Attitudes, Self-Monitoring, and Appraisal Behaviors

I. M. Jawahar
Illinois State University

This study examined the influence of attitudes and self-monitoring on leniency (elevation accuracy) of performance ratings and personnel decisions. In addition, moderating effects of self-monitoring on the relationship between attitudes and accuracy of ratings and decisions were investigated. Attitudes and self-monitoring tendency of 210 managers–professionals were measured, and ratings provided and decisions made by them were used to test 3 sets of hypotheses. Moderated regression and follow-up split-group analyses indicated that self-monitoring moderated the relationship between attitudes toward accurate appraisal and rating accuracy. Self-monitoring significantly influenced rating and decision accuracy such that accuracy declined with increasing level of self-monitoring. Results highlight the influence of rater’s personality on appraisal behaviors. Implications of results and directions for future research are discussed.

A major shortcoming of most appraisal systems seems to be the failure of raters to be critical in their evaluations of ratees. Indeed, many surveys and interview-based studies indicate leniency in rating to be the norm rather than the exception in most organizations (e.g., Bernardin & Villanova, 1986; Bretz, Milkovich, & Read, 1992; Longenecker, Jackson-Jaccoud, Sims, & Gioia, 1992). The pervasiveness of leniency has prompted researchers to devote substantial attention to understanding the causes of leniency. Over the past 50 years, researchers have investigated three potential causes of leniency: rating formats, raters’ ability and cognitive processes, and raters’ motivation. In recent years, factors with potential to influence raters’ motivation to provide accurate ratings have received increased attention (Harris, 1994; Murphy & Cleveland, 1995). For instance, several researchers have proposed that raters’ beliefs and attitudes (e.g., Murphy & Cleveland, 1995; Tziner, Murphy, Cleveland, Beaudin, & Marchand, 1998) are likely to influence raters’ motivation to provide accurate ratings.

Recent models of the performance appraisal process (e.g., Cleveland & Murphy, 1992; Murphy & Cleveland, 1991, 1995) have emphasized the importance of studying raters’ attitudes as a key to understanding rating behavior. For instance, Murphy and Cleveland (1995) have suggested that raters’ attitudes and beliefs that are immediately relevant to the task of evaluating performance as well as more general beliefs and attitudes about the organization could substantially influence rating behaviors. Raters’ level of trust in the appraisal process (Bernardin & Orban, 1990), perceptions of consequences (Napier & Latham, 1986), and perceptions of the prevalence of political considerations in performance appraisal (Tziner, 1999) could all affect attitudes toward the task of evaluating performance. Raters’ general beliefs and attitudes about the organization could affect raters’ commitment to the organization (Mayer & Schoorman, 1992), which in turn could encourage raters to put forth considerable effort for the benefit of the organization, including effectively evaluating performance (see Tziner et al., 1998). For these reasons, raters who have favorable attitudes toward performance appraisal and the organization could be expected to invest more effort in and be more willing to do an effective job of evaluating performance than raters with less favorable attitudes. Thus, attitudes toward performance appraisal and toward the organization have the potential to either increase or decrease raters’ willingness or motivation to accurately evaluate the performance of others (see Tziner & Murphy, 1999).

In spite of such a compelling rationale, very few studies have examined the influence of attitudinal factors on the quality of performance ratings (e.g., Tziner & Murphy, 1999). One objective of the present study was to build on this stream of research and investigate the effects of raters’ attitudes on the leniency of performance ratings and decisions. Even though almost half a century ago Guilford (1954) asserted that leniency is a stable characteristic of raters and noted that it can be predicted from measures of individual differences, only a handful of studies (e.g., Kane, Bernardin, Villanova, & Peyrefitte, 1995) have directly tested Guilford’s assertion. If, in fact, some raters are predisposed to be lenient, then it is important to identify the specific predisposition because the inability to discriminate inaccurate from accurate raters has been noted as the most significant barrier to enhancing the accuracy of performance ratings (e.g., Murphy & Cleveland, 1991, 1995). Self-monitoring is a stable dispositional characteristic that is also related to individuals’ tendencies to act according to their attitudes (Snyder, 1974, 1979). Therefore, a second objective of this study was to investigate the influence of self-monitoring on leniency of ratings and decisions. On the basis of prior research on self-monitoring conducted in the context of the attitude–behavior relationship (e.g., Ajzen, Timko, & White, 1982), a third objective of this study was to investigate whether self-monitoring moderates the relationship between attitudes and leniency of ratings and decisions.

Attitudes Toward Accurate Appraisal

Researchers have investigated and documented a relationship between attitudes and behaviors in many areas within the disci-
pline of organizational behavior. For instance, attitudes such as job satisfaction have been used to predict diverse behaviors, including absenteeism (Zaccaro, Craig, & Quinn, 1991) and turnover (George & Jones, 1996). In performance appraisal research, however, the attempt to link attitudes and appraisal behaviors is still new and novel.

