

Influence of Context Effects on Various Rating Measures

Robert D. Costigan
St. John Fisher College

■ ABSTRACT

This performance appraisal study examined the influence of the within-ratee context effects bias (i.e., contrast effects and assimilation effects) on various rating measures. Previous research in this area has produced inconsistent findings and is fraught with methodological problems. The present study contrasted ratings made in two contextual conditions (i.e., no time delay between evaluation of contextual and target videotapes and a seven-day interval between evaluation of contextual and target videotapes) with control group ratings (i.e., evaluation of the target videotape only). Eighty-one undergraduate students rated videotape(s) of a fictitious college instructor's performance. The results of this study indicated evidence for the presence of contrast effects in both the no-delay and seven-day interval conditions when ratee performance improved substantially from the first to the second videotape. Furthermore, rater affect was found to mediate performance ratings. Implications and directions for future research are also discussed.

■ INTRODUCTION

Kravitz and Balzer (1992) stated that it is impossible to have context-free appraisals. Supervisory performance ratings are thought to be influenced by many environmental or contextual variables (Murphy & Cleveland, 1991). Namely, environmental influences (e.g., social and legal system) and organizational influences (e.g., group, leader, and task characteristics) are likely to affect the rating process. Murphy and Cleveland (1991) identified the ratee's own past performance as a micro-level context variable. More specifically, an employee's prior performance may provide a context in which this same employee's current performance is evaluated, yielding distorted ratings. Kravitz and Balzer (1992) referred to this bias as context effects.

There are two kinds of context effects: assimilation effects and contrast effects. Assimilation effects are said to be present when an employee's past performance ratings tend to engulf his or her current ratings (Kravitz & Balzer, 1992). That is, current ratings are biased in the direction of the earlier ratings. Contrast effects, on the other hand, are present when a contrast between a ratee's past and current performance causes current ratings to be exaggerated away from prior performance ratings. Because of the contrast, current ratings move towards the extreme of the rating scale, in the opposite direction of the previous performance ratings (Kravitz & Balzer, 1992).

Contrast and assimilation effects are within-ratee effects, meaning that the ratee's previous performance provides a backdrop for the evaluation of his or her subsequent performance. Researchers have also examined between-ratee contrast effects in employment interviews, work samples, assessment-center activities, and performance appraisals. Between-ratee contrast effects suggest that the performance of one or more individuals provides a context for the evaluation of another individual's performance.

Requests for reprints or correspondence concerning this article should be sent to Dr. Robert D. Costigan, 3690 East Avenue, Rochester, NY 14618.

Murphy, Balzer, Lockhart, and Eisenman (1985) found a contrast-effects bias when raters evaluated a videotape of average ratee performance immediately after rating two videotapes of either excellent performance or poor performance. Specifically, both behavior and evaluative ratings of the ratee's performance in the average performance videotape were distorted in the direction away from the performance level depicted in the initial videotapes. Smither, Reilly, and Buda (1988) replicated Murphy et al.'s (1985) contrast-effects finding, using the same tapes, rating scales, and procedures. Paradoxically, they found assimilation effects when raters reviewed previous written performance ratings prior to observation of the target videotape. In a 1986 study, Balzer found support for contrast effects in the entries made in a diary format. More specifically, he reported that raters are more likely to record behavioral incidents that are inconsistent with prior performance than behavioral incidents that are consistent.

Murphy et al. (1985) also introduced a one-day delay procedure in their study. Raters observed either two videotapes of excellent performance or poor performance, followed by the average-performance videotape. Raters immediately evaluated either the poor performance or good performance videotapes, and then returned the next day to evaluate the ratee's performance in the average performance videotape. When rating the third, average-performance tape from memory, the contrast-effects bias did not surface in either behavior or evaluative ratings. Murphy et al. (1985) concluded that the presence or absence of contrast effects in ratings may be contingent upon the memory demands required of the raters and that this bias may be less of a problem in the more typical memory-based rating situations.

