

Unit Climate, Rater Goals and Performance Ratings in an Instructional Setting



KEVIN R. MURPHY, JEANETTE N. CLEVELAND,
TED B. KINNEY, AMIE L. SKATTEBO,
DANIEL A. NEWMAN AND HOCK PENG SIN

Department of Psychology, Penn State University

ABSTRACT

A series of studies was conducted to analyse the relationships between rating behaviour, rater goals and contextual variables, namely unit or class climate. Cleveland and Murphy (1992) suggested that errors and inter-rater disagreements in performance ratings could be understood in terms of differences in unit climate and the goals pursued by raters. In two studies, substantial correlations were found between self-rated goals and evaluations of instructor's performance; our second study provided evidence that goals measured before raters have an opportunity to observe performance are related to ratings obtained after observing instructor performance. A third study investigated Murphy and Cleveland's (1995) proposal that contextual variables, specifically organisational climate, affect rating behaviour. To test this hypothesis, data reflecting perceptions of the climate of college level courses and ratings of instructor performance were collected. Ratings of participative and co-operative climates showed a strong relationship with student ratings of instructors' performance. Average correlations between climate and ratings are substantially larger than those between rater goals and ratings; the mediation hypotheses tested in this study were not supported. Results of the three studies are discussed in relation to past research.

UNIT CLIMATE, RATER GOALS AND PERFORMANCE RATINGS IN AN INSTRUCTIONAL SETTING

Traditionally, theorising about the performance rating process has suggested that there exist observable and measurable objective differences in performance and that performance ratings are an appropriate method to measure that reality

(Judge and Ferris, 1993). In this tradition, performance appraisal has been treated as a measurement process, where the main focus is to collect reliable and valid measures of performance (Murphy and Cleveland, 1995). From this perspective, research has typically focused on improving rater ability, for example how to make raters more accurate (Bernardin and Beatty, 1984; Banks and Murphy, 1985; Harris, 1994). Rating problems are typically considered to result from poor rating scales or a lack of rater training. While it is certainly true that taking this perspective can lead to heightened understanding of the performance rating process, it ignores other sources of rating variance, such as rater goals and characteristics of the work context, which also hold an important role in the performance ratings process.

Murphy and Cleveland (1995) suggest that it may be better to think of the performance rating process as a goal-directed communication process. These authors suggest that when evaluating performance, raters attempt to use the ratings they assign to communicate information consistent with their personal interests. This idea is important because it implies that raters are not always motivated to provide accurate ratings. Rather, raters consider the consequences of rating in a particular manner and adjust the ratings to facilitate the attainment of their personal goals (for example, the goal of helping a particular subordinate to earn a promotion). Longenecker, Sims and Goia note, "Accuracy is not the primary concern of the practicing executive in appraising subordinates. The main concern is how best to use the appraisal process to motivate and reward subordinates" (1987: 191). To gain a deeper understanding of performance ratings, it is crucial to consider the rater goals and the context (for example, climate) within which raters evaluate performance.

Bjerke, Cleveland, Morrison and Wilson (1987) point out that the choice of exactly what message the rater communicates with performance ratings depends both on the goals being pursued by rater and the context in which rating occurs. If the impact of rater goals on performance ratings is not considered, consumers of performance appraisal information will not gain an accurate understanding of the messages communicated by the performance-rating process (Newman, Kinney and Farr, in press). Further, Murphy and Cleveland (1995) suggest that contextual variables, such as perceptions of climate, also may influence ratings. To date, little empirical research has examined these relationships. The purpose of the current paper is to present three studies examining the relationships among contexts, goals and ratings. Two studies investigate the links between rater goals and performance ratings. A third study examines the link between unit climate, rater goals and evaluations.

RATERS WHO PURSUE DIFFERENT GOALS GIVE DIFFERENT RATINGS
Several theories of personality, motivation and leadership focus on the role of goals in directing and energising behaviour (for example, Miner, 1984; Dweck and Leggett, 1988; Locke and Latham, 1990; Cropanzano, James and Citera, 1992). Goals are central to modern theories of work motivation (Roberson et al., 1989; Locke and Latham, 1990; 2002) and they provide a key to understanding

both levels of effort and strategies for performance in achievement-oriented situations (Dweck and Leggett, 1988). Goal-oriented frameworks have proved useful for predicting the performance and biases of financial analysts (Cianci, 2001), for understanding grade expectations in the classroom (Gaultney and Cann, 2001) and for understanding practice and participation in sports (Nasir, 2000).

Locke and Latham (2002) summarise three decades of research on goal setting. Their review makes several points that are useful for understanding the role of rater goals in performance appraisal. First, goals define end states that individuals hope to accomplish. As Cleveland and Murphy (1992) note, supervisors in organisations may have more important goals in mind when completing performance appraisals than simply providing accurate ratings (see also Murphy and Cleveland, 1995). Thus, inaccurate or distorted ratings might not be an indication that raters cannot accurately evaluate performance. Rather, they may be an indication that raters are trying to do something other than simply record employee performance when completing their ratings.