Only recently, a few studies have investigated the influence of raters’ beliefs and attitudes on the quality of ratings. For instance, in a field study, Bernardin and Orban (1990) reported that raters’ “trust in the appraisal process” was inversely related to the level of performance ratings. In another study, Villanova, Bernardin, Dahnmus, and Sims (1993) found that raters who reported high levels of appraisal discomfort were more likely to provide lenient ratings than raters who reported low levels of appraisal discomfort.

Tziner et al. (1998) examined the influence of proximal and distal factors on three measures of rating quality. Proximal factors studied included raters’ trust in the appraisal process, rating efficacy, feedback discomfort, and outcome expectancies. Of those factors, only trust in the appraisal process and outcome expectancies were significantly related ($p < .05$) to the overall level of ratings. Organizational commitment, a general attitude, was also positively correlated with overall level of ratings ($r = .22, p < .10$). Using data collected from 29 managers, Tziner and Murphy (1999) recently reported that instrumental organizational commitment was positively related to the overall level of ratings ($r = .29, p < .05$). Although the correlation between affective organizational commitment and overall level of ratings was positive, it failed to reach statistical significance ($r = .22, ns$).

Of the studies just reviewed, only two studies, both by Tziner and his colleagues (Tziner & Murphy, 1999; Tziner et al., 1998) measured “attitudes” per se. Even in those two studies, general attitudes, instrumental commitment and affective commitment, were used to predict the level of ratings, a specific outcome. From a methodological standpoint, Ajzen and his colleagues (Ajzen, 1991; Ajzen & Fishbein, 1977, 1980) argued that the relationship between attitudes and behavior could be enhanced by improving the correspondence, in terms of specificity, between the measurement of attitudes and behaviors or outcomes. In general, Ajzen and his colleagues noted that behaviors can be best predicted by attitudes that specifically relate to those behaviors rather than more global and general attitudes. Substantial evidence for the predictive validity of attitudes has been found under conditions of methodological correspondence between attitudinal and behavioral measures (Jawahar & Elango, 1998).

Relative to general attitudes such as affective commitment, raters’ attitudes toward accurate appraisal match raters’ behavior of rating accurately or inaccurately in terms of specificity and, therefore, should be strongly related. Although the purpose of the present study was also to investigate the effects of attitudes, unlike previous studies, in this study attitudes that specifically relate to appraisal behaviors were measured. Two important appraisal behaviors are evaluating performance and making decisions about pay increases. Therefore, this study investigated the relationship between attitudes toward accurate appraisal and the accuracy of ratings and decisions.

Hypothesis 1a: Attitudes toward accurate appraisal will be positively related to the accuracy of ratings.

Hypothesis 1b: Attitudes toward accurate appraisal will be positively related to the accuracy of decisions.

In addition to attitudes toward accurate appraisal, the general attitudes of instrumental and affective commitment were also measured to replicate the results of Tziner and Murphy (1999). I expected that the relationship between specific attitudes and ratings (and decisions) would be stronger than the relationship between the general attitudes and ratings (and decisions).

Self-Monitoring

A few studies have provided evidence to support Guilford’s (1954) assertion that leniency is a stable characteristic of raters (e.g., Borman & Hallam, 1991; Villanova et al., 1993). For instance, in one study, Kane et al. (1995) investigated the stability of raters’ leniency in three actual appraisal situations and found leniency to be a relatively stable response tendency across the three situations. On the basis of their results, Kane et al. also recommended the use of measures of individual differences to predict leniency of performance ratings and related decisions.

One individual difference that not only has the potential to influence leniency but also is related to an individual’s tendency to act in accordance with his or her attitudes is self-monitoring. According to Snyder (1979, 1987), people differ in the extent to which their behavior is susceptible to situational or interpersonal cues, as opposed to inner states or dispositions. Self-monitoring theory and research suggest that high self-monitors are adept at deciphering cues in the social environment and are capable of tailoring their behaviors to fit the social context. In contrast, the behaviors of low self-monitors reflect their feelings and attitudes without regard to the situational or interpersonal consequences of those behaviors (Ajzen et al., 1982; Snyder, 1979).

In one study, for instance, anticipated future interaction induced the “situationally guided” high self-monitors to become even more attentive to situational cues when deciding how to act while prompting low self-monitors to rely even less on situational cues and more on personal thoughts and evaluations (Shaffer, Ogden, & Wu, 1987). In another study, White and Gerstein (1987) found that when social rewards were contingent on helping others, high self-monitors, in contrast to low self-monitors, were more likely to help others, but when consequences were nonexistent, high self-monitors were less likely to help. Whereas high and low self-monitors are equally aware of their attitudes’ action implications, in contrast to low self-monitors, high self-monitors seem to be easily deflected by situational factors (Ajzen et al., 1982). Indeed, research on self-monitoring conducted with college students as well as working adults suggests that, in contrast to low self-monitors, high self-monitors behave in anticipation of consequences of those behaviors (see Caldwell & O’Reilly, 1982a, 1982b; Fandt & Ferris, 1990; Jawahar & Stone, 1997; Snyder, 1979, 1987).

In organizations, important administrative decisions regarding pay increases, promotions, and retention are to a large extent determined on the basis of performance ratings. Ratings obtained to make such administrative decisions have consequences for the ratee (e.g., pay increase or no pay increase), and hence the rater-ratee relationship. Theory and research on self-monitoring suggest that, in contrast to low self-monitors, high self-monitors choose
behaviors most likely to maximize approval and minimize disapproval of others (Jawahar & Stone, 1997; White & Gerstein, 1987). When important consequences are contingent on performance ratings, lenient ratings as opposed to ratings that are accurate or stringent are likely to elicit the approval of ratees (see Harris, 1994; Murphy & Cleveland, 1995). Therefore, in a performance appraisal situation, high self-monitors, in contrast to low self-monitors, may be expected to provide lenient ratings. Using data collected from 29 managers, Tziner and Murphy (1999) reported that self-monitoring was negatively related to the level of ratings, but this relationship was not statistically significant ($r = -.15, ns$), perhaps due to the small sample. In Jawahar and Stone’s study, ratings of high self-monitors were significantly more lenient than ratings of low self-monitors. However, in Jawahar and Stone’s study, undergraduate students with little practical experience served as raters. The present study overcame the limitations of these two studies that had previously investigated the relationship between self-monitoring and performance ratings.

Actual decisions such as pay increase decisions have even more severe consequences for the rater–ratee relationship than performance ratings because sometimes performance ratings may or may not be closely linked to the size of one's pay increase. Consequently, pay increase decisions of high self-monitors may be expected to be more lenient than pay increase decisions of low self-monitors.

**Hypothesis 2a:** Self-monitoring will be negatively related to the accuracy of ratings.

**Hypothesis 2b:** Self-monitoring will be negatively related to the accuracy of decisions.

Because high self-monitors are easily deflected by situational influences and behave in anticipation of consequences, their behaviors are likely to be inconsistent with their previous attitudes. In contrast, low self-monitors, whose behaviors reflect their feelings and attitudes without regard to situational or interpersonal consequences, could be expected to exhibit substantial attitude–behavior correspondence. Previous research has provided empirical support for these predictions (e.g., Ajzen et al., 1982). In one study, for instance, Ajzen et al. found that the relationship between attitudes and behavior was significantly stronger for low self-monitors than for high self-monitors. In a more recent study, Dobbins, Farh, and Werbel (1993) found that inflated self-reports of grade point averages could be better predicted by attitudes toward inflation of low self-monitors than by attitudes of high self-monitors. These studies provide the basis for expecting the relationship between attitudes toward accurate appraisal and accuracy of ratings and decisions to be stronger for low self-monitors than for high self-monitors.

**Hypothesis 3a:** Self-monitoring will moderate the relationship between attitudes toward accurate appraisal and accuracy of ratings such that the relationship will be stronger for low self-monitors than for high self-monitors.

**Hypothesis 3b:** Self-monitoring will moderate the relationship between attitudes toward accurate appraisal and accuracy of decisions such that the relationship will be stronger for low self-monitors than for high self-monitors.

**Method**

**Subjects**

Data for this study were collected at two state universities located in the midsection of the United States. Subjects were full-time employees who were also part-time graduate students. Potential subjects were informed that the purpose of the study was to investigate the effectiveness of a new method of performance appraisal, the journal entry (i.e., diary-keeping) method, and were requested to participate in the research study. At University A, the initial sample consisted of 89 subjects, but because of incomplete responses, the sample was reduced to 74. At University B, although the initial sample consisted of 154 subjects, only 136 subjects provided usable data. Because the subjects from these two universities did not differ on the independent and dependent variables and were comparable in terms of type of employment and experience with performance evaluations, the two samples were combined to increase the number of subjects. Thus, the final sample consisted of 210 subjects. These subjects were employed in either professional or managerial positions in diverse settings including the government, manufacturing, and service sectors. The average age of the subjects was 32.6 years, and the average number of years of full-time employment was 8.2. All subjects had experience with performance appraisals as raters. Of the 210 subjects, 82 were female and 128 were male.

**Study Design and Procedure**

This study was conducted in two phases separated by 6 weeks. In Phase 1, subjects completed a scale developed to measure attitudes toward accurate appraisal and the self-monitoring scale constructed by Snyder and Gangestad (1986). A subset of subjects, subjects from University B, also completed Mayer and Schoorman's (1992) scales of instrumental organizational commitment and affective organizational commitment.

In the second phase of the study, subjects were mailed an information packet that contained a letter, a scenario, performance stimuli, and a rating form. The scenario contained a brief description of SM, Inc., a fictitious mail-order company specializing in a wide range of outdoor products, and a job description of sales representatives. Performance stimuli were presented to subjects in the form of critical incidents (journal entries). For instance, one critical incident read "made a recommendation about adding Spencer fishing poles because of numerous customer suggestions," and another read “lost temper when dealing with an upset customer.” Twenty-five such incidents captured the performance of each of two subordinates, Pat and Chris. In addition, the order in which critical incidents capturing the performance of Pat and Chris were presented was randomized. Although performance information of two subordinates, Pat and Chris, was provided, all raters were instructed to rate the performance of Pat only. Performance information on Chris was provided so that raters would have a standard for evaluating Pat. Also, performance information was rigged so that Chris was clearly the better performer of the two. Pat’s performance relative to that of Chris was portrayed as poor so as to avoid a ceiling effect and allow room for raters to inflate ratings.