In a 1986 study, Murphy, Gannett, Herr, and Chen showed how the recall of subsequent ratee performance biased the behavior ratings of previous ratee performance in the one-day delay condition but not in the immediate rating condition. That is, an assimilation effect was detected in the delayed condition but not in the immediate rating condition. Memory-based behavior ratings of the ratee's prior performance tended to reflect the rater's schematic impression of the ratee that was formed by subsequent performance. Schemas that raters implicitly rely on to make their ratings contain both behavioral and the more general category-related information. Murphy et al.'s (1986) findings are consistent with well-accepted cognitive rating models (e.g., Feldman, 1981) that suggest that rating delays produce suboptimal, schema-influenced evaluations. With increased time delays between observation and recall, the main characteristics of the category to which the ratee has been assigned should account for more of the rating variance. In sum, Murphy et al. (1986) concluded that, when memory demands are great, assimilation effects are more likely to occur while contrast effects are more likely when memory demands are minimal.

With the exception of a few published studies (e.g., Salvemini, Reilly, & Smither, 1993; Steiner & Rain, 1989), most researchers examining context effects, such as Becker & Villanova (1995), have used flawed research designs (Kravitz & Balzer, 1992). Raters in these studies first observe a ratee's videotaped performance that provides a context, perhaps either excellent or poor performance; and then they are exposed to a target videotaped performance, often times depicting average performance. Significant differences in the ratings of the target performance produced by the contextual manipulation (i.e., excellent and poor prior performance conditions) were interpreted as evidence for either assimilation or contrast effects, depending upon the direction of the mean differences. In their methodological critique of context effects studies, Kravitz and Balzer (1992) showed the inadequacy of this research design. Simply put, Murphy et al. (1985) and Murphy et al. (1986), among others, did not include a random context control condition. Raters in such a control-group condition observe and evaluate only the target videotape, without exposure to either set of prior-performance videotapes. Kravitz and Balzer (1992), therefore, concluded that, due to a faulty research design, it is impossible to determine unequivocally that contrast or assimilation effects were present or absent in past studies.

In one study that did incorporate a control group condition, Steiner and Rain (1989) found support for assimilation effects when poor performance followed average performance videotapes. This bias did not emerge until the raters observed three contextual videotapes (i.e., three tapes depicting average performance); in other words, one or two contextual videotapes did not produce the bias in this study. Assimilation or contrast effects was noticeably absent when average performance videotapes were followed by one good performance videotape. In a second study, Steiner and Rain (1989) included a one-day interval between the showing of each videotape, contending that observation of ratee performance over a period of time reflects, to a greater extent, a realistic appraisal environment. Raters in this study evaluated each tape immediately after observation. They found that ratings of the last videotape showing inconsistent performance were biased with assimilation effects in both good and poor subsequent performance conditions.

In another control-group study, Kravitz and Balzer (1992) reported unexpected results. Contrast effects were apparently

found in the poor-performance contextual condition while assimilation effects were found in the good performance contextual condition. Salvemini et al. (1993) found robust assimilation effects for both the poor and good performance context conditions after providing raters with written information about the ratee's prior performance instead of showing them videotapes. Interestingly enough, Salvemini et al. (1993) reported that, when raters were offered a monetary incentive to rate accurately, assimilation effects did not surface in the ratings. Foti and Hauenstein (1993) found support for the presence of assimilation effects in ratings when raters were asked to perform two distinct tasks simultaneously (i.e., observing the ratee's performance while performing another mental exercise).

Considering the conflicting results reported in these context effects studies and the methodological problems in many of them (i.e., absence of a random context control condition), more research is necessary in this area. The overall purpose of this paper is to further explore within-ratee context effects in performance ratings. An increased understanding of how and when a prior-performance context systematically influences subsequent performance ratings is needed. Two time-of-rating conditions were included in the present study, considering that ratee performance improvement (decrement) does not usually occur in real-world settings without a time interval. Thus, control group ratings were contrasted with ratings made in two contextual conditions (i.e., no time interval between the evaluation of the contextual videotape and the target videotape and a seven-day interval between the evaluation of the contextual and target videotapes). Another purpose was to determine whether a seven-day time interval between the evaluation of the poor performance videotape and the videotape depicting good performance lessens the context effects bias relative to the other two-tape condition. Other than Smither et al.'s (1988) study which had raters reviewing ratee performance over a three-week time period, no other context effects study has incorporated such a long interval between the contextual and target videotapes. Note that Smither et al. (1988) did not find support for the presence of context effects; they failed, however, to include a control group condition in their study.

