 394–409.
- Boyatzis, R. 1999. Self-directed change and learning as a necessary meta-competency for success and effectiveness in the 21st century. In R. Sims & J. G. Veres, III (Eds.), *Keys to employee success in the coming decades*. Westport, CT: Greenwood Publishing.
- Campbell, D. T. & Stanley, J. C. 1963. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Clark, K. E., Clark, M. B., & Campbell, D. P., 1992. *Impact of leadership*. Greensboro, NC: The Center for Creative Leadership.
- Conway, J. M., & Huffcutt, A. I. 1997. Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10, 331–360.
- Furnham, A., & Stringfield, P. 1994. Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology*, 67, 57–67.
- Goleman, D., Boyatzis, R., & McKee, A. 2002. *Primal leadership: Realizing the power of emotional intelligence*. Cambridge, MA: Harvard Business School Press.
- Harris, M., & Schaubroeck, J. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43–61.
- Heidemeier, H., & Moser, K. 2009 Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, 94(2):353-370.

- Kirkpatrick, D. L. 1959. Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13(6): 3–9, 21–26.
- Kruger, J., & Dunning, D. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.
- Law, B. M., 2004. Training the brain to fix itself. *Monitor on Psychology*, 35(5): 36-37
- Leslie, J. B., & Fleenor, J. W., 1998. *Feedback to managers: A review and comparison of multi-rater instruments for management development* (3rd ed.). Greensboro, NC: Center for Creative Leadership.
- Lombardo, M. M. & McCauley, C. D. 1988. The dynamics of management derailment. Technical Report No. 34. Greensboro, NC: Center for Creative Leadership.
- London, M. & Smither, J. W. 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(4): 803-839.
- Mabe, P., & West, S. 1982. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67, 280–296.
- McCall, M. W., Jr. & Lombardo, M. M. 1983. Off the track: Why and how successful executives get derailed. Technical Report No. 21. Greensboro, NC: Center for Creative Leadership.
- Morrison, A. M., & McCall, J., & DeVries, D. L. 1978. *Feedback to managers: A comprehensive review of twenty-four instruments*. Greensboro, NC: Center for Creative Leadership.
- Nunnally, J. C. 1978. *Psychometric theory*. New York: McGraw-Hill.
- O'Brien, Robert M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality and Quantity*, 41(5)673-690.
- Resick, C. J., Whitman, D. S., Weingarden, S. M., & Hiller, N. J. The bright-side and the dark-side of personality: Examining core self-evaluations, narcissism, transformational leadership, and strategic influence. *Journal of Applied Psychology*, 94(6): 1365-1381.
- Rosti, Jr., R. T., & Shipper, F. 1998. A Study of the Impact of Training in a Management Development Program Based on 360 Feedback. *Journal of Managerial Psychology*, 13(1): 77-89.
- Seifert, C. F., Yukl, G., & McDonald, R. A. 2003. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*, 88(3): 561-569.

- Shipper, F. 1995. A study of the psychometric properties of the managerial skill scales of the survey of management practices. *Educational and Psychological Measurement*, 55(3): 468–479.
- Shipper, F. 2004. A cross-cultural, multi-dimensional, nonlinear examination of managerial skills and effectiveness. *Journal of Leadership and Organizational Studies*, 10(3): 89–102.
- Shipper, F. 2009. Investigating the sustainability of a sustained 360 process. **The Best Papers Proceedings of the Academy of Management Annual Meeting.**
- Shipper, F., & Davy, J. 2002. A model and investigation of managerial skills, employees' attitudes and managerial performance. *Leadership Quarterly*, 13(2): 95–120.
- Shipper, F., & Dillard, J. E. 2000. A study of impending derailment and recovery among middle managers across career stages. *Human Resource Management Journal*, 39(4): 331–345.
- Shipper, F., Hoffman, R. C., IV, & Rotondo, D. M. 2007. Does the 360 feedback process create actionable knowledge equally across cultures? *Academy of Management Learning & Education*, 6(1): 33–50.
- Shipper, F., Kincaid, J. F., Rotondo, D. M., & Hoffman, R. C., IV. 2003. A cross cultural exploratory study of the linkage between emotional intelligence and managerial effectiveness. *The International Journal of Organizational Analysis*, 11(3): 171–191.
- Shipper, F., & Neck, C. P. 1990. Subordinates' observations: Feedback for management development. *Human Resource Development Quarterly*, 1(4): 371-384.
- Shipper, F., & White, C. S. 1999. Mastery, frequency & interaction of managerial behaviors relative to subunit effectiveness. *Human Relations*, 52(1): 49–66.
- Shipper, F., & Wilson, C. L. 1992. The impact of managerial behaviors on group performance, stress and commitment. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), *Impact of leadership*: 119–129. Greensboro, NC: The Center for Creative Leadership.
- Smither, R. D., Houston, J. M., & McIntire, S. A. 1996. *Organization development: Strategies for changing environments*. New York: Harper Collins College Publishers.
- Smither, J. W., London, M., & Reilly, R. R., 2005. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*. 58(1): 33-66.

- Smither, J. W., London, M., Vasilopoulos, N. L., Reilly, R. R., Millsap, R. E., & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, 48, 1–34.
- Tallal, P., Miller, S. L., Bedi, G., Byrna, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271 (5245): 81 – 84.
- Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. 2005. A metaanalytical review of behavior modeling training. *Journal of Applied Psychology*, 94(4): 692–709.
- Tolman, E. C. 1932. *Purposive behavior in animals and men*. New York: Appleton Century Crofts.
- Vickery, K. S., Reynolds, V. A., & Cochran, S.W, 1987. Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia*, 37(1): 189-200.
- Van Velsor, E., & Leslie, J. B. 1991. *Feedback to managers, Volume II: A review and comparison of sixteen multi-rater feedback instruments*. Greensboro, NC: Center for Creative Leadership.
- Vecchio, R. P., & Anderson, R. J. 2009. Agreement in self-other ratings of leader effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment*. 17(2): 165-179.
- Wilson, C. L. 1975. Multi-level management surveys: Feasibility studies and initial applications. *JSAS: Catalog of Selected Documents in Psychology*, Vol. 5 (Ms. No. 1137). Washington: American Psychology Association.
- Wilson, C. L. 1978. The Wilson multilevel management surveys: Refinement and replication of the scales. *JSAS: Catalog of Selected Documents in Psychology*, Vol. 8 (Ms. No. 1707). Washington: American Psychology Association.
- Wilson, C. L. 2003. **How and Why Effective Managers Balance Their Skills**. Silver Spring, MD: The Clark Wilson Group.
- Wilson, C. L., O'Hare, D., & Shipper, F. 1990. Task cycle theory: The process of influence. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership*: 185–204. West Orange, NJ: Leadership Library of America, Inc.
- Wilson, C. L., & Shipper, F. 1992. *Task cycle management: A competency-based course for operating managers* (Rev. ed.). Silver Spring, MD: The Clark Wilson Group.

Wilson, C. L., & Wilson, J. L. 1991. *Teams and leaders: A manual for the Clark Wilson Publishing Company training and development programs*. Silver Spring, MD: Clark Wilson Publishing Company.

Yukl, G & Lepsinger, R. 1995. How to get the most out of 360-degree feedback. *Training*, 32(12): 45-50.