Second, goals are likely to have subconscious effects, but on the whole, people appear to be conscious of and able to describe the goals they are pursuing. Studies of performance evaluation in the military (Bjerke et al., 1987) and of executive appraisal (Longenecker et al., 1987) suggest that decision makers can articulate, after the fact, the goals they pursued when making decisions about their subordinate (which, can often be at odds with the goals of the organisation). Locke and Latham (2002) suggest that raters might be able to identify the goals they are pursuing at the time they complete appraisals, or even before they make decisions about others' performance. This suggests that it is reasonable to attempt to obtain information from raters about the goals they intend to pursue when completing performance evaluations and that information about these goals should help to predict the ratings they actually give.

Murphy and Cleveland (1995) note that it can sometimes be difficult to isolate the effects of rater goals on ratings, because rater goals appear to be influenced by the ratee's level of performance (see also Kerst, 1993). That is, raters appear to pursue different goals when evaluating truly good performers (for example, they might use ratings to try to maintain current performance levels) than when evaluating truly marginal performers (for example, they might inflate their ratings in an attempt to encourage better performance). Rater goals are also probably influenced by a number of contextual variables, especially those that have to do with ongoing and expected future relationships between the rater and the ratee. Thus, for example, raters who supervise a group of mainly good performers may focus their rating behaviour on maintaining that high level of performance and they may be less inclined to give ratings that highlight differences in ratees' performance (Murphy and Cleveland, 1995). Raters may be reluctant to give low ratings (even when they are richly deserved) if they believe that those ratings will lead to resentment, decreased motivation and even worse performance in the future.

One way to disentangle the influences of rater goals and ratee performance

levels on ratings is to hold ratee performance constant, by examining the relationships between goals and performance appraisals provided by raters who are asked to evaluate the same ratee. It is well-known that if several raters are asked to evaluate the same ratees, their ratings are likely to vary (Thornton, 1980; Bernardin and Beatty, 1984; Harris and Schaubroeck, 1988; Bretz, Milkovich and Read, 1992; Viswesvaran, Ones and Schmidt, 1996; Conway and Huffcutt, 1997; Gregarus and Robie, 1998; Murphy, Cleveland and Mohler, 2001). Research and theory reviewed by Cleveland and Murphy (1992) suggest that this variance might sometimes be a reflection of the different goals pursued by different raters. In particular, if two raters emphasise different goals when completing performance appraisals, they will tend to give systematically different ratings.

The first two studies presented here examine the links between rater goals and performance ratings assigned to instructors in end-of-term teacher ratings. All raters in each class have multiple, similar opportunities to view the same sample of performance and all raters in each class evaluate the same ratee. If the traditional psychometric model is correct, differences in the ratings assigned by different students should be independent of any differences in rater goals, reflecting random measurement error. If the goal-oriented approach described above is correct, information about the goals pursued by different raters who are evaluating the same sample of job performance should account for variance in their ratings. Thus, the main hypothesis of these first two studies is: H_1 – measures of the rating goals most strongly emphasised by raters will account for a substantial portion of the variance in the performance ratings they assign.

Following the theoretical lead of Locke and Latham (2002), we hypothesise that goals precede and have a causal impact on ratings. The design of our first study does not allow us to evaluate this possibility, but our second study includes time-lagged data that allow us to determine whether goals assessed before performance is observed and evaluated predict evaluations after a substantial period of performance observation.

STUDY 1: METHOD

In our first study, we obtained measures of rater goals from students in several college classrooms at the same time as end-of-term teacher evaluations were completed. After students provided their regular teacher evaluations, we asked them to complete a questionnaire that evaluated the extent to which raters considered each of several goals when rating their instructor.

Procedures and Participants

Students in seven separate classes (N ranged from 19 to 187), each with a different instructor, were asked to complete the questionnaire in class at the same time they completed a standardised University-wide, end-of-term teacher evaluation form. Participants were assured that their responses to both the teacher evaluation form and the goal questionnaire were anonymous, and questionnaires were counter-balanced (that is, half of the questionnaires asked about goals first, then about instructor performance, the other half asked about

instructor performance first). Each participant provided ratings for only one instructor.

We obtained usable questionnaires from 895 respondents (338 male, 500 female, 58 did not indicate gender).

Goal Questionnaire Development

A nineteen-item questionnaire was developed to assess the goals pursued by raters in evaluating their instructors. In developing our questionnaire, we used items and item statistics from a study by Kerst (1993), which investigated performance rating as a function of rating purpose, ratee performance level and rater goals. As part of her study, Kerst asked managers and supervisors in a mid-sized organisation to generate examples of goals they might pursue when completing performance evaluations. She then had other supervisors retranslate the goals into categories based on the purpose for which appraisals may be conducted (to provide feedback or to make administrative decisions) and the performance level of the ratee (poor, average, outstanding). Although anecdotal studies of rater goals exist (Bjerke et al., 1987), Kerst's study is unique in the sense that it provides a systematic record of rating goals articulated by actual users of performance appraisal. We used a sample of rating goals articulated by supervisors and managers as the basis for creating a rating goal questionnaire.

We selected nineteen goal statements from Kerst's (1993) study (rewording some slightly to fit the teacher rating context) that: (a) spanned the range of goal types and levels included in that study and (b) were most consistently scaled with regard to the type of goal and performance levels they corresponded to.

The nineteen goal items presented in Table 4.1 could be grouped on a conceptual basis around four main themes.