Researchers of context effects have mostly used a target videotape depicting average target performance (Becker & Villanova, 1995). The thinking is that average ratee performance should produce increased rater uncertainty, allowing either assimilation or contrast effects to surface more readily. High performance or low performance in the target videotape, on the other hand, creates a more certain rating environment for raters, thereby enhancing rater accuracy. This assumption remains largely untested, especially with a proper research design. Unlike other studies, the present study examined the context effects produced by a more extreme variation in performance (i.e., bad performance videotape followed by a good performance videotape). Raters in real-world, 360-degree appraisal environments (i.e., ratings are collected all around the ratee, from external and internal customers, peers, subordinates, and bosses) may at times see significant variation in employee performance. For instance, supervisors and peers may see shaky performance on the part of a new employee followed by improved performance. Also, employees (e.g., auditors, bank tellers, hotel housekeepers) may have a bad day serving customers followed by greatly improved performance on another day. Formal or informal performance feedback from customers may be solicited at brief intervals. To repeat, a sequence of poor ratee performance followed by good performance does happen in our encounters with others, providing these raters with more of a contrast than does a poor performance context followed by average performance.

The extent to which the context manipulation in this study influenced various rating measures is also worth investigating. These measures included: (a) behaviors showing substantial improvement from the context to the target videotape, (b) ratee behaviors remaining constant over the two videotapes, (c) liking (affect) of the ratee, and (d) general impression of the ratee's overall performance. General impression is an important variable to look at, considering that traditional cognitive rating models (e.g., Feldman, 1981) hold that the rater's overall impression will increasingly affect both behavioral and dimensional ratings with longer time delays between observation and evaluation.

Also in contrast to previous research, this study incorporated an affect measure to determine whether rater liking mediates between the context manipulation and performance ratings. Dipboye (1985) claimed that affect or supervisory liking has been overlooked by performance appraisal researchers as a potential contaminant of ratings. In a revealing study, Alexander and Wilkins (1982) found that supervisory liking of the ratee accounted for more of the rating variance than did the ratee's more objective level of performance (i.e., quality and quantity of ratee performance). Likewise, Cardy and Dobbins (1986) found that rating accuracy suffers when the rater either strongly likes or dislikes the ratee. Wayne and Ferris (1990) found support for their model which predicts, in part, that supervisory liking mediates between ratee impression management (i.e., self-enhancing communication) and supervisory ratings. Murphy and Cleveland (1991) reported the results of Williams, Allinger, and Pulliam's (1988) study which found that increased rater liking leads to more lenient performance ratings.

The male ratee in the present study showed dramatic improvement in his performance from the first videotape to the second. Just as the more superficial impression management tactics led to increased supervisory liking and eventually higher ratings in the Wayne and Ferris' (1990) study, the context manipulation in this study was predicted to lead to increased rater liking relative to the control group condition. In other words, if superficialities (e.g., impression management) can increase rater liking, then substantive performance improvement on the part of the ratee should likewise result in increased rater affect. Drawing further support from Williams et al.'s (1988) findings, it was predicted that, with increased rater liking due to substantive performance improvement, raters in both two-tape conditions (i.e., no delay between the poor and good videotapes and seven-day interval between the same two videotapes) will rate the target performance more leniently than will raters in the one-tape, control-group condition. In essence, substantial performance improvement on the part of the ratee should lead to an increase in rater liking which, in turn, should produce contrast effects and not assimilation effects. Raters in the control group condition were predicted to like the ratee less, leading to lower ratings in the one-tape condition.

■ METHOD

SAMPLE

The sample consisted of 44 male and 37 female college students enrolled in undergraduate business and psychology courses at a small eastern college ($N = 81$). They received extra credit for participating. Research participants were randomly assigned to one of three experimental conditions (evaluated the contextual videotape [poor performance] and then the target videotape [good performance] without a delay vs. evaluated the contextual videotape [poor performance] and then seven-days later returned to evaluate the target videotape [good performance] vs. evaluated the target videotape only [good performance]).