The letter stated that “pay increases at SM, Inc. are based strictly on performance and because of unusually high profits SM, Inc. has plenty of funds/money for pay increases.” The letter also instructed the subjects to first familiarize themselves with the scenario, the performance appraisal form, and the critical incidents and then evaluate the performance of Pat on a behaviorally anchored rating scale. This scale was developed from the job description of raters (sales representative) to measure performance on the five dimensions of interpersonal and communication skills, dependability, quality of work, knowledge of company products and sales procedures, and initiative. After evaluating Pat’s performance, subjects were instructed to make a pay increase decision for Pat, furnish demographic information, and answer some additional questions. Subjects were requested to directly
mail the completed surveys to the researcher in the stamped and addressed envelope and were thanked for their participation.

Independent Measures

In Phase 1 of the study, subjects' attitudes and self-monitoring tendency were assessed. Subjects completed Gangestad and Snyder's (1985) 18-item self-monitoring scale. Sample items include "I find it hard to imitate the behavior of other people" and "In different situations and with different people, I often act like very different persons." The reliability of the scale in this study was higher (α = .79) than that (α = .70) reported by Gangestad and Snyder (1985). Although a detailed review of the literature is beyond the scope of this article, note that there has been substantial research in support of the construct validity of self-monitoring and the self-monitoring scale (see, e.g., Gangestad & Snyder, 1985, 1991; Snyder, 1979, 1987). Consistent with past research on self-monitoring (see Larkin & Pines, 1994), subjects with self-monitoring scores greater than or equal to 10 were regarded as high self-monitors, and those with scores less than or equal to 9 were regarded as low self-monitors. The self-monitoring scale has a classification accuracy (i.e., accuracy of correctly identifying an individual as a low or a high self-monitor) of 87% (Gangestad & Snyder, 1985). Furthermore, consistent with prior research (e.g., Gangestad & Snyder, 1985; Snyder, 1987), the distribution of self-monitoring scores in this study was bimodal with peaks at scores of 8 (low self-monitors) and 11 (high self-monitors).

On the basis of a review of the literature on "performance appraisal purpose" (see Cleveland, Murphy, & Williams, 1989; Jawahar & Williams, 1997) and two pilot tests, 10 items were developed to measure attitudes toward accurate appraisal. These items were administered to 292 students enrolled in introductory management courses, and these data were used for conducting reliability analyses. One of the items had an item-total correlation of less than .30, and the reliability analyses indicated that dropping this item would increase the reliability of the scale to an alpha of .76. Consequently, the item was dropped from the scale. The final attitudes toward accurate appraisal scale consisted of 9 items with scale points of 1 (completely disagree), 2 (strongly disagree), 3 (disagree), 4 (neither disagree nor agree), 5 (agree), 6 (strongly agree), and 7 (completely agree). Sample items include "Performance ratings given to an employee should accurately reflect employee's level of performance," "Personal biases or preferences of the supervisor should not influence performance ratings he/she may give to an employee," and "Feedback given to an employee regarding his or her performance should be as accurate as possible." Reliability of the attitudes toward accurate appraisal scale estimated with data collected for this study was .77.

The same scales used by Tziner and Murphy (1999) to measure instrumental and organizational commitment were used in this study. Instrumental organizational commitment was measured with an 8-item scale designed by Mayer and Schoorman (1992). Sample items include "The longer I stay with this organization, the harder it is to leave" and "It would be hard on my family if I decided to leave this organization at this time." Reliability of the instrumental commitment scale was .64. Although a reliability of .64 is lower than desirable (Nunnally & Bernstein, 1994), it is comparable to the reliability of .53 reported by Tziner and Murphy for this scale. Affective organizational commitment was measured with a 9-item scale developed by Mayer and Schoorman. Sample items include "I am proud to tell others that I am part of this organization" and "I talk up this organization to my friends as a good organization to work for." Reliability of the affective commitment scale was .73.

Dependent Measures

True scores. The most widely used procedure for computing true score estimates was developed by Borman (1977). In brief, this procedure involves the use of multiple experts who evaluate performance under optimal conditions. If these true score measures are collected, it becomes possible to assess a subject's accuracy in rating a ratee's performance on several dimensions by comparing the ratings with these true scores. Using expert ratings as true scores is widely accepted in the appraisal literature (see Jawahar & Stone, 1997; McIntyre, Smith, & Hassett, 1984; Murphy & Cleveland, 1991).