1. Eight of the items described goals that related to identifying the instructor's weaknesses (items 1, 3, 5, 6, 11, 12, 13 and 18; Cronbach's $\alpha=.91$);
2. Six items described goals that related to identifying and conveying information about strengths (items 4, 8, 9, 15, 16 and 17; Cronbach's $\alpha=.74$);
3. Three items dealt with providing fair and informative ratings (items 2, 7 and 14; Cronbach's $\alpha=.60$);
4. Two items dealt with motivating ratees (items 10 and 19; Cronbach's $\alpha=.67$).

A confirmatory factor analysis suggested that this model fit the goal rating data very well (GFI=.98, NFI=.97, RMSEA=.08), and we created scales to represent each of these correlated factors.

In this study, participants were asked to indicate the extent to which they agreed with the statement that each of the nineteen goals influenced or was important in the evaluation of their instructor's performance, using a scale from 1 (strongly disagree) to 5 (strongly agree).

Measures of Instructor Performance

The teacher evaluation form used by the university included three items that referred specifically to the effectiveness of the instructor and/or the course ("Overall, I would rate this course as good"; "The teacher organised the course effectively"; "Overall, I would rate this teacher as good"). In addition to these three items, we asked students to complete a one-item overall rating of their instructor's performance in the course ("The instructor performed very well in this course"). Ratings on all four items were highly correlated and a four-item composite provided a highly reliable index of instructor performance (Cronbach's $\alpha = .94$).

Results

Table 4.1 lists the means, standard deviations and the correlations with composite performance rating for each of the nineteen goals. Table 4.2 lists means, standard deviations and correlations of the four rating goal scales with performance ratings, as well as the intercorrelations among these scales.

Relationship between Goals and Ratings

Ratings of the importance of the nineteen goals were significantly and substantially related to performance ratings in six of seven classes (one class had only nineteen students, making it impossible to fit a regression model with nineteen predictors); shrunken R^2 values ranged from .15 to .61, with a median value of .49. Scores on the four goal scales were also significantly related, albeit less strongly, to ratings in all of the classes, with R^2 values ranging from .07 to .36, with a median value of .16.

We next examined relationships between goals and performance ratings in the total sample, using hierarchical regression. We first entered a set of dummy-coded variables representing the seven instructors and found that differences in instructors/courses (instructors are nested within courses) accounted for 18.7 per cent of the variance in teacher ratings. We then entered ratings of the importance of the nineteen goals and found a significant and substantial increase in the percentage of variance explained (incremental $R^2 = .31$). We used the same procedure to examine the predictive power of the four goal scales, independent of any differences across courses in mean performance ratings, and found a significant and substantial increase in the proportion of variance explained (incremental $R^2 = .24$).

Discussion

This study provides empirical support for Cleveland and Murphy's (1992) hypothesis that information about the goals pursued by raters can be used to predict the performance ratings they assign. Depending on the method of analysis, R^2 values ranged from the .20s to the .30s, indicating relatively strong links between goals and ratings. Raters who emphasised different goals did seem to assign different ratings to the same teacher's performance. Although goals probably are affected by the performance of the subordinate, it is likely that the

Table 4.1 Goals: Importance and Correlations with Performance Ratings: Study 1

Goal	Mean	SD	r ^a
1 Identify areas in which the instructor might need improvement	3.34	1.20	-.10*
2 Rate my instructor fairly	4.29	.98	.43*
3 Identify areas where the instructor needs more training	3.14	1.25	-.01
4 Convey my satisfaction with the instructor's performance	4.22	1.01	.48*
5 Identify areas that the instructor should focus on improving	3.34	1.23	-.06
6 Indicate where the instructor fell short in terms of performance	3.27	1.30	-.10*
7 Give my instructor a rating that s/he will realise is based on performance, rather than my judgment of him/her as a person	4.19	1.08	.30*
8 Identify my instructor's strengths and weaknesses	3.81	1.12	.20*
9 Highlight my instructor's performance so that his/her success is visible to his/her department head	3.84	1.16	.34*
10 Improve my instructor's confidence	3.25	1.32	.15*
11 Make it clear to my instructor that there is room for improvement	3.18	1.34	-.16*
12 Identify my instructor's performance deficiencies	3.13	1.33	-.09*
13 Challenge my instructor to improve her/his performance	3.19	1.33	-.07*
14 Clarify expected performance levels to the instructor	3.17	1.27	-.01
15 Evaluate the instructor in a manner that clearly indicates what was done well and what was done poorly	3.25	1.32	.12*
16 Indicate where instructor has exceeded performance expectations	3.81	2.03	.34*
17 Encourage the instructor's current level of performance	3.89	1.09	.43*
18 Encourage the instructor to improve performance	3.32	1.33	-.06
19 Motivate the instructor	3.48	1.26	.22*

a Correlation with performance rating composite.

* $p < .05$

Table 4.2 Correlations between Goals Scales and Performance Ratings: Study 1

Measure	Mean	SD	Correlations			
			1	2	3	4
1 Strength scale	3.91	.87				
2 Weakness scale	3.23	1.06	.38*			
3 Fairness scale	3.91	.86	.69*			
4 Motivate scale	3.37	1.12	.47*	.61*	.51*	
5 Performance Ratings	4.38	.85	.42*	-.10*	.29*	.21*

* $p < .05$

goals pursued by raters also reflect some stable rating tendencies that are not solely a function of the ratees or the rating context (Dweck and Leggett's 1988; Murphy and Cleveland, 1995). Our second study investigates the relationship between goals measured prior to observing the ratee's performance and ratings obtained after observing the ratee's performance.