STIMULUS MATERIALS

Two videotapes, each 12-15 minutes long, were developed by Costigan and Dossett (1991) with a common lecture theme, *Landsaped Office Designs*. The same qualified male actor portrayed a fictitious college instructor in the two videotapes. The behaviors presented in the two videotapes were related to a college instructor's duties. Except for the planned variations in the quality of performance of four behaviors (i.e., good vs. poor performance), the lecture content was the same for the videotapes. The four behaviors that varied in performance between the two videotapes are: (a) "Gives real-life examples to illustrate technical terms," (b) "Obtains up-to-date material to supplement text material," (c) "Provides outline that will be followed during lecture," and (d) "Uses body movements during lecture." These four teaching behaviors were interspersed in the same order throughout the two videotapes. Each behavior was clearly portrayed six times in each videotape. Again, the performance level of all other behaviors shown in the two videotapes (e.g., "Explain concepts and definitions clearly and completely") was held constant. A pilot study was conducted to insure that the performance levels for the four behaviors were representative of relatively low performance in the first videotape and relatively high performance in the second videotape. For this pilot study, a sample of 26 undergraduate day students (10 for the low performance tape and 16 for the high performance tape) evaluated the four varied instructor behaviors in the two videotapes. Pilot-study raters used graphic scales with evaluative anchors ([1] being "poor" and [5] being "excellent") to evaluate the instructor's effectiveness on these four behaviors. They studied the rating scales prior to observation and evaluation of the videotape. Independent-samples *t*-tests were used to contrast the ratings of each good-performance behavior with the ratings of each poor-performance behavior. The results of all *t*-tests were statistically significant, $p < .001$. These tests support the position that there are substantive performance differences in the four behaviors between the two videotapes. That is, the ratee's performance on these four behaviors in the first context videotape is relatively poor and the same ratee's performance on these same behaviors in the second target videotape is greatly improved.

PROCEDURES

All participants were told that the purpose of this research project was to use their ratings to test the effectiveness of a proposed new teacher evaluation form for a neighboring university. They were instructed that the instructor in the videotape would not see their ratings. Groups with approximately five students were shown the videotape(s). Unlike the pilot-study raters, participants did not see the rating scales prior to videotape observation. One third of the students observed and rated a poor performance videotape (contextual videotape). Next, they observed and rated a good performance videotape (target videotape) showing the same instructor giving the same *Landsaped Office Design* lecture as in the previous videotape. There was no delay

between showings of the two videotapes. These students were told that the first lecture (i.e., context performance) was videotaped in the previous semester whereas the second lecture (i.e., target performance) was videotaped in the current semester. Another third of the participants was given the same instructions as the previous group. They observed and rated the same poor performance videotape (context performance) and returned one week later to observe and rate the good performance videotape (target performance). They were also informed that the first videotaped lecture took place in a previous semester while the second videotaped lecture occurred in the current semester. The final third of the participants observed and rated only the good performance videotape (target performance). The instructions given to this third group of raters were the same as the two previous conditions, without referring to a first videotape. All subjects were debriefed after evaluating the good performance videotape.

DEPENDENT MEASURES

All participants rated the good performance videotape with the same five-point graphic rating scales (with [1] being "poor" and [5] being "excellent") previously used in the pilot study. Specifically, they rated the instructor's performance on the four behaviors that showed improvement in the second videotape. They also rated his performance on two other behaviors that did not vary across the two videotapes. These behaviors were: (a) "Explains concepts and definitions clearly and completely" and (b) "Presents material in an enthusiastic way." Students rated the instructor on two general impression items with the same five-point graphic scale: (a) "How would you rate the instructor's overall performance?" and (b) "What is your general impression of this instructor's total performance?" Lastly, they rated the instructor on a measure of affect with the same graphic scale: "How much do you like this instructor?"

■ RESULTS

The primary purpose of this study was to determine whether sharply improved performance by a particular ratee produces context effects in performance ratings. A further interest was whether a seven-day time interval between the evaluation of the poor performance videotape and the videotape depicting good performance lessens context effects relative to a no-delay condition. To test these concerns, a single factor (evaluation of two videotapes with no delay between videotapes vs. a seven-day interval between the evaluation of the two videotapes vs. evaluation of the good performance videotape only) multivariate analysis of variance (MANOVA) was conducted. Four behavior ratings of the ratee's performance in the second videotape only (i.e., behaviors that showed sharply improved performance) served as dependent measures in this analysis. Ratings made on the first videotape (i.e., poor performance in the context tape) were not considered for this or subsequent analyses since these ratings were not pertinent. Significantly higher means for one or more of the contextual conditions relative to the means for the control group condition would suggest the presence of a contrast effects bias. On the other hand, significantly lower means for one or more of the contextual conditions relative to the means for the control group condition would suggest the presence of an assimilation effects bias. No significant mean differences would suggest that both contrast and assimilation effects did not influence the ratings in this study.

The results of the MANOVA indicated that there were significant multivariate main effects, $F(8, 148) = 2.26, p < .05$. Table 1 shows the descriptive statistics for this analysis. As shown in Table 2, univariate test results were significant for three of the four improved behaviors. As shown in Table 1, planned independent-samples *t*-tests on the means indicated that both two-tape conditions produced significantly higher ratings than the control group condition for certain improved behaviors. The direction of the mean differences for these behaviors (as reported in Table 1) suggests support for contrast effects and not assimilation effects.

Another MANOVA was conducted on the two behaviors that remained constant across the two videotapes. The results of this MANOVA were marginally significant, $F(4, 152) = 1.99, p = .10$. Descriptive statistics for this analysis are also shown in Table 1. As reported in Table 2, univariate test results were significant for one of the two constant behaviors (i.e., "Presents material in an enthusiastic way"). Planned independent-samples *t*-tests on this behavior's means (see Table 1) indicated that the two-tape, no-delay condition produced significantly higher ratings than did the control group condition, providing additional evidence for contrast effects.

An analysis of variance (ANOVA) was conducted on a composite of the two items assessing the rater's general

TABLE 1
Means and Standard Deviations for Context Effects

Measures	No Delay / Two Videos		7-Day Interval / Two Videos		Good Performance Video	
	M	SD	M	SD	M	SD
Improved behaviors:						
"Gives real life examples ..."	4.00	.83	4.07 ^x	.73	3.84 ^y	1.12
"Obtains up-to-date material..."	4.07 ^v	.92	3.89 ^x	.89	3.33 ^{w,y}	1.11
"Uses body movements..."	4.11 ^v	.70	3.79	.80	3.44 ^w	1.05
"Provides outline..."	4.33	.73	4.22	.90	4.37	.84
Constant behaviors:						
"Explains concepts..."	4.11	.89	4.03	.92	3.73	1.00
"Presents material..."	3.15 ^x	1.20	2.73	.92	2.35 ^y	1.16
General impression composite:	3.63 ^v	.91	3.24	.95	2.85 ^w	.97
Affect (liking) item:	3.33 ^v	1.07	2.93 ^v	1.07	2.15 ^w	.82

Note: Values with the different superscript (x or y) are significantly different at $p < .05$ level. Values with the different superscript (v or w) are significantly different at $p < .01$ level.

$n = 27$ for each condition.

TABLE 2
ANOVAs of Context Effects Manipulation for Behavioral, General Impression and Affect Ratings

Measures	MS	F (2,78)	W ²
Improved behaviors:			
"Gives real life examples ..."	2.81	3.40*	5.59%
"Obtains up-to-date material..."	4.01	4.20*	7.32%
"Uses body movements..."	3.00	4.03*	7.68%
"Provides outline..."	.16	.24	.00%
Constant behaviors:			
"Explains concepts and definitions..."	1.07	1.22	.55%
"Presents material in enthusiastic..."	4.26	3.51*	5.97%
General impression composite:	4.08	4.59*	8.14%
Affect (liking) item:	9.79	9.88**	17.99%

* $p < .05$ ** $p < .01$.

impression of the ratee's performance in the target videotape. Coefficient alpha was .96 for these items. As shown in Table 2, the results of the ANOVA were significant, $p < .05$. Descriptive statistics for this analysis are shown in Table 1. As reported in Table 1, planned independent-samples *t*-tests on the general impression composite means indicated that once again the two-tape, no-delay condition produced significantly higher ratings than the control condition, indicating support for contrast effects. The two-tape, seven-day interval condition did not, however. An ANOVA was conducted on the measure of affect (i.e., liking the ratee). As shown in Table 2, the ANOVA results were significant, $p < .001$. As reported in Table 1, planned independent-samples *t*-tests on the affect means indicated that both two-tape conditions (i.e., no-delay and seven-day interval) produced significantly higher ratings than did the control condition, providing further evidence for contrast effects. The *t*-test results also revealed that the means in both two-tape conditions did not differ significantly for all of the analyses in this study.

The large standard deviations for ratings made in the control group condition (as reported in Table 1) suggest potential for the violation of the assumption of homogeneity of variance. The Bartlett-Box *F* test results for homogeneity of variance (see Howell, 1982) were not significant, abating concern over the improper use of ANOVAs in this study's analyses in this study.

To determine whether rater affect (i.e., liking) mediates between the context manipulation and the performance ratings, a series of hierarchical regressions were conducted. The affect rating (i.e., liking the ratee) was entered first into each regression equation. The context manipulation was dummy coded (two-tape, no-delay condition = 00; two tape, seven-day interval condition = 10; one-tape control condition = 01) and then entered second into each regression equation. If rater affect is operating as a mediator, then this liking rating should explain a significant portion of the rating variance while the incremental effects of the context manipulation being negligible. As shown in Table 3, the consistently high R^2 's for the rater affect variable with non-significant incremental effects of the context manipulation variable suggest support for the notion that rater affect mediates performance ratings in this study.

■ DISCUSSION

This study was designed to advance our understanding of within-ratee context effects bias by exploring further the conditions that produce such effects. It adopted Kravitz and Balzer's (1992) suggestion that, when studying context effects, a

TABLE 3
Hierarchical Regression of Behavioral and General Impression Ratings
with Affect as a Mediating Variable

Measures	Stage 1		Stage 2*	
	Affect (Liking) R^2	F	Incremental Effect of Context Manipulation R^2 Change	F Change
"Gives real life examples ..."	.36	44.86*	.02	1.00
"Obtains up-to-date material..."	.35	43.43*	.00	.20
"Uses body movements..."	.30	34.59*	.01	.36
"Provides outline..."	.16	15.30*	.06	3.07
"Explains concepts..."	.37	46.42*	.01	.71
"Presents material..."	.61	123.75*	.01	.57
General impression composite:	.72	199.56*	.01	1.12

Note: $N = 81$. $dfs = 1,79$ for *F*-test in Stage 1; $dfs = 3, 77$ for *F*-test in Stage 2.

* Incremental effect given that the complementary effect was previously entered in the regression equation at Stage 1.

* $p < .001$.

control group condition should be included in the experimental design. The findings of this study provided some support for the presence of contrast effects in ratings after the ratee's performance showed substantial improvement. More specifically, certain behavioral ratings, a general impression composite, and a rating of affect were influenced by the context manipulation in the no-delay condition and, to some extent, in the seven-day interval condition. Note that, even though statistical significance was not reached in some univariate tests, the direction of the mean differences, as reported in Table 1, consistently favored contrast effects and not assimilation effects. The context manipulations in previous research quite often incorporated average performance in the target videotape. This study deviated from this by showing substantial improvement in the target videotape in key behavioral areas. Sharp improvement in ratee performance may have abetted the potential for contrast effects over assimilation effects. The strong performance contrast between the two videotapes, with greater certainty as to the performance level of each improved behavior in the target videotape, apparently produced firm memory traces, allowing the effects of the contrast bias to occur after a seven-day interval between the two performances. The robustness of the contrast effects bias surfacing in the performance-improvement conditions of this study appears to be strengthened by the results of this study, not only because of its presence in some ratings made with a seven-day interval, but also because it generalizes to other rating measures (i.e., behaviors held constant across videotapes, general impression, and affect). Note that the effects of the contrast bias were generally stronger in the no-delay than in the condition with the seven-day interval.

Support for the contrast effects bias in ratings obtained here appears to be consistent with Kravitz and Balzer's (1992) finding. In other words, performance improvement produces contrast effects. This study did not examine whether a performance decrement produces assimilation effects, as did Kravitz and Balzer (1992). Future research should attempt to replicate their work in the performance-decrement area. If replicated, explanations as to why performance improvement conditions yield contrast effects and performance decrement conditions yield assimilation effects is worth investigating.

The underlying explanation to this phenomenon may be linked to the strong mediating effects of the affect measure found in this study. In line with the findings of Wayne and Ferris' (1990) and Williams et al.'s (1988) studies, the present study found that, with a significant performance improvement on the part of the ratee, raters seemed to like the ratee more than raters who saw him in the good videotape only. Perhaps, raters may have liked the ratee better due to their perception that he made a strong effort to improve his performance. Attribution theory may provide the clue as to why this occurs. Rater affect should also be tracked when ratee performance is declining to determine its mediating potential under such circumstances. To sum up, rater affect was found to mediate performance ratings in this study, suggesting that ratees may at first perform poorly followed by substantial performance improvement and fare better in their performance ratings than if they had performed consistently well from the beginning to end. Future research should give further attention to this surprising proposition as well as its implications to other person and product improvement domains.

Within-ratee contrast effects are thought to be less troublesome for real-world raters than other rating biases since it has surfaced in studies without a time interval between observation and judgment. Rating subordinates without a time delay is usually atypical of real-world rating conditions. This study incorporated a strong manipulation by having raters observe poor performance on one day and return seven days later to observe and evaluate substantially improved performance. This manipulation may better reflect real-world rating conditions. As mentioned, managers may at times see substandard performance on the part of a new organizational member followed later by improved performance. When this performance pattern occurs, managers should, at a minimum, be aware of the tendency to inflate their ratings of subsequent performance. Raters in 360-degree appraisal environments (e.g., internal and external customers) may also encounter substantial variation in employee performance over a period of time. A pattern of poor ratee performance followed, in a few days, by improved performance occurs in some customer-service exchanges. Customer ratings of the employee may then be vulnerable to the bias found in this study. It is worth investigating whether certain rating formats and rater training programs mitigate the deleterious effects of contrast bias.

This study is limited in a number of regards. One limitation regarding the mediating role of rater affect is that raters in the control group condition observed only one videotape performance (i.e., good performance). Perhaps, if these control-group research participants were shown the same high performance videotape twice, their affect level might possibly reach the affect level of subjects in the performance improvement conditions. Future research should consider this possibility. Another obvious drawback is the use of college students as raters in performance appraisal research because they lack real-world rating experience. Nathan and Lord's (1983) counter-argument was that, when studying cognitive processes in the realm of performance appraisal, the use of student raters does not seriously detract from the generalizability of the study. This premise

should hold for research examining unintentional rater biases as well.

■ REFERENCES

- Alexander, E.R., & Wilkins, R.D. (1982). Performance rating validity: The relationship of objective and subjective measures of performance. *Group & Organization Studies*, 7, 485-496.
- Balzer, W.K. (1986). Biases in the recording of performance-related information and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes*, 37, 329-347.
- Becker, G.A., & Villanova, P. (1995). Effects of rating procedure and temporal delay on the magnitude of contrast effects in performance ratings. *The Journal of Psychology*, 129, 157-166.
- Cardy, R.L., & Dobbins, G.H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672-678.
- Costigan, R.D., & Dossett, D.L. (1991). *Underlying causes of rating accuracy in memory-based behavior ratings: Prototypicality vs. holistic-processing*. Paper presented at the meeting of the International Academy of Management, Detroit, MI.
- Dipboye, R.L. (1985). Some neglected variables in research on discrimination in appraisals. *Academy of Management Review*, 10, 116-127.
- Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Foti, R.J., & Hauenstein, N.M.A. (1993). Processing demands and the effects of prior impressions on subsequent judgments: Clarifying the assimilation/contrast debate. *Organizational Behavior and Human Decision Processes*, 56, 167-189.
- Howell, D.C. (1982). *Statistical methods for psychology*. Boston, MA: Duxbury Press.
- Kravitz, D.A., & Balzer, W.K. (1992). Context effects in performance appraisal: A methodological critique and empirical study. *Journal of Applied Psychology*, 77, 24-31.
- Murphy, K.R., Balzer, W.K., Lockhart, M.C., & Eisenman, E.J. (1985). Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology*, 70, 72-84.
- Murphy, K.R., & Cleveland, J.N. (1991). *Performance appraisal: An organizational perspective*. Boston, MA: Allyn and Bacon.
- Murphy, K.R., Gannett, B.A., Herr, B.M., & Chen, J.A. (1986). Effects of subsequent performance on evaluations of previous performance. *Journal of Applied Psychology*, 71, 427-431.
- Nathan, B.R., & Lord, R.G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology*, 68, 102-114.
- Salvemini, N.J., Reilly, R.R., & Smither, J.W. (1993). The influence of rater motivation on assimilation effects and accuracy in performance ratings. *Organizational Behavior and Human Decision Processes*, 55, 41-60.
- Smither, J.W., Reilly, R.R., & Buda, R. (1988). Effect of prior performance information on ratings of present performance: Contrast versus assimilation revisited. *Journal of Applied Psychology*, 73, 487-496.
- Steiner, D.D., & Rain, J.S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, 74, 136-142.
- Wayne, S.J., & Ferris, G.R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. *Journal of Applied Psychology*, 75, 487-499.
- Williams, K.J., Allinger, G.M., & Pulliam, R. (1988). *Rater affect and performance ratings: Evidence for the moderating effects of rater perceptions*. Presented at Annual Convention of Society for Industrial and Organizational Psychology, Dallas, TX.

■ AUTHOR'S BIOGRAPHY

Robert D. Costigan is an assistant professor of management at St. John Fisher College. He earned a doctorate in applied psychology at the University of Missouri-St. Louis. His research interests include recruitment, selection, and performance appraisal.