In this study, two management faculty members with a combined experience of more than 34 years in teaching, research, and consulting related to performance appraisal served as expert raters. These expert raters were familiarized with ratee's job description, exhibited job behaviors, the nature and contents of performance dimensions, and the performance appraisal instrument. A written copy of this information was also provided, and the experts were encouraged to refer to this material when evaluating performance. Agreement between raters (reliability) was very high (r = .91). Consistent with prior studies, mean expert ratings generated through this procedure were used as true score measures of performance.

Performance ratings. Given the hypotheses tested in this study, leniency or elevation accuracy is the most appropriate measure of accuracy. Leniency is defined as a rater's tendency to assign ratings that are higher than the true scores. Ratings provided by subjects were manipulated using the formula presented in McIntyre et al. (1984) to compute leniency. Basically, the rating obtained on each performance dimension is subtracted from the corresponding true score, and this difference is divided by the number of performance dimensions. This procedure is repeated for each performance dimension, and the resulting scores are summed across dimensions to obtain an overall estimate of leniency for each rater. In general, leniency and accuracy are distinct concepts (Murphy & Balzer, 1989). However, because leniency computed using McIntyre et al.'s formula also captures elevation accuracy (Cronbach, 1955; see Sulsky & Balzer, 1988, p. 500), and because both refer to the extent to which ratings provided by raters are higher or more elevated than true scores, lenient ratings are indicative of inaccuracy. Higher negative values indicate higher levels of leniency and lower levels of accuracy. For instance, −1.2 would indicate the ratings to be more lenient and less accurate than −0.4, and a score of 0 would indicate absence of leniency.

Pay increase decision. Besides evaluating Pat's performance, subjects made a pay increase decision for Pat by using a 9-point scale ranging from 1 (strongly oppose [pay increase]) to 9 (strongly support). Using McIntyre et al.'s (1984) formula and the approach described earlier to generate true scores, subjects' pay increase decisions were manipulated to compute leniency—elevation accuracy of those decisions.

Other measures. Both perceptions of consequences and the extent to which subjects considered those consequences when evaluating performance and making pay increase decisions were measured. Perceptions of consequences were measured with two items: "The performance ratings that you just provided will significantly affect your subordinate" and "The performance ratings you provided will not have any consequences for your subordinate." Both items were rated on a 9-point scale ranging from 1 (strongly disagree) to 9 (strongly agree). Two items, "To what extent did you consider consequences of performance ratings while evaluating ratee's performance?" and "To what extent did you consider how your pay increase decision would affect ratee?" were used to measure the extent to which subjects considered consequences. Both items were rated on a 5-point scale ranging from 1 (did not consider at all) to 5 (considered it a great deal).

Results

Manipulation Checks

In this study, performance was evaluated for pay increase purposes only, and all subjects correctly identified the purpose for which they evaluated performance. As stated in the Study Design and Procedure section, subjects were led to believe that SM, Inc.,
had plenty of funds—money for pay increases. Two items were used to check subjects’ perceptions about the availability of funds for pay increases. The items were “Plenty of funds are available for providing pay increases” and “Due to lack of funds, pay increases cannot be provided.” Subjects used a 9-point scale ranging from 1 (strongly disagree) to 9 (strongly agree) to respond to these two items. Subjects’ responses to the second item were reverse-scored and then averaged with their responses to the first item. Subjects’ responses (M = 8.10, SD = 1.37) indicated that they in fact believed that funds for pay increases were plentiful. In addition, the performance information presented to subjects was rigged so that Chris was a better performer than Pat. This manipulation was also effective. Subjects’ evaluation of Chris was much higher than that of Pat. Subjects evaluated overall performance on a 7-point scale ranging from 1 (very poor) to 7 (very good). As expected, Chris’s performance (M = 4.80) was rated significantly higher, t(135) = 11.43, p < .001, than Pat’s performance (M = 3.21). In addition, two items, “Chris is a better performer than Pat” and “Pat is a better performer than Chris,” were used to assess subjects’ views of the performance levels of Pat and Chris. Subjects used a 9-point scale ranging from 1 (strongly disagree) to 9 (strongly agree) to respond to these two items. Subjects’ responses to the item “Pat is a better performer than Chris” were reverse-scored and then averaged with their responses to the item “Chris is a better performer than Pat.” A mean of 7.79 indicates that subjects perceived the performance of Chris to be much higher than that of Pat. Means and standard deviations of attitudinal measures, accuracy of ratings, and pay increase decisions for low and high self-monitors are presented in Table 1.

**Hypotheses Testing**

Hypothesis 1 was not supported because attitudes toward accurate appraisal were not related to either accuracy of ratings (r = .09, ns; β = .087, ns) or decisions (r = −.02, ns; β = −.026, ns). Instrumental organizational commitment was also unrelated to either ratings (r = .03, ns; β = −.016, ns) or decisions (r = .09, ns; β = .091, ns). Likewise, affective organizational commitment was not related to either ratings (r = −.05, ns; β = −.052, ns) or decisions (r = .04, ns; β = .041, ns). Because none of the attitudes were significantly related to either ratings or decisions, the expectation that specific attitudes would be more predictive than general attitudes was not upheld.

Hypothesis 2 was fully supported because self-monitoring was negatively related to accuracy of ratings (r = −.72, p < .001; β = −.720, p < .001) and decisions (r = −.63, p < .001; β = −.635, p < .01). Subjects were classified as high or low self-monitors (see the Independent Measures section). Although low and high self-monitors did not differ in their attitudes toward accurate appraisal (M = 5.99 vs. M = 5.94, respectively), t(208) = 0.41, ns, ratings of high self-monitors (M = −1.37) were significantly more lenient and inaccurate (p < .001) than those of low self-monitors (M = −0.37). In a similar manner, decisions of high self-monitors (M = −2.86) were significantly more lenient and inaccurate (p < .001) than those of low self-monitors (M = −0.23).

To test Hypotheses 3a and 3b, moderated regression analyses were conducted to determine if self-monitoring moderated the attitude–behavior relationship. Following the procedure articulated by Cohen and Cohen (1983), the dependent variable was regressed on attitudes and self-monitoring. Next, the cross-product vector of attitudes and self-monitoring was computed and added to the equations. A significant beta weight for the interaction term would indicate that self-monitoring moderated the relationship between attitudes and behavior. A negative regression coefficient for the interaction term would indicate that the relationship between attitude and behavior is stronger at lower levels of the moderator than at higher levels of the moderator.

Hypothesis 3a predicted that self-monitoring would moderate the relationship between attitudes and accuracy of ratings such that the relationship would be stronger for low than for high self-monitors. Because the correlation between the cross-product (of attitudes and self-monitoring) and self-monitoring was unusually high (r = .90, p < .001), the centering technique was used to properly test the interaction term. Centering a variable involves simply replacing it by its deviation from the mean. The variables, attitudes and self-monitoring, were centered; the product of the two centered variables was used as the new cross-product term; and the moderated regression analyses were repeated (Cohen & Cohen, 1983, p. 238).

Results of moderated regression for rating accuracy are presented in Table 2. As one can see in Table 2, the interaction term added to the prediction of accuracy (β = −.143, p < .005, ΔR² = .02). Following the recommendations of Cohen and Cohen (1983) and others (e.g., Aiken & West, 1991; Stone & Hollenbeck, 1989), split-group regression analysis was performed. First, the sample was split into low and high self-monitors, and then simple regres-

---

Table 1

<table>
<thead>
<tr>
<th>Means and Standard Deviations of Attitudinal Measures, Accuracy of Ratings, and Pay Increase Decisions for Low and High Self-Monitors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attitudes</strong></td>
</tr>
<tr>
<td><strong>ATAA</strong></td>
</tr>
<tr>
<td><strong>Self-monitor</strong></td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>High</td>
</tr>
</tbody>
</table>

Note. ATAA = attitudes toward accurate appraisal; IC = instrumental commitment; AC = affective commitment.
also investigated the moderating effects of self-monitoring on the relationship between affective commitment and decisions ($\beta = - .057, p < .001, R^2 = .135$) but not for high self-monitors ($\beta = - .147, ns$). These results indicate that self-monitoring not only had a direct and significant effect on accuracy of ratings but also moderated the relationship between attitudes toward accurate appraisal and accuracy of ratings such that the relationship was stronger at lower levels than at higher levels of self-monitoring. Self-monitoring did not moderate the relationship between instrumental commitment and ratings ($\beta = .055, ns$) or affective commitment and ratings ($\beta = - .053, ns$).

Hypothesis 3b predicted that self-monitoring would moderate the relationship between attitudes toward accurate appraisal and accuracy of decisions such that the relationship would be stronger for low than for high self-monitors. Results of moderated regression analyses for decision accuracy are presented in Table 3. Self-monitoring was significantly predictive of decision accuracy ($\beta = -.635, p < .001, R^2 = .403$), but attitudes were not ($\beta = - .026, ns$). The interaction term did not add to the prediction of decision accuracy ($\beta = -.083, p = .128, \Delta R^2 = .007$). Self-monitoring did not moderate the relationship between instrumental commitment and decisions ($\beta = .001, ns$) or the relationship between affective commitment and decisions ($\beta = -.057, ns$).

Discussion

In addition to investigating the direct effects of attitudes and self-monitoring on accuracy of ratings and decisions, this study also investigated the moderating effects of self-monitoring on the relationship between attitudes and accuracy of ratings and decisions. Contrary to expectations, attitudes were not related to accuracy of ratings or decisions. Using data collected from 29 managers, Tziner and Murphy (1999) reported that instrumental commitment and affective commitment, both general attitudes, were related to the overall level of ratings. Although the sample size used in this study was much larger than Tziner and Murphy’s sample, the data failed to yield significant results. Moreover, the expectation that attitudes specific to appraisal behaviors, such as attitudes toward accurate appraisal, would be related to ratings and decisions was also not upheld. This latter finding is surprising because the sample size of 210 subjects afforded a power of .90 to detect even small effects of magnitude ($r = .20$ at the .05 $\alpha$ level; Cohen, 1988). Results of this study indicate that attitudes, regardless of whether those attitudes are general or specific, are not related to accuracy of ratings and decisions.

Moderated regression analyses indicated that self-monitoring moderated the relationship between attitudes and rating accuracy such that the relationship was stronger for low self-monitors than for high self-monitors. In addition, the follow-up split-group regressions indicated that attitudes of low self-monitors were positively related to accuracy of ratings. Together, these results suggest that accuracy of ratings of low self-monitors but not high self-monitors can be predicted from their attitudes toward accurate appraisal.

The most robust result of these data is the strong support for self-monitoring. Self-monitoring significantly affected rating accuracy ($\beta = -.720, R^2 = .53$) and decision accuracy ($\beta = -.635, R^2 = .40$) such that accuracy declined with higher scores on self-monitoring. The strong support for self-monitoring in this

### Table 2

Results of Moderated Regression Analyses for Rating Accuracy

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\beta$</th>
<th>$t$ (one-tailed)</th>
<th>$p$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>.087</td>
<td>1.24</td>
<td>.216</td>
<td>.007</td>
<td>.007</td>
<td>.003</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>.054</td>
<td>1.11</td>
<td>.268</td>
<td>.001</td>
<td>.517</td>
<td>.520</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>-.720</td>
<td>-14.87</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>.063</td>
<td>1.33</td>
<td>.184</td>
<td>.000</td>
<td>.020</td>
<td>.538</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>-.716</td>
<td>-15.06</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes $\times$ Self-Monitoring</td>
<td>-.143</td>
<td>-2.99</td>
<td>.003</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3

Results of Moderated Regression Analyses for Decision Accuracy

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\beta$</th>
<th>$t$ (one-tailed)</th>
<th>$p$</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>-.026</td>
<td>-0.371</td>
<td>.711</td>
<td>.001</td>
<td>.001</td>
<td>-.004</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>-.056</td>
<td>-1.032</td>
<td>.303</td>
<td>.403</td>
<td>.403</td>
<td>.397</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>-.635</td>
<td>-11.731</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>-.050</td>
<td>-0.932</td>
<td>.352</td>
<td>.410</td>
<td>.007</td>
<td>.401</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>-.633</td>
<td>-11.715</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes $\times$ Self-Monitoring</td>
<td>-.083</td>
<td>-1.528</td>
<td>.128</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
study is consistent with the results of Jawahar and Stone (1997). In a lab study, Jawahar and Stone manipulated appraisal purpose and availability of funds for pay increases to vary the severity of consequences across the three experimental conditions of severe consequences, less severe consequences, and less or no consequences. In the first two conditions, ratings of subjects who were high self-monitors were more lenient than ratings of low self-monitors, and high self-monitors in the severe-consequences condition were most lenient. Similar to subjects in the severe-consequences condition of Jawahar and Stone's study, subjects in this study were informed that ratings were collected for pay increase purposes and were led to believe that the company had plenty of funds—money for pay raises. The present study then replicated Jawahar and Stone's finding that even perceptions of consequences prompt high self-monitors, in comparison to low self-monitors, to rate leniently and inaccurately.

The finding that high self-monitors tend to produce lenient and inaccurate ratings is consistent with prior studies (e.g., Longecker, Sims, & Gioia, 1987; Tziner, 1999) that have indicated that in the process of performance appraisal, at least some raters report being inclined to invoking political considerations. It is possible that sanctions against lenient ratings found in at least some ongoing organizations might mask tendencies of raters who are high self-monitors to inflate their appraisals of others. Alternatively, one could argue that the politics and other interpersonal aspects of appraisals would accentuate differences between low and high self-monitors so that ratings and decisions of high self-monitors will be even more lenient than those of low self-monitors. For instance, in a field experiment involving customer service employees, decision makers who were high self-monitors engaged in more information manipulation to justify their decisions than low self-monitors. More important, these effects were more pronounced under conditions of high accountability (Fandt & Ferris, 1990) and are consistent with results of much of the earlier research on self-monitoring (see Caldwell & O'Reilly, 1982b; Snyder, 1979).

It appears that when ratings and decisions have consequences, high self-monitors who also have low self-efficacy as raters may invoke political considerations and manipulate ratings and decisions (Tziner, 1999).

Strong support for self-monitoring suggests that leniency can be predicted from measures of individual differences (Guiford, 1954). Results also corroborate Kane et al.'s (1995) assertion that given equal situational parameters, differences in the tendency to rate leniently or severely might reflect personality differences among raters. It is important to uncover such individual differences because the inability to discriminate inaccurate from accurate raters and reward the latter has been noted as the most significant barrier to enhancing accuracy of performance ratings (c.g., Murphy & Cleveland, 1991, 1995).

One potential limitation of this study is the use of the "paper people" design. Paper people studies in which raters read performance vignettes and then rate performance of hypothetical ratees are considered to be less realistic than behavior observation studies in which ratings are based on direct or indirect (e.g., through videotape) observation of ratees' behavior. However, one could argue that in many professional jobs (e.g., loan officer), outcomes of behaviors (e.g., number of loans issued, dollar value of loans) are likely to be just as important, if not more important, than observable behaviors in influencing performance evaluations. Thus, paper people studies may in fact simulate important features of actual appraisals. In a meta-analysis, Murphy, Herr, Lockhart, and Maguire (1986) found that the average effect sizes were slightly larger in paper people studies ($d = .42$) than in behavior observation studies ($d = .31$), but this difference was largely restricted to studies of the effects of variation in true performance level and the effects of purpose of appraisal. With respect to appraisal purpose, Murphy et al. contrasted 2 paper people studies with 7 behavior observation studies and found that paper people studies yielded larger effect sizes than behavior observation studies. However, a more recent meta-analysis that contrasted 4 paper people studies with 16 behavior observation studies found the opposite—effect sizes were larger in behavior observation studies than in paper people studies (Jawahar & Williams, 1997). Given such inconsistent results, the use of the paper people design may not be a serious limitation.

**Implications**

The support for self-monitoring is noteworthy for several reasons. First, given the 60/40 split of high and low self-monitors in the general population (Gangestad & Snyder, 1991; Snyder, 1987), it is not surprising that leniency of ratings and decisions is the norm in the industry (Bretz et al., 1992). Second, self-monitoring is a stable dispositional characteristic, and results of this study indicate that it is substantially predictive of leniency of ratings ($\beta = -.720, R^2 = .53$) and decisions ($\beta = - .635, R^2 = .40$). Finally, although leniency may be adaptive and perhaps the most rational behavior from the rater's perspective, over time, the effects of lenient ratings when those ratings are also inaccurate (as in this study) are likely to be more dysfunctional than functional for the organization. Therefore, identifying rater characteristics such as self-monitoring that influence the quality of ratings has practical significance.

Results of this study clearly indicate that performance ratings and decisions made by high self-monitors are lenient and inaccurate. It is also clear that high self-monitors inflate ratings when there are consequences contingent on ratings but evaluate accurately in the absence of such consequences (see Jawahar & Stone, 1997). Obviously, ratings obtained for administrative purposes (e.g., pay increases, retention decisions) have consequences, and results of at least one study suggest that high self-monitors may manipulate information even more when they are required to justify their evaluations of others to their own superiors (Fandt & Ferris, 1990). So it is not clear what practitioners can do to keep high self-monitors from leniently evaluating others. Results of White and Gerstein's (1987) study suggest one possibility. In their study, high self-monitors engaged in behaviors that were most likely to avoid disapproval of others. Perhaps by making high self-monitors feel secure, the motivation of high self-monitors to engage in behaviors for the purpose of avoiding disapproval could be reduced. One avenue to increase high self-monitors' sense of security with respect to performance appraisal would be to require all raters, including high self-monitors, to meet with and review performance of their subordinates, albeit informally, several times a year. This procedure could reduce high self-monitors' tendency to fear disapproval at the time of the formal appraisal and encourage them to more accurately evaluate performance of others.
Future Directions

Only two other studies (Tziner & Murphy, 1999; Tziner et al., 1998) have investigated the influence of attitudes on ratings. The present study failed to replicate the marginal support for attitudes reported by Tziner and his colleagues. These inconsistent results, coupled with the relationship reported between attitudes and diverse behaviors (e.g., absenteeism, turnover), will, hopefully, encourage more research on the influence of attitudes on appraisal behaviors. This study should be replicated with subjects from ongoing organizations using actual ratings collected for administrative purposes. If data from ongoing organizations evidence support for attitudes, then future research should examine factors that shape attitudes toward accurate appraisal. For instance, one may expect low levels of trust in the appraisal process (Bernardin & Orpan, 1990) and perceptions of prevalence of political considerations in performance appraisal (Tziner, 1999; Tziner, Latham, Prince, & Haccou, 1996) to lead to unfavorable attitudes. The influence of attitudes on dependent variables other than accuracy of ratings and decisions should also be examined. For instance, raters with favorable attitudes could be expected to spend more time observing, reviewing, and providing feedback to employees. Such behaviors could encourage employees to increase their performance and also enhance their satisfaction with performance appraisal and their trust in the appraisal process. Also, future research should uncover other personality factors or stable individual differences with the potential to influence appraisal behaviors. For instance, raters who have high scores on the Conscientiousness factor of the five-factor personality model could be expected to appraise accurately relative to raters who are less conscientious. In terms of moderators, perceptions of risks associated with distorting ratings (Kane & Kane, 1992) could moderate the relationship between attitudes toward accurate appraisal and the accuracy of ratings. Identifying antecedents of attitudes and moderators of the relationship between attitudes and appraisal behaviors could then serve as the basis for designing interventions to curb the tendency to inflate ratings.

References


Kane, J. S., & Kane, K. F. (1992). The analytic framework: The most...