Fall 1983

Employee Perceptions of Performance Appraisal Systems: Causal Determinants of Fairness, Accuracy, and Acceptability

Michael Dale Secunda
Old Dominion University

EMPLOYEE PERCEPTIONS OF PERFORMANCE APPRAISAL SYSTEMS: CAUSAL DETERMINANTS OF FAIRNESS, ACCURACY, AND ACCEPTABILITY

by

Michael Dale Secunda
B.A. June 1978, Florida State University
M.S. August 1980, University of Central Florida

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

Old Dominion University
September, 1983

Approved by:

Glynn D. Coates (Director)

ABSTRACT

EMPLOYEE PERCEPTIONS OF PERFORMANCE APPRAISAL SYSTEMS: CAUSAL DETERMINANTS OF FAIRNESS, ACCURACY, AND ACCEPTABILITY

Michael Dale Secunda
Old Dominion University
Director: Dr. Glynn D. Coates

The purpose of this research was to develop an initial model of employee perceptions of performance appraisal systems which would integrate available literature and provide a point of departure for future research endeavors. To accomplish these goals, this study had three objectives: (1) integrate the large body of literature to develop constructs that adequately describe employee perceptions of appraisal processes and systems, (2) integrate these constructs into a causal model that is consistent with current literature, and (3) test the model using linear structural modeling.

Seven constructs hypothesized as representing various aspects of employee perceptions were conceptualized and operationalized, and multiple indicators were generated for each construct. Questionnaires containing these items were distributed to two samples--non-exempt employees in a university setting, and police officers in a large metropolitan police department. Confirmatory factor

analyses, which resulted in a six-factor solution that was successfully replicated on a hold-out sample, were used to demonstrate and improve construct validity.

These constructs, as well as several other measures, were integrated into a causal model of employee acceptance of their appraisal systems. This model was then tested using the LISREL V computer program (Joreskog & Sorbom, 1981). Results indicate substantial support for the proposed model, with system acceptability found to be a function of the perceived fairness, accuracy, and use of the appraisal system. Furthermore, both perceived accuracy and fairness varied as a function of the supervisor (i.e., trust in supervisor; supervisor's knowledge of performance), satisfaction with both the content and atmosphere of the performance review session, and of the level of performance rating received. These findings were discussed in terms of limitations, future research directions, and implications for practice.

ACKNOWLEDGMENTS

I would like to thank my dissertation committee members, Drs. Glynn Coates, Raymond Kirby, and Bruce McAfee, for their contributions and support throughout this endeavor. I would also like to thank Dr. Ben Morgan; in many ways he served as a role model, and I have benefited personally and professionally by working with him. Dr. Albert Glickman, in his own inimitable way, has made his mark upon the manner in which I approach/conceptualize/assail both research and broader social problems/issues. He is a profound and prodigious gentleman, and I have gained from his friendship. Above all, however, I owe the major part of my development to Eileen McDonald, who shared these experiences with me during the past five years. I have grown significantly as a result of her guidance, friendship and love.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER	
1. INTRODUCTION	1
2. EMPLOYEE ACCEPTANCE OF APPRAISAL SYSTEMS: A CAUSAL MODEL	26
The Model	26
Constructs	29
Supervisor's Knowledge of Performance	29
Trust in Supervisor	32
Perceptions of Feedback Session Devoted to Goalsetting	32
Feedback Session Atmosphere	33
Perceived Accuracy	34
Perceived Fairness	35
Acceptability	36
Other Variables	36
Formal System Devoted to Improving Performance	37
3. METHOD	40
Subjects and Questionnaire Administration Procedure	40
Organization A	40
Organization B	40

Page

Questionnaires 43

Analysis Procedure 44

Samples 44

Confirmatory Factor Analyses 44

Computation of Construct Scores and Reliability Estimation 48

Path Analysis 49

Summary of Analytical Procedures 55

4. RESULTS AND DISCUSSION 56

Confirmatory Factor Analysis 56

Primary Sample 56

Hold-Out Sample 64

Construct Score Statistics and Reliabilities 65

Path Analyses 69

Model Modification 69

Path Analyses Results - Primary Sample 70

Summary of Path Analysis - Primary Sample 81

Cross-Validation of Path Model 81

5. CONCLUSIONS 84

Summary of Causal Effects 84

Perceptions of the Supervisor 85

Frequency of Evaluation 88

Perceptions of Performance Review Session 88

Acceptability of Appraisal Systems 89

Performance Ratings 90

Page

Limitations of Research Findings 91

Implications for Future Research and Practice 92

REFERENCES 95

APPENDICES

A. Abbreviated Items for Each of the Seven Constructs 104

B. Police Sample Questionnaire 107

C. University Sample Questionnaire 113

D. Factor Loadings for the Six-Factor Model 119

LIST OF TABLES

Table		Page
1.	Summary of Perceptions of Appraisal System Studies	9
2.	Definitions of Appraisal System Constructs	30
3.	Percentage of Respondents in Each Sample by Race and Sex	42
4.	Summary Statistics and Tests of Hypotheses for Factor Analyses (Primary Sample)	59
5.	Summary Statistics and Tests of Hypotheses for Factor Analyses (Hold-Out Sample)	66
6.	Summary Statistics for Items in Both Primary and Hold-Out Samples	67
7.	Intercorrelations and Summary Statistics for Primary and Hold-Out Samples	68
8.	Summary Statistics and Tests of Hypotheses for Path Models	72

LIST OF FIGURES

Figure		Page
1.	Hypothetical Model of the Perceptions of Appraisal System Acceptability	7
2.	Barr, Brief, and Aldag's (1981) Causal Model of Perceived Fairness/Accuracy	20
3.	Hypothetical Model of the Causal Determinants of Employee Perceptions of Appraisal System Fairness, Accuracy, and Acceptability	28
4.	Revised Model of Causal Determinants of Employee Perceptions of Appraisal System Fairness, Accuracy, and Acceptability	71
5.	Final Model of Causal Determinants of Employee Perceptions of Appraisal System Fairness, Accuracy, and Acceptability	76

CHAPTER ONE

Introduction

The evaluation of human performance in organizations has been a topic of great interest and importance to the science and practice of Industrial/Organizational Psychology. A vast amount of research and literature concerning performance appraisal has accumulated over the past several decades; an indication of this extensive interest can be seen by the fact that in the past three years alone (1980-1982), over ninety articles dealing with this general topic have been published in the Journal of Applied Psychology, Organizational Behavior and Human Performance, Personnel Psychology, and the Academy of Management Journal and Review. Based on this extensive body of literature, one could not deny that the appraisal of performance is one of the critical problems organizations must face.