STUDY 2

In Study 1, ratings of goal importance were collected at about the same time as ratings of instructor performance, which could possibly lead to inflated estimates

of their relationships. This research design does not allow us to determine whether rater goals are an antecedent or a consequence of the rater's overall evaluation of the instructor's performance. For example, it is plausible that a student who has already formed a somewhat negative opinion about his or her instructor's performance will tend to rate goals that refer to identifying the instructor's weaknesses as being important. What is needed is a longitudinal design, in which information about goals is obtained before raters have formed clear opinions about ratee performance. Our second study provides an opportunity to test the hypothesis that rater goals assessed before raters have meaningful information about the ratee's performance can predict the ratings they will give after being exposed to the ratee's performance.

Method

Ratings of goal importance were collected during the first week of the semester ($N=303$) and ratings of teacher performance ($N=232$) were collected at the end of the semester in five separate courses. We were able to match goal importance and teacher ratings for 186 participants.

Results

All five instructors received very high ratings, with a mean rating of 4.14, and individual instructor ratings ranged from 4.03 to 4.22, on a five-point scale. Differences across instructors accounted for 14 per cent of the variance in teacher ratings.

Relationships between Goals and Ratings

We first examined the relationship between ratings of goal importance collected at the beginning of the semester and ratings of teacher performance obtained at the end of the semester separately in each class. As in Study 1, we found substantial links between goal ratings and performance ratings in each classroom, with shrunken \underline{R}^2 values ranging from .23 to .52, with a median of .35 (in one course, there were fewer students than predictors and we were not able to estimate the shrunken \underline{R}^2). We next examined relationships in the total sample, pooling across instructors.

As in our previous study, we found a large and statistically significant multiple correlation between the nineteen goals measured at the beginning of the semester and performance ratings obtained at the end of the semester (adjusting for instructor mean differences, the incremental $\underline{R}^2=.26$). We found significant and moderately large relationships between goal scale scores obtained at the beginning of the semester and ratings collected at the end of the semester (incremental $\underline{R}^2=.08$).¹

Discussion

Ratings of goal importance collected at the beginning of the semester, before students have had an opportunity to observe the instructor's performance, predicted ratings of instructor performance collected at the end of the semester.

Because all raters in each class are evaluating the same target, this means that raters who entered the semester with different orientations toward performance rating (for example, more emphasis on providing information about ratee strengths, less emphasis on fairness) provided systematically different evaluations of the same instructor's performance.

Our first study documented the correlation between measures of goal importance and evaluations of teaching performance, but because ratings of goals and of performance were obtained at the same time, there is the possibility of a spurious correlation. Our second study helps to rule out this explanation for the goal-rating link and suggests that raters' goal orientations might account for variability in performance ratings that is independent of ratee performance. In particular, ratings of goal performance collected at the very beginning of the semester (before raters have had a chance to observe the ratee's performance) predict end-of-semester performance evaluations. The lag between collecting goal information and performance evaluations virtually rules out the possibility of reverse causation (that is, end-of-semester evaluations collected in Study 2 cannot have a direct causal influence on goal ratings collected at the beginning of the semester and there is too little information available at the beginning of the semester about the instructor's performance to make it likely that evaluations of performance at that early point have a substantial influence on goal ratings). Correlational research, even with longitudinal designs, rarely provides conclusive evidence for causation, but the findings of these two studies are consistent with a model that assumes that rater goals have an impact on the evaluations they provide.

Relationship between Organisational Climate, Rater Goals and Performance Ratings

Murphy and Cleveland (1995) argue that performance appraisal cannot be adequately understood outside its organisational context. These researchers classify contextual factors into a continuum from proximal to distal factors. According to Murphy and Cleveland, "*Proximal* factors are those that impinge directly on the individual rater, whereas *distal* factors affect the rater indirectly" (1995: 32). Distal factors include climate perceptions, which are hypothesised to affect the behaviour of raters and ratees through intervening variables, such as standards of performance. Below, we suggest that climate may also affect rating behaviour, possibly as a result of the influences of climate on rater goals.

According to Litwin and Stringer (1968), organisational climate describes a quality of the work environment perceived by its members that influences expectancies and incentives. These authors propose a model that defines organisational climate as an intervening variable, mediating between organisational system factors and motivational tendencies. System and leadership features are seen as generating an organisational climate, which in turn arouses or suppresses particular motivational tendencies. It seems reasonable to expect that perceptions of organisational climate, generated from system features and other contextual factors, may similarly arouse or suppress raters' motivational

tendencies in the form of rater goals. These goals, in turn, affect rating behaviour.

Murphy and Cleveland (1995) suggest that organisational context in which rating occurs is a major determinant of goal-oriented rating behaviour. Consistent with Litwin and Stringer's (1968) theory of organisational climate, they suggest that contextual information from the environment of the organisation, such as climate and culture, is integrated and interpreted by raters to formulate goals. According to these researchers, "...culture might determine what approach to appraisal and what behaviours are evaluated positively or negatively in the organisation" (1968: 227).

It may be possible that rater goals reflect perceptions of the rating environment, namely climate, thereby acting as an intervening variable, mediating the affect of climate perceptions on performance ratings. Therefore, in this study it is predicted that perceived climate will not only affect ratings directly, but that the influence of perceived climate on performance ratings will be mediated by rater goals.

STUDY 3: METHOD

Measures

In order to assess perceptions of climate, Ostroff's (1993) measures of co-operative climate and participative decision-making climate were used. The decision-making climate scale comprises four items measuring student voice in decisions that are made regarding the class ($\alpha=.73$), whereas the co-operative climate scale measures mutual helpfulness among students and between the students and the instructor, as well as the patience and listening behaviour of the instructor (13 items, $\alpha=.86$).

Study 3 employed the same measures of instructor performance ($\alpha=.93$) and rater goal importance (weakness goals $\alpha=.71$, strength goals $\alpha=.64$, fairness goals $\alpha=.33$ and ratee motivation goals $\alpha=.62$) as did Studies 1 and 2.

Procedure

As with Study 2, the measurement of goal importance ($N=631$) was separated in time from ratings of teacher performance ($N=592$); these two sets of measures were obtained two weeks apart. Ratings of climate for participation and co-operative climate ($N=631$) were collected contemporaneously with the goal importance measures three weeks prior to the end of the semester, while performance ratings were made at the end of the semester. Climate, goal importance and teacher ratings could be matched for 523 participants, representing fifteen courses/fifteen instructors.

Results

In this design, raters are nested within both courses and ratees. Thus, the design factors of course-level climate and ratee are perfectly confounded, which prohibits the unambiguous use of cross-level modelling of the relations between unit-level climate and individual-level ratings in this data. Rather, the critical issue addressed by this study involves the variability of ratings *within ratees*: the

source of variance that is generally considered error under true score theory. As such, we sought to demonstrate covariance of individual-level climate perceptions and individual-level ratings and rater goals within-classes.

Homogeneity of Ratings Within-Class

Estimation of between-instructor/between-class versus within-instructor/within-class variability in climate perceptions, goal importance and performance ratings can provide a guide to understanding the multi-level nature of these constructs. As can be observed in Table 4.3, one-way ANOVA and eta squared values indicate substantial between-instructor variability compared to within-instructor variability for both climate measures and performance evaluations, but less so for goal ratings. This result suggests that climate and teacher ratings are a course-level construct that vary across courses, whereas rater goals vary as much within courses as between them.

Relationships among Climate Perceptions, Goals and Ratings

Our next step was to examine relationships between facets of climate, rater goals and ratings. Although the most statistically powerful test of these relationships would come from calculating correlations for all 523 participants combined across all fifteen courses, doing so would confound relationships at the class level with those at the individual level (Rousseau, 1985; George and James, 1993; Ostroff and Harrison, 1999).

Because this study is primarily focused on within-class variance in ratings of a single rater, it was appropriate to analyse the relationships among variables within each class separately and then to aggregate results across replications. In order to estimate correlations among climate perceptions, rater goals and ratings made on a single rater, weighted mean correlations from the fifteen classes were estimated using meta-analysis. Results of these analyses for all variables appear in Tables 4.4a and 4.4b.

The average correlations between climate and performance ratings (Mr values are .42 and .46 for decision climate and co-operative climate, respectively) are substantially larger than those of rater goals with performance ratings (Mr estimates for rater goals range from .13 to -.04). Correlations among climate facets (Mr=.64) and among goal types (Mr estimates range from .26 to .53) were generally stronger than those between climate perceptions and goals (most mean correlations between climate and goals were below .20). Co-operative climates were more strongly related to rater goals of 'identifying strengths' and 'rating fairly' than were decision-making climates. Correlations involving 'identifying weaknesses' goals were not homogenous across classes/studies, with significant Q values suggesting possible moderators for nearly all of the relationships involving weakness goals.

Tests of Moderation and Mediation Hypotheses

Murphy and Cleveland (1995) suggested that the influence of distal antecedents (for example, organisational climate) on rating behaviour may be mediated by

more proximal antecedents, such as rater goals. The theoretical explanation noted earlier for climate effects on rating behaviour implies that climate may also moderate the influence of goals on rating behaviour, such that rater goals more strongly affect performance evaluations when climate is perceived to be participative.

In order to test the mediation hypothesis, performance ratings were regressed on rater goals, on raters' climate perceptions and then on both rater goals and climate perceptions. These regression analyses were based upon the meta-analytic correlations in Table 4.4 (cf. Viswesvaran and Ones, 1995). Multiple R^2 for goals was .05, for climate was .24 and for goals and climate together was .255 (all statistically significant for $\alpha = .05$). The large increase in R^2 with the addition of climate perceptions to the regression equation (R^2 change = .21) indicates that goals do not mediate the influence of climate on performance ratings, as the distal-proximal model of rating behaviour (Murphy and Cleveland, 1995) would suggest. Rather, the small increase in R^2 with the addition of rater goals (R^2 change = .01; NS) is consistent with the interpretation that rater climate perceptions may mediate the influence of rater goals on rating behaviour.

In order to test for possible interactions of climate and rater goals on ratings given, cross-product terms were tabulated. Moderation hypotheses were tested on the combined matrix of correlations across all fifteen classes ($N = 523$). When cross-product terms between the two climate scales and the four goals scales are added to rater goals and climate in the regression equation, R^2 increased by a non-significant amount (R^2 change = .01; NS), failing to support the moderation hypothesis.

Table 4.3 Within-Group Homogeneity of Climate Perceptions, Goal Importance and Performance Ratings: Study 3

	F	eta squared	med r_{wg}
Climate			
Decision-making climate	17.8	.29	.77
Co-operative climate	20.9	.32	.95
Ratings of Performance			
Performance ratings	12.9	.24	.85
Goals			
Weakness goals	1.8	.04	.92
Strength goals	2.1	.05	.91
Fairness goals	2.1	.05	.84
Ratee Motivation goals	2.8	.05	.59

Discussion

The results of this study are clear. Measures of climate and performance vary substantially across instructors, whereas measures of goals do not. This is the result that would be expected given our understanding of rater goals as individual differences of the raters themselves, which are not influenced by the ratee about whom the performance judgements are made.

Study 3 also shows that the relation of performance ratings to climate

Table 4.4a Meta-Analysis Results for Antecedents of Instructor Performance Ratings: Study 3

Correlation with Ratings	M_r^a	SD_{weighted}	95% CI	Q
Climate				
Decision-making climate	.42	.18	.35-.49	15.51
Co-operative climate	.46	.15	.39-.53	10.58
Goals				
Weakness goals	-.04	.25	-.13-.04	30.34*
Strength goals	.13	.21	.04-.21	20.95
Fairness goals	.07	.18	-.02-.16	15.10
Ratee Motivation goals	.13	.15	.04-.21	11.02

* $p < .05$ $df=14$, $k=15$, $N=523$

a Mean correlation is calculated using an inverse Fisher Z transformation.

Table 4.4b Meta-Analysis Results for Climate and Goals: Study 3

Correlations	M_r	SD_{weighted}	95% CI	Q
Decision-making climate – Co-operative climate	.64	.22	.59-.69	28.55*
Decision-making climate – Weakness goals	.06	.22	-.02-.14	28.91*
Decision-making climate – Strength goals	.18	.18	.11-.26	19.80
Decision making climate – Fairness goals	.09	.14	.01-.17	12.17
Decision-making climate – Ratee Motivation Goals	.19	.13	.11-.27	10.49
Co-operative climate – Weakness goals	.09	.22	.01-.17	29.16*
Co-operative climate – Strength goals	.33	.20	.25-.40	24.11*
Co-operative climate – Fairness goals	.22	.18	.15-.30	19.73
Co-operative climate – Ratee Motivation Goals	.17	.15	.09-.25	13.33
Weakness goals – Strength goals	.53	.24	.47-.58	33.54*
Weakness goals – Fairness goals	.49	.18	.43-.55	17.87
Weakness goals – Ratee Motivation goals	.26	.19	.18-.33	20.00
Strength goals – Fairness goals	.51	.16	.45-.57	14.31
Strength goals – Ratee Motivation goals	.34	.15	.26-.41	13.22
Fairness goals – Ratee Motivation goals	.29	.14	.22-.37	11.88

* $p < .05$ $df=14$, $k=15$, $N=631$

perceptions is much stronger and more consistent than the relation of rater goal importance to performance ratings, with confidence intervals for the latter relation including zero in the case of two goals: (a) identifying ratee weaknesses and (b) giving ratings fairly.

Our data suggest that co-operative climates support rater goals such as 'identifying strengths' and 'wanting to rate fairly'. Perhaps the link between co-operative climate and identifying strengths is the result of employee appreciation for the manager/instructor's patience and helpfulness (which are likely to be considered strengths). Another partial explanation is that individuals in an environment that they perceive to be mutually helpful are more accustomed to giving positive feedback as part of the normal communication in that setting. Intentions to rate fairly fit with the due process metaphor and may reflect a desire on the part of raters to reciprocate the consideration and patience afforded by

the instructor throughout the semester. The finding of an association between course climate perceptions and rater goals indicates that some types of rater goals may depend upon the context within which the ratings are made and also may result in part from a matching of particular goals to particular rater goals. While the latter explanation is unresponsive of our theory and inconsistent with the results of Study 2 (in which rater goals independent of rater behaviour influenced ratings), the idea that rater goals may change to suit the context is an interesting prospect that merits future study.

One unexpected finding from Study 3 was that climate perceptions appeared to mediate the influence of rater goals on performance ratings. Three alternative explanations might be used to explain these unexpected findings. First, climate perceptions may be an active mechanism through which rater goals impact the evaluations raters give. Specifically, raters who begin with the rating goals of identifying strengths and weaknesses may selectively attend to positive and negative aspects of performance, respectively, in alignment with their goal orientations toward rating. This selective attention and interpretation could result in the emergence of more positive or negative perceptions of course climate, which then contribute to the positive or negative ratings summoned when evaluations are elicited in class. Another explanation for the mediation finding relates to the fact that the effects of rater goals on performance ratings were small even before climate was introduced to the regression equation and that a small decrease in unique R square would be expected even in the absence of a substantive mediation process. Essentially, it did not take much to make the small effect of rater goals become non-significant and the small drop in R square that changed the level of statistical significance of rater goals when climate was introduced could easily be attributed to the low-level multi-collinearity between the climate and goal measures, which were measured with common methods. Finally, for the job being studied (university instructor), it may make sense to think of the measures of climate perceptions as measures of specific aspects of performance. For instance, whether the instructor is patient and listens to students may be considered an aspect of the criteria on which teachers should be rated when overall evaluations are made. If not a job requirement, engendering class participation and helpfulness may at least be considered a form of contextual performance (Borman and Motowidlo, 1993). If the ratings of climate perceptions in this study were indeed considered by students to be specific aspects of the more general construct of 'overall performance', then climate and performance ratings may simply represent two overlapping criteria that are each predicted by rater goals.

OVERALL DISCUSSION

This set of studies has led to several findings regarding the role of goals and context in the performance rating process. First, Studies 1 and 2 test the hypothesis that performance ratings can be better understood if the goals that raters pursue are considered. These studies show that rater goals predict the evaluations they will give; this finding holds up even when goals are measured

well in advance of evaluations of performance. Murphy and Cleveland (1995) predicted strong links between goals and performance ratings; the results of Studies 1 and 2 support this prediction.

Study 3 accomplishes two purposes: first, it provides yet another replication of the rater goals and performance-rating relationship; second, it tests the hypothesis that contextual factors influence performance ratings and that they mediate goal-rating relationships. This study provides evidence that contextual factors do in fact contribute unique variance to performance ratings. However, this study fails to support the hypothesised role of climate as a mediator of the relationship between goals and rating. If anything, our data are more consistent with the hypothesis that goals mediate climate-rating relationships. Future studies should continue to investigate the nature of this goal-climate-rating relationship.

CONCLUSION

The studies discussed in this report have provided researchers with a platform from which several lines of research can be started. Variables such as rater goals and the rating context have been discussed in regard to performance appraisal for quite some time (Murphy and Cleveland, 1995). These studies represent one of the first empirical tests of the influence of these variables. In that vein, these studies do show that goals and context are important. We are confident that future research will serve to further elaborate on the nature of these effects. Given the ubiquitous use of performance appraisal in organisations, it is time that researchers focus on what these ratings mean. The meaning of performance ratings cannot be understood without consideration of the goals that raters pursue when giving ratings and the context in which ratings are given.

Authors

Kevin Murphy is a Professor of Psychology at Pennsylvania State University. He has served as Editor of the Journal of Applied Psychology and President of the Society for Industrial and Organizational Psychology, and has held visiting appointments at University of California, Berkeley, Navy Personnel Research and Development Center, and University of Limerick. He is the author of numerous books, chapters and papers dealing with personnel assessment and performance appraisal.

Jeanette N. Cleveland is a Professor of Psychology at Pennsylvania State University. She has served on numerous editorial boards including the Journal of Organizational Behavior, Academy of Management Journal, Journal of Applied Psychology, Personnel Psychology, and the Journal of Management, and has held visiting appointments at University of California, Berkeley, Occidental College and University of Limerick. She is the author of books, chapters and papers dealing with performance appraisal, gender and diversity, and work and family issues.

Ted Kinney is a graduate student of Industrial/Organizational Psychology at the Pennsylvania State University. His research focuses on performance rating, specifically on

identifying systematic sources of variance contributing to differences in the ratings assigned to the targets of performance evaluation.

Amie Skattebo is a graduate student of Industrial/Organizational Psychology at the Pennsylvania State University. Her research interests include performance appraisal, organisational strategy and human computer interaction.

Daniel A. Newman is a graduate student of Industrial/Organizational Psychology at the Pennsylvania State University. His research focuses on job attitudes, performance measurement and longitudinal and multi-level data analysis.

Hock-Peng Sin is a graduate student in Industrial and Organizational Psychology at Pennsylvania State University. His research interests include performance appraisal and management, work group effectiveness, work motivation, as well as levels and temporal issues in organisational research.

-
1. If time 2 goals are entered first, goals measured at time 1 account for an additional 19.8 per cent of the variance in ratings, which is also significant.

REFERENCES

- Banks, C.G. and Murphy, K.R. (1985) "Toward Narrowing the Research-Practice Gap in Performance Appraisal", *Personnel Psychology*, Vol. 38, pp. 335-45.
- Bernardin, H.J. and Beatty, R.W. (1984) *Performance Appraisal: Assessing Human Behavior at Work*, Boston: Kent.
- Bjerke, D.G., Cleveland, J.N., Morrison, R.F. and Wilson, W.C. (1987) *Officer Fitness Report Evaluation Study* (Navy Personnel Research and Development Center Report, TR 88-4), San Diego: NPRDC.
- Borman, W.C. and Motowidlo, S.J. (1993) "Expanding the Criterion Domain to Include Elements of Contextual Performance", in N. Schmitt and W. Borman (eds) *Personnel Selection in Organizations*, San Francisco, CA: Jossey-Bass.
- Bretz, R.D., Milkovich, G.T. and Read, W. (1992) "The Current State of Performance Research and Practice: Concerns, Directions, and Implications", *Journal of Management*, Vol. 18, pp. 321-52.
- Cianci, A. (2001) "Motivational and Cognitive Explanations for Financial Analysts' Optimistic Earnings Forecasts", *Dissertation Abstracts International Section A: Humanities and Social Sciences*, Vol. 61, No. 12-A, pp. 4845.
- Cleveland, J.N. and Murphy, K. (1992) "Analyzing Performance Appraisal as Goal-Directed Behavior", in G. Ferris and K. Rowland (eds) *Research in Personnel and Human Resources Management*, Greenwich, CT: JAI Press, Vol. 10, pp. 121-85.
- Conway, J.M. and Huffcutt, A.I. (1997) "Psychometric Properties of Multisource Performance Ratings: A Meta-Analysis of Subordinate, Supervisor, Peer, and Self-Ratings", *Human Performance*, Vol. 10, pp. 331-60.
- Cropanzano, R., James, K. and Citera, M. (1992) "A Goal Hierarchy Model of Personality,

- Motivation and Leadership", in B. Staw and L. Cummings (eds) *Research in Organizational Behavior*, Vol. 15, pp. 267-322.
- Dweck, C.S. and Leggett, E.L. (1988) "A Social-Cognitive Approach to Motivation and Personality", *Psychological Review*, Vol. 95, pp. 256-73.
- Gaultney, J.F. and Cann, A. (2001) "Grade Expectations", *Teaching of Psychology*, Vol. 28, No. 2, pp. 84-7.
- George, J.M. and James, L.R. (1993) "Personality, Affect and Behavior in Groups Revisited: Comment on Aggregation, Levels of Analysis, and a Recent Application of within and between Analysis", *Journal of Applied Psychology*, Vol. 78, No. 5, pp. 798-804.
- Gregarus, G.J. and Robie, C. (1998) "A New Look at Within-Source Interrater Reliability of 360-Degree Feedback Ratings", *Journal of Applied Psychology*, Vol. 83, pp. 960-8.
- Harris, M.M. (1994) "Rater Motivation in the Performance Appraisal Context: A Theoretical Framework", *Journal of Management*, Vol. 20, No. 4, pp. 737-56.
- Harris, M.H. and Schaubroeck, J. (1988) "A Meta-Analysis of Self-Supervisory, Self-Peer, and Peer-Supervisory Ratings", *Personnel Psychology*, Vol. 41, pp. 43-62.
- Harter, J. (1986) "Appraisal from Aristotle on", *Management Education and Development*, Vol. 17, No. 1, pp. 34-9.
- Judge, T.A. and Ferris, G.R. (1993) "Social Context of Performance Evaluation Decisions", *Academy of Management Journal*, Vol. 36, No. 1, pp. 80-105.
- Kerst, M. (1993) "Rating Goals as a Function of Appraisal Purpose and Ratee Performance Level", Unpublished Thesis, Colorado State University.
- Litwin, G.H. and Stringer, R.A. Jr (1968) *Motivation and Organizational Climate*, Boston, MA: Division of Research, Graduate School of Business Administration, Harvard University.
- Locke, E. and Latham, G. (1990) *A Theory of Goal Setting and Performance*, Englewood Cliffs, NJ: Prentice Hall.
- Locke, E.A. and Latham, G.P. (2002) "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey", *American Psychologist*, Vol. 57, pp. 705-17.
- Longenecker, C.O., Sims, H.P. and Goia, D.A. (1987) "Behind the Mask: The Politics of Employee Appraisal", *Academy of Management Executive*, Vol. 1, pp. 183-93.
- Miner, J.B. (1984) "The Validity and Usefulness of Theories in an Emerging Organizational Science", *Academy of Management Review*, Vol. 9, pp. 296-306.
- Murphy, K. and Cleveland, J. (1995) *Understanding Performance Appraisal: Social, Organizational and Goal-Oriented Perspectives*, Newbury Park, CA: Sage.
- Murphy, K., Cleveland, J. and Mohler, C. (2001) "Reliability, Validity and Meaningfulness of Multisource Ratings", in D. Bracken, C. Timmreck and A. Church (eds) *Handbook of Multi-Source Feedback*, San Francisco: Jossey-Bass, pp. 130-48.
- Nasir, N.S. (2000) "Points ain't everything: Emergent Goals and Average and Percent Understandings in the Play of Basketball among African-American Students", *Anthropology and Education Quarterly*, Vol. 31, pp. 283-305
- Newman, D.A., Kinney, T. and Farr, J.L. (in press) "Job Performance Ratings", in J.C. Thomas (ed.) *Comprehensive Handbook of Psychological Assessment, Vol. 4: Industrial Organizational Assessment*, New York: John Wiley & Sons, pp. 956-1008.
- Ostroff, C. (1993) "The Effects of Climate and Personal Influences on Individual Behaviors and Attitudes in Organizations", *Organizational Behavior and Human Decision Processes*, Vol. 55, No. 1, pp. 56-90.
- Ostroff, C. and Harrison, D.A. (1999) "Meta-analysis, Level of Analysis, and Best Estimates of Population Correlations: Cautions for Interpreting Meta-Analytic Results in Organizational Behaviour", *Journal of Applied Psychology*, Vol. 84, No. 2, pp. 260-70.
- Roberson, L., Houston, J.N. and Diddans, M. (1989) "Identifying a Valued Work Outcome

- through a Content Analysis of Personal Goals", *Journal of Vocational Behavior*, Vol. 35, pp. 30-45.
- Rousseau, D.M. (1985) "Issues of Level in Organizational Research: Multi-Level and Cross-Level Perspectives", *Research in Organizational Behavior*, Vol. 7, pp. 1-37.
- Thornton, G.C. III (1980) "Psychometric Properties of Self-Appraisals of Job Performance", *Personnel Psychology*, Vol. 33, pp. 263-71.
- Viswesvaran, C. and Ones, D.S. (1995) "Theory Testing: Combining Psychometric Meta-Analyses and Structural Equations Modally", *Personnel Psychology*, Vol. 78, pp. 865-85.
- Viswesvaran, C., Ones, D.S. and Schmidt, F.L. (1996) "Comparative Analysis of the Reliability of Job Performance Ratings", *Journal of Applied Psychology*, Vol. 81, pp. 557-74.