While there has been a tremendous amount of research published relative to performance appraisal, as a whole this literature has been unsystematic and "as subject to fads as any other aspect of personnel research and practice" (DeCotiis & Petit, 1978, p. 644). Historically, the largest part of this literature has been concerned with psychometric aspects of appraisal, primarily with improving the reliability and validity of performance ratings (Jacobs,

Kafry, & Zedeck, 1980; Landy & Farr, 1980). As Landy and Farr (1980) voluminously point out, the extant literature on factors which have been studied as potential influencers of performance ratings includes a lengthy list; among these factors are (a) rater and ratee characteristics such as sex, race, age, and educational level, psychological variables, such as intelligence and cognitive complexity, tenure, etc.; (b) rating formats (BARS, BOS, Graphic), and dimensions (traits, performance results, or observable behaviors), (c) contextual variables, such as intended use of the ratings (administrative, developmental, or research), (d) rating process variables, such as rater training and anonymity, and (e) the determination and reduction of rating effects (leniency, halo, central tendency, differential accuracy, etc.).

Since judgmental indices of performance are highly susceptable to both intentional and inadvertent biases, the extensive research efforts focused on improving the quality of these measures is not without merit. However, these research efforts have been both fragmented and disappointing. For example, most rater and ratee characteristics have been shown to have small, if any, effects on performance ratings (Landy & Farr, 1980). Psychological variables typically have had the same fate, although Landy and Farr report that "cognitive complexity may be an important variable to examine" (p. 78). However, there is now mounting support against its effects on performance ratings (Bernardin &

Boetcher, 1978; Bernardin & Cardy, 1981; Borman, 1979; Sauser & Pond, 1981). Finally, the extensive rater-training literature indicates that training is not always effective in reducing psychometric errors (e.g., Hedge, 1982). However, in this large body of literature there are many conflicting results. Meta-analytic techniques could be used to help clarify some of these inconsistencies.

Based on a perusal of the published literature, it would appear that there is tremendous momentum to continue this mostly psychometric and individual differences line of research. DeCotiis and Petit (1978) state it best:

> ...the continued emphasis in performance appraisal research and publication on the development and advocacy, respectively, of new appraisal formats [is] as though the primary problems in performance appraisal could be solved simply by changes in the format of appraisal. In short, performance appraisal research has focused on instrumentation at the expense of other, perhaps more important, variables. (p. 644)

Researchers in performance appraisal have typically concentrated their efforts on measurement strategies that deal with rating formats and psychometric considerations; there has been a relative lack of research dealing with "non-measurement aspects of appraisal" (Kane & Lawler, 1979, p. 458) such as the more MACRO or systemically-oriented

issues (i.e., the acceptability of performance appraisal systems or the effectiveness of appraisal systems in helping organizations meet their various goals). Since businesses spend tremendous amounts of time, effort, money and other resources on the development, implementation and continued usage of their appraisal systems, it would appear that more research is needed on these larger issues.

Why has there been so much research on measurement issues to the relative exclusion of these systemic ones? Surely, "Acceptable psychometric evidence is a necessary but not sufficient condition for the acceptance and continued use of an appraisal system" (Kavanagh & McAfee, 1982, p. 7). Appraisal systems include numerous other components besides those dealing with formats and psychometrics; these include variables such as the intended use of the system, development and implementation of the system, issues concerning the content and process of the system, and the acceptability of the system to the users. All of these factors combine to determine the effectiveness of the performance appraisal system in meeting its intended goals, whether these goals concern improving personnel and administrative decisions or individual functioning and satisfaction.

While there has been considerable research interest on the performance appraisal issues discussed above, this emphasis on techniques has recently begun to shift towards a concern for more global appraisal issues. One of the attitudinal parameters that impacts the effectiveness of an

appraisal system is user acceptance. It is quite possible that this criterion has dramatic impact on the extent to which a developed system will be used, whether it will be used without attempts to "game" the system, and whether the system will continue to be used once it is implemented (Kavanagh & McAfee, 1982). As Bass and Barrett (1981) noted, in order for any performance evaluation system to be used successfully, it must be both relevant and acceptable to the users. Furthermore, user acceptance and confidence in the system are critical to its effect on employee motivation as well as on management control (Kavanagh, 1981).

Several writers have recently begun to include acceptance as one of their evaluative criteria for appraisal programs. For example, Kavanagh (1981) considers acceptability to be "...one of the most frequently overlooked, and perhaps most important", system criteria (p. 30). Beer, Ruh, Dawson, McCaa, and Kavanagh (1978), reporting on the development, implementation and evaluation of a performance management system for Corning Glass Works, made extensive efforts to tie the performance appraisal facet of the project to all other parts of the personnel system in order to increase the chances for its acceptance and use by the employees. In his discussion of performance appraisal, Cascio (1978) states that in order for any appraisal system to be used successfully, it must be both relevant and acceptable. Other researchers have also suggested that employee opinions regarding the appraisal

process are as crucial to its long-term effectiveness and acceptability as the reliability and validity of the measures themselves (DeCotiis, 1977; Dipboye & de Pontbriand, 1981; Lawler, 1967; Schneier, 1977). To the extent, then, that the appraisal system is not accepted by its users, maladaptive organizational effects could ensue, such as appraisal system "gaming," lower employee morale and motivation, and other factors detrimental to organizational goals. Furthermore, since appraisal systems drive many personnel functions, overall organizational effectiveness could suffer from a lack of system acceptability.

A conceptual model of appraisal system acceptability can be seen in Figure 1. It is clear that acceptability means various things at different levels of the organization. At the organizational level, the system must provide what it was implemented for; that is, whether it was designed to provide information for making administrative decisions (i.e., salary increases, promotions), to improve employee skill levels, or for legal (i.e., validation) purposes, the system should help the organization be more effective. At the supervisory level, the system must be manageable in terms of time and effort requirements and should help develop supervisor/subordinate relations. At the subordinate level, the appraisal process should be relevant to the specific job and must provide intrinsic and/or extrinsic benefits, depending on its purpose. It is this last level which is the concern of this research project.

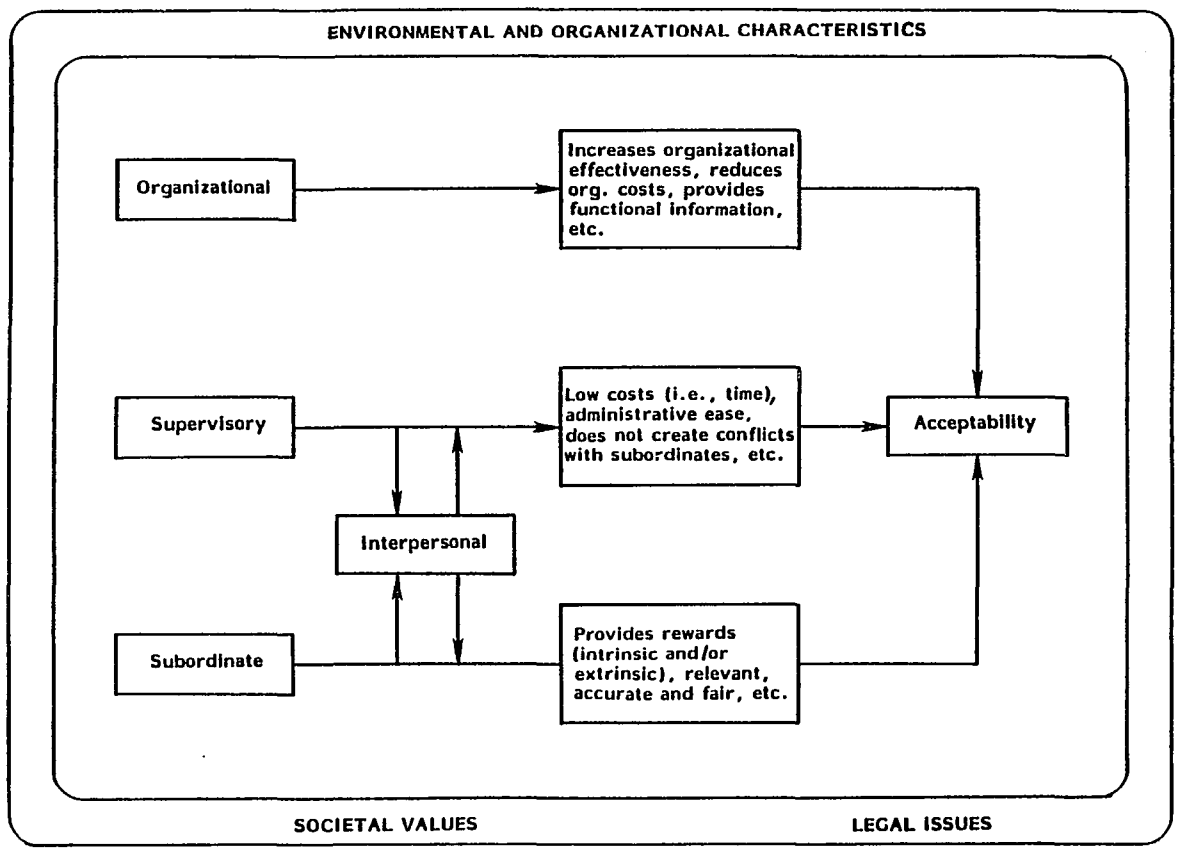

Figure 1. Hypothetical model of the perceptions of appraisal system acceptability.

An important issue for research that could help improve the current state-of-the-art in performance appraisal research would be an identification of those parameters that determine employees' perceptions of performance appraisal acceptability. In fact, some progress in this area has already been made. Table 1 presents a summary of the available empirical research that has examined employee perceptions of either perceived fairness/accuracy of ratings or the acceptability of appraisal systems. The fact that only 5 studies are contained in this table demonstrates that this is an area fertile for additional conceptualization and research. In a study concerning the measurement aspects of various types of performance ratings, Lawler (1967) proposed a hypothetical model of the factors that affect the reliability and validity of performance appraisal systems. His model indicates that attitudes towards the fairness and acceptability of rating systems are a function not only of objective content characteristics (such as who rates and what rating dimensions are used), but also of process and contextual characteristics such as organizational and individual-differences variables. Lawler also stated that the ultimate success of a performance evaluation system depends on the confidence of the person being evaluated in the evaluation process itself; this confidence can result from such process and contextual factors as meaningful interactions with significant others (i.e., supervisors) and participation in various systemic characteristics.

Table 1

Summary of Perceptions of Appraisal System Studies

Authors	Sample and size	Criteria	Significant predictors	R^2
Landy, Barnes, & Murphy (1978)	Manufacturing (711)	Fairness/Accuracy	Formal Program	
			Frequency of Evaluation	
			Supervisor's Knowledge	
			Opportunity to Express Feelings	
			Action Plans made	
			Goalsetting	.29
Barr, Brief, & Fulk (1981)	R & D Engineers (198)	Fairness/Accuracy	Trust in Supervisor	
			Action Plans made	
			Supervisor's Knowledge	.42
		Action Plans made	Supervisor's Knowledge	.13
		Trust in Supervisor	Supervisor's Knowledge	.15

(table continues)

Table 1 (Continued)

Authors	Sample and size	Criteria	Significant predictors	$\underline{R}^2$
Dipboye & de Pontbriand (1981)	R & D (474)	Opinion of Appraisal	Favorability of Rating (accuracy)	
			Relevance of Dimensions	
			Opportunity to Participate	.31
		Opinion of System	Favorability of Rating (accuracy)	
			Goalsetting	
			Relevance of Dimensions	
			Opportunity to Participate	.16

(table continues)

Table 1 (Continued)

Authors	Sample and size	Criteria	Significant predictors	R^2
Vance, Winne, & Wright (1982)	Police (94)	Fairness/Accuracy[a]	Received Ratings Expected	
			Supervisor's Knowledge	.66
		System Acceptability	Goalsetting	
			Adequacy of Dimensions	
			Received Ratings Expected	
			Job Provides Adequate Feedback	.48
Kavanagh, Hedge, & DeBiasi (1983)	Health Care (323)	Acceptability of Form	Fair/Accurate Appraisal	
			Satisfied with Feedback	
			Understand Standards	.47

[a] Fairness/Accuracy and System Acceptability correlated .49.

Building upon this model proposed by Lawler (1967), a study conducted by Landy, Barnes, and Murphy (1978) was directed towards identifying various attitudinal correlates of perceived fairness/accuracy of performance evaluations. In particular, they were not concerned with the physical characteristics of the rating instrument, but were more interested in the perceptions of system fairness and accuracy in terms of descriptive information regarding rating system processes and reactions to contextual system characteristics. Data were gathered from 711 exempt employees in a large, multi-division manufacturing organization using a 12-item questionnaire concerning the frequency, quality, and consequences of performance evaluation. A single-item measure assessing the fairness and accuracy of evaluation was designated the dependent variable. Results from their multiple regression analysis (see Table 1) indicated that perceptions of fairness/accuracy of evaluations were significantly related to (1) the presence of a formal evaluation program, (2) the frequency of evaluation, (3) the perception that raters were familiar with an individual's performance levels, (4) the perceptions that the subordinate had an opportunity to express his/her feelings when evaluated, and (5) the perception that goalsetting took place. In addition, a reanalysis of their data indicates that the development of action plans designed to improve weaknesses also entered the regression equation ($p < .01$). Furthermore, three other variables (agreement with supervisor

on duties; feelings when performance criticized; salary discussed during evaluation?) correlated significantly with the dependent measure, but did not enter the regression equation due to moderate intercorrelations within the set of independent variables. Added support for these results is indicated by a cross-validation of their regression equation on a hold-out sample which resulted in minimal shrinkage (primary sample $\underline{R}$ = .54; hold-out sample $\underline{R}$ = .51). Finally, Landy, Barnes-Farrell, and Cleveland (1980) tested the hypothesis that those employees who had more favorable (higher) ratings were also those who had more positive attitudes towards the evaluation process. If this were true, it would imply that if a rater was interested in increasing favorable attitudes towards the system he should be more lenient than accurate in his employee evaluations! Their results indicate that performance rating levels did not moderate the perceptions of system fairness/accuracy.

In a more recent article, Dipboye and de Pontbriand (1981) studied the correlates of employee reactions towards both performance appraisals and appraisal systems. They cogently argued that an important research issue is the determination of factors that influence both the appraisal itself and the system such that these factors could be modified to increase employee acceptability. They specifically predicted that employees would have favorable impressions towards their last appraisal and the system of appraisal to the extent which (1) they were allowed to

participate in the feedback session, (2) they perceived the ratings factors to be job-related, (3) the appraisals were goal-oriented, and (4) they perceived their ratings to be favorable. Data were collected from a sample of 474 scientists, engineers, and technicians employed by a research and development firm. A three-item composite relating to opinions of their latest appraisal and a four-item composite concerned with opinions of the appraisal system were designed the dependent variables. Six other items measuring perceived appraisal session participation, employees' perceptions of the job-relatedness of the rating factors, goal orientation of the appraisal process, and perceived favorability of the appraisal were designated the independent variables. Results indicate that perceived relevance of the job factors, perceived discussion of plans and objectives, and the perceived opportunity to express one's own side in the performance feedback session are all significantly related to both opinions of the appraisal and opinions of the system.

Perceived favorability of the appraisal also significantly related to both criteria used in this study. Dipboye and de Pontbriand interpret this to mean that an employee's opinion of the appraisal and of the system tended to be negative to the extent that performance feedback was experienced as being negative. It should be noted, however, that the item used to measure this variable dealt not with the LEVEL of the rating received, but with the perceived ACCURACY of measurement (i.e., the congruency between the

rating received versus what rating was expected). Therefore, this result should not be interpreted (as they did) to mean that employee opinions of the appraisal and appraisal system depend upon the level of rating received; instead, the correct interpretation is that opinions of the appraisal and appraisal system were moderated to the extent the subordinate felt his/her rating was inaccurate (i.e., received a rating that were not in agreement with what was expected).

The results also indicate that employees are more receptive of negative feedback if they are allowed to participate in the feedback session, if plans and objectives are discussed, and if they feel they are being evaluated on factors relevant to their work. Dipboye and de Pontbriand (1981) state, "Although negative feelings may not be eliminated entirely, actions on the part of the supervisor to enhance these perceptions of the appraisal process may increase employee acceptance of the feedback and the appraisal system" (p. 251).

Finally, although Dipboye and de Pontbriand report that goalsetting was not a significant predictor of the acceptability of appraisals and appraisal systems, questionable analytical techniques leave this interpretation suspect. An inspection of the interitem correlation matrix indicates significant correlations between goalsetting and both criteria. In addition, multicollinearity in the independent variables suggest that this nonsignificant result may be misleading.

In a study designed to investigate correlates of performance appraisal system acceptability, Kavanagh, Hedge and DeBiasi (1983) obtained both process and content data relating to employee perceptions and attitudes towards fairness of the system, the frequency and accuracy of appraisal, and the quality and quantity of feedback obtained through the use of their system (these items were similar in nature to those used in Landy et al., 1978). Acceptability of the current appraisal form was designated the dependent variable. Data were collected from 323 administrative, professional, and clerical employees of nursing departments in two urban hospitals where a new performance appraisal system was being developed and implemented. Results from a multiple regression analysis indicate that perceptions of a fair and accurate appraisal was the most significant predictor of the criterion; in addition, satisfaction with feedback and a clear understanding of performance standards were also significantly related to acceptability of the form. Confidence in the results is strengthened by a shrinkage estimate of only .ØØ3 (from $\underline{R}$ = .684 to $\underline{R}$ = .681) in the multiple correlations between a primary and a hold-out sample respectively.

Vance, Winne, and Wright (1982) investigated the correlates of rater and ratee reactions to a performance appraisal system in a large metropolitan police department. A survey assessing perceptions of fairness/accuracy of ratings, a number of attitudinal measures concerning

perceptions of goalsetting, feedback session atmosphere, and other contextual and process variables related to reactions to the rating system were collected from 94 police officers with the rank of officer, corporal, or sergeant. In addition, performance ratings were available. Dependent variables included a single-item measure of perceived fairness/accuracy of ratings; acceptability of the rating system was measured by a five-item composite with an internal consistency reliability of .87.

Hierarchical regression entering performance ratings first, followed by a stepwise procedure for the remaining independent variables, resulted in nonsignificant effects for performance ratings. That is, there was no relationship between performance ratings and perceptions of either fairness/accuracy or system acceptability. This result agrees with the findings reported by Landy et al. (1980) as well as with the reinterpreted findings of Dipboye and de Pontbriand (1981). The contextual variable concerned with whether the subordinate received the rating expected was the best predictor of fairness/accuracy. In addition, the process variable relating perceptions of the supervisor's opportunities to observe the ratee's job behavior was also significantly related to perceived fairness and accuracy.

Vance et al. (1982) also report results for the determinants of appraisal system acceptability. Perceptions of goalsetting was the most significant predictor of system acceptability, followed by rating factor adequacy,

expectations of ratings, and the degree of feedback provided. As Vance et al. discuss, these results support Dipboye and de Pontbriand's (1981) finding that relevance of rating factors and discussion of plans and objectives were related to favorable opinions of appraisal systems. Finally, the two dependent measures used in this study--perceived fairness/accuracy and system acceptability--correlated .49 ($p < .005$).

Barr, Brief, and Fulk (1981) also studied the correlates of perceived fairness/accuracy of performance appraisal systems. Their research was an attempt to cross-validate Landy et al.'s (1978) findings. A refinement of Landy et al.'s model was also made in that dynamic relationships among five of Landy's evaluation process variables were incorporated into a causal model. In addition, a variable that reflects the quality of the overall relationship between the rater and ratee (what they call Trust in Supervisor) was included in the path model. It was hypothesized that trust in the supervisor would be an important source of perceived fairness and accuracy in that fairness and accuracy perceptions are likely to arise not only from formal system characteristics, but also from the manner in which the supervisor and the subordinate use the performance appraisal system. As Huse (1967) states, in a situation which is as potentially sensitive as performance evaluation, the establishment of a climate of trust is critical to the skillful use of the formal system.

Barr et al.'s (1981) sample consisted of the 198 research and development engineers in a large electronics firm. The five process variables reported in Landy et al.'s (1978) article were used (formal program, evaluation frequency, opportunity to express feelings, supervisor's knowledge, and plans related to performance), as well as Roberts and O'Reilly's (1974) three-item Trust in Supervisor scale.

The cross validation of Landy et al.'s findings resulted in nearly identical estimates of R in this setting. Therefore, it was concluded that Landy's model is accurate.

The refined (causal) model was tested by the method outlined in Duncan (1966) and was followed by Heise's (1969) "theory trimming" procedure. This method involves the identification and deletion of nonsignificant paths--essentially creating a new, more accurate model. Finally, the degree of consistency of this trimmed model was assesed by the method recommended by Kerlinger and Pedhazur (1973). Results of these procedures can be seen in Figure 2.

The path analysis procedure showed significant paths from supervisor's knowledge of subordinate's performance to the development of action plans and from there to perceived fairness/accuracy. Also, the hypothesized paths between supervisor's knowledge and both trust in supervisor and perceived fairness/accuracy were significant. Finally, a significant path was found between trust and perceived fairness/accuracy. In a personal communication with the

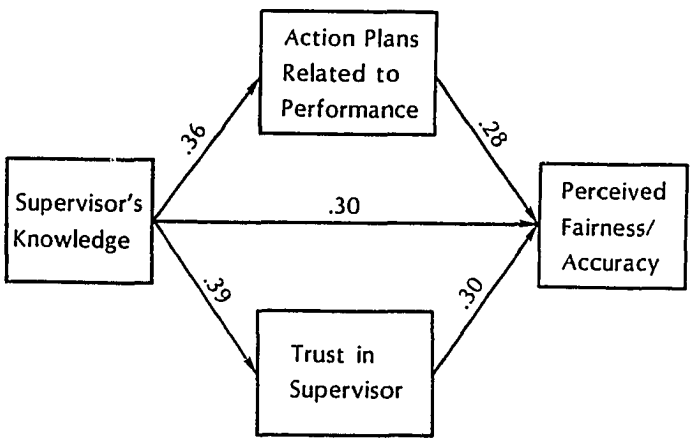

Figure 2. Barr et al.'s (1981) causal model of perceived fairness/accuracy.

primary author, it was discovered that there were other significant paths in the original model (i.e., between trust and opportunities to express feelings; between evaluation frequency and opportunity to express feelings; between formal program and frequency of evaluation; etc.). However, insignificant intermediate links resulted in their being dropped from the model. It is possible that there was insufficient power to detect true differences in this case, either due to the small sample size, to inadequacy of the item (construct validity), or to possible low reliability of the single-item measures.

Results from this study indicate that this refined model explains significantly more variance in the perceptions of fairness and accuracy (42%) than does Landy et al.'s (1978) model (29%). In addition, the inclusion in the model of trust, along with Landy et al.'s "plans related to performance" and "supervisor's knowledge of performance" indicate that supervisor/subordinate readiness to implement the performance appraisal process appropriately is a particularly important factor in determining perceptions of fairness and accuracy. The revised model also suggests that a subordinate's perceptions of the supervisor's knowledge, in addition to being directly related to perceived fairness and accuracy, act indirectly through plans related to performance and trust in supervisor. Thus, supervisor's knowledge seems to have many points of influence.

Several shortcomings or inadequacies can be seen in previous research on perceived fairness, accuracy, and acceptability of performance appraisal systems. The first shortcoming is concerned with basic problems in the measurement of variables.

Most of the past research used single-item measures to quantify the variables studied; there are obvious shortcomings inherent in this approach (Nunnally, 1978). Not only does the possible low reliability of single-item measures limit content and predictive validity, but such a limited sampling of the construct domain leads one to question whether the item actually measures the underlying theoretical construct (Drasgow & Miller, 1982).

The use of single-item measures in past research may have also inadvertently hindered a more indepth conceptualization of the constructs involved in perceptions of performance appraisal and the processes by which these constructs may be intertwined (or, as Cronbach and Meehl, 1955, describe it, the nomological net). The process of identifying multiple items, or indicators, as they are frequently called (see, for example, Sullivan & Feldman, 1979), forces the conceptualization of a common denominator between items chosen to measure the same "thing." This common denominator is the essence of the hypothetical construct that one is interested in investigating. By viewing these multiple items as indicators of higher-level constructs, one can begin to understand better the larger

issues that are relevant to perceptions of the appraisal system, and perhaps even the processes that relate one to the other. For example, several of the items used in Landy et al. (1978) measure perceptions of events that are related to goalsetting within the performance review session (Action plan developed? Action plan related to performance weaknesses? Progress on goals reviewed?). Several other items measure the atmosphere of the review session (Opportunity to express feelings? Feelings when performance criticized?). Therefore, by considering these items as indicators of constructs, one can begin to think in terms of larger issues which can be useful for developing conceptual models.

Taking a construct, rather than single-item perspective, may even lead to a reevaluation regarding the relationships between variables. For example, Vance et al. (1982) found their item "received rating expected" to be the best predictor of the criterion "fairness/accuracy of ratings." This is not at all surprising, however, since the predictor "received rating expected" can easily be considered an indicator (rather than a predictor) of their criterion. The best predictor of a construct should be one of its own indicators, and in their case, this is precisely the outcome. Therefore, Vance et al. may be confusing the boundaries between the predictor and criterion spaces.

Finally, this process of construct development in performance appraisal has the added benefit of assisting in

the development of conceptual models which could direct and/or focus both applied and research efforts in a more organized way than has occurred in the past. While models exist for such subtopics as rating accuracy (DeCotiis & Petit, 1978) and rating processes (Landy & Farr, 1980), no such model has been developed that identifies and integrates the various factors that potentially influence the perceptions of appraisal system acceptability.

To summarize, even though there is an extensive body of literature in the area of performance appraisal, very little has been written concerning employee perceptions of either the appraisal itself or the system in general. Furthermore, the available research has not been of particularly high quality. This research has focused on two different criteria: correlates of perceived fairness/accuracy and correlates of performance appraisal acceptability. The research on fairness/accuracy cited above have treated both constructs as a single-item dependent measure; that is, the typical question asked of subjects was, "Has performance been fairly and accurately evaluated?" However, there is no reason to assume that fairness and accuracy are identical constructs. Perceived accuracy of the performance rating received by a subordinate corresponds most likely to a discrepancy between what rating he/she believes is deserved and what rating was given by the supervisor. Perceived fairness is a function of (a) the ratings that the subordinate believes he/she deserved, (b) perceived ratings

obtained by others (normative), and (c) the use made of the ratings. Therefore, the research described here will treat these constructs separately.

The construct of perceived fairness/accuracy has been causally treated as a dependent variable, and performance appraisal acceptability has independently been shown to be related to fairness/accuracy. However, there has been no research to combine this information causally to show the correlates of acceptability as moderated through perceived accuracy and fairness. Based on the literature to date, this would seem a plausible undertaking. Therefore, the research outlined herein will develop and test a causal model of the perceived correlates of performance appraisal acceptability. Regardless of whether the appraisal system is used for administrative or developmental feedback purposes, knowledge of the causal determinants of system acceptability has direct implications since the overall effectiveness of the system (and therefore its utility) will depend on how acceptable it is to the users.

CHAPTER TWO

Employee Acceptance of Appraisal Systems: A Causal Model

The purpose of this research was to develop and test a literature-based causal model of the determinants of performance appraisal system fairness, accuracy, and acceptability (from the employee's viewpoint). Many researchers have concluded that a model of some sort will be necessary in the area of performance assessment before any significant advances will be made in understanding and controlling this process in organizations (DeCotiis, 1977; DeCotiis & Petit, 1978; Jenkins & Taber, 1977; Kane & Lawler, 1978). Such systematic efforts are needed to tie the available literature together and provide a framework from which to view performance appraisal in more systematic terms. This dissertation research represents an initial attempt at such an endeavor. The model and its constructs are described below.

The Model

The model depicted in Figure 3 demonstrates the hypothesized constructs and causal paths which determine performance appraisal acceptability. This section briefly describes the overall model; the following section describes each construct individually.

A subordinate's feelings about system Acceptability are determined by both perceptions of Fairness and Accuracy, which are in turn determined by various dyadic relationships with his/her supervisor and formal system processes. The dyadic relationships are factors that typically occur around, but are not necessarily within, the performance feedback session. Included here are the constructs Feedback Session devoted to Goalsetting (FBGS), feedback session Atmosphere, Trust in Supervisor (Trust), and Perceptions of the Supervisor's Knowledge of Performance (Knowledge).

In addition, the perceived Use of the appraisal system, as well as the Frequency of appraisal, are seen to influence factors within the feedback session. The model also indicates that to the extent a subordinate perceives his supervisor to be knowledgeable about his strengths and weaknesses, the appraisal system will be seen as both fair and accurate. This effect is also moderated by positive supervisor/subordinate relationships (i.e., Trust), since the establishment of a climate of trust is critical to the skillful use of the formal system (Huse, 1967). The direct effect for positive dyadic relationships is consistent with the findings of Ilgen, Fisher, and Taylor (1979) who reported that the acceptance of feedback depends on source credibility, of which one dimension is the recipient's trust in the source's motive (Barr et al., 1981, p. 156). In addition, the effect of supervisor's Knowledge impacts on perceived Fairness and Accuracy indirectly through plans that

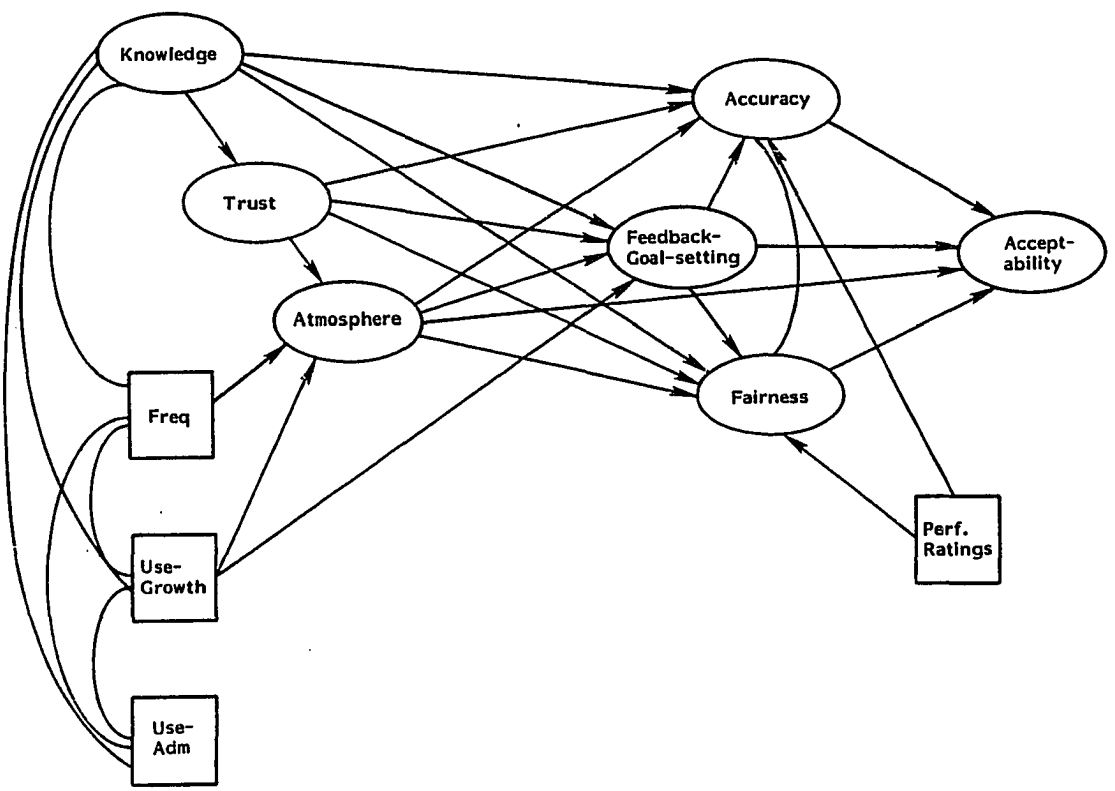

Figure 3. Hypothetical model of the causal determinants of employees' perceptions of appraisal system fairness, accuracy, and acceptability.

are made in the feedback session related to improving performance deficiencies. The potential for two-way dyadic interactions within the feedback session (i.e., feedback session Atmosphere) also is hypothesized to impact on perceived Fairness and Accuracy, as well as impacting directly on overall system Acceptability. The opportunity for performance appraisal feedback sessions are of course moderated by the Frequency of evaluation. Finally, the Acceptability of the system is seen as a direct effect of the opportunity for two-way discussions (Atmosphere) and the perceived Accuracy and Fairness of the appraisal system.

Constructs

The constructs identified for this study were developed based on sound psychometric and theoretical principles. By using these constructs instead of single-item measures as has typically been done in the past, chances for improved reliability, content and construct validity, specificity, and discriminability are increased (Joreskog & Sorbom, 1981; Nunnally, 1978). Table 2 presents a definition for each of the constructs identified in the model of employee perceptions of the fairness, accuracy, and acceptability of appraisal systems. What follows is a brief description and rationale for each construct.

Supervisor's Knowledge of Performance (Knowledge). This construct is an extension of an item used by Landy et al. (1978) of the same name. However, its conceptualization has been broadened somewhat to include not only the

Table 2

Definitions of Appraisal System Constructs

Construct	Definition
Knowledge	Perceptions of supervisor's awareness of job requirements and performance.
Trust	The quality of the interpersonal relationship between a subordinate and her supervisor.
Atmosphere	The quality of the interpersonal communications within the performance review session.
FBGS[a]	Extent to which a subordinate accepts as reasonable both performance feedback and future goals.
Accuracy	Discrepancy between performance rating received and performance rating expected.
Fairness	A belief that supervisors will rate job performance in an unbiased manner.
Acceptability	Satisfaction with appraisal system content, process, and outcomes.

[a]FBGS = Feedback - Goalsetting.

perception of a supervisor's knowledge of performance, but also a belief by the subordinate that the supervisor knows the requirements of successful job performance.

In terms of the inclusion of this construct in the causal model, DeCotiis and Petit (1978) state that the accuracy of ratings is a function of various rater abilities, two of which are the supervisor's opportunities and ability to observe ratee job behavior. Landy and Farr (198Ø) also report that the accuracy of rating is partially a function of the rater's knowledge of the ratee's performance and his job. Therefore, a path from this construct to Accuracy is included in the model. In support of this path, Landy et al. (1978), Barr et al. (1981), Dipboye and de Pontbriand (1981), and Vance et al. (1982) have shown an item measuring this construct to be a significant predictor of fairness/accuracy. Based upon this research, a path has also been added to Fairness. Barr et al. (1981) also suggest that greater supervisor's knowledge of subordinate performance provides a stronger basis for a constructive helping role. Therefore, it is also expected that this relationship with Fairness and Accuracy is partially moderated by favorable perceptions of dyadic relationships with the supervisor (i.e., the Trust and FBGS constructs), as has causally demonstrated by Barr et al. (1981). For this reason, paths to these two constructs have also been included. (The items used to measure this and the following constructs can be found in Appendix A.)

Trust in Supervisor (Trust). This construct is included in the model since perceptions of fairness, accuracy, and acceptability are likely to arise not only from formal system characteristics, but also from the manner in which the supervisor and subordinate use the performance appraisal system. Since evaluation can be very sensitive, the establishment of a positive relationship is seen as critical to the skillful use of the formal system (Huse, 1967).

Barr et al. (1981) have shown the significant causal effect this construct has on Fairness/Accuracy. This construct also has strong empirical and theoretical support between the constructs of Accuracy (Graen & Schiemann, 1978; Roberts & O'Reilly, 1974), FBGS (i.e., acceptability of feedback, Ilgen, 1981), and Atmosphere (Ilgen, Peterson, Martin & Boescher, 1981). Therefore, causal paths from this construct to both Fairness and Accuracy of ratings, as well as to the Atmosphere of and processes within the performance review session, are hypothesized.

Perceptions of Feedback Session Devoted to Goalsetting (FBGS).

This is a multidimensional construct assessing satisfaction with both the feedback and goalsetting components in performance appraisal. The literature in both these areas is extensive; only a cursory review in terms of the proposed model will follow.

The results of Landy et al.'s (1978) research indicate that the discussion of future action plans (i.e.,

goalsetting), especially if these action plans were related to performance weaknesses, significantly predicted perceived fairness/accuracy of ratings. In addition, Barr et al. (1981) also reported a significant path from "action plans related to performance weaknesses" to perceived fairness/accuracy of ratings. Therefore, paths from FBGS to both Fairness and Accuracy are included in the proposed model.

Dipboye and de Pontbriand (1981) report that the opportunity to participate in discussions within the performance review sessions as well as the discussion of plans and objectives for the future, were significant predictors of opinions of the appraisal system. Vance et al. (1982) indicate that a multi-item measure of feedback and goalsetting processes and outcomes was significantly related to appraisal system acceptability. Finally, Kavanagh et al. (1983) reported that satisfaction with feedback obtained in the performance review sessions correlated significantly with their criterion of appraisal form acceptance. Therefore, a path to Acceptability has also been included in the model.

Feedback Session Atmosphere. This construct is an extension of Landy's "Opportunities to Express Feelings when Evaluated." The outcomes of the performance review session are a function not only of interactions between characteristics of the supervisor, subordinate, and performance levels, but also of the atmosphere of the review session. Since the atmosphere is partially under control of

the supervisor, he should attempt to create as favorable an environment as possible in order to carry out the functions prescribed for the review session. (This might entail taking a helping rather than an evaluative attitude, and she might actively encourage subordinate participation.) Clearly, a more relaxed atmosphere should be more constructive than a tense one.

Landy et al. (1978) found that "opportunities to express feelings" during the performance review session was a significant predictor of fairness/accuracy. Dipboye and de Pontbriand (1981) also reported this item to be significantly related to opinions of both the appraisal and of the appraisal system. Therefore, paths have been included to the four constructs FBGS, Accuracy, Fairness, and Acceptability.

Perceived Accuracy. As noted earlier, previous researchers have considered the accuracy and fairness of ratings to be a single construct. One of the hypotheses of this model is that these two should be treated as separate constructs. As discussed earlier, perceived accuracy of rating is a function of the discrepancy between the rating received and the rating the subordinate believes is deserved.

The accuracy of ratings has received extensive attention in the research literature. DeCotiis and Petit (1978) have developed a model of the determinants of the accuracy of performance ratings and cogently integrated a wide body of literature. Landy and Farr (1980) treat accuracy at great length, and Borman's (1977; 1980) work in this area is

perhaps the most systematic. Ilgen et al. (1979) discuss the effects of feedback on accuracy, and Graen and Scheimann (1978) as well as Roberts and O'Reilly (1974) discuss the impact of positive relationships with the supervisor on this construct (thus adding further support for these previously documented paths). Vance et al. (1982) have shown that perceptions of fairness/accuracy of performance ratings to be the best predictor of acceptability. In addition, Dipboye and de Pontbriand (1981) found that "favorability of rating" (reinterpreted above as perceived accuracy) was the best predictor of opinions of the appraisal system. Therefore, a direct causal path from accuracy to acceptability is included in the model.

Perceived Fairness. In contrast with perceived accuracy, this construct deals more with an individual's comparative processes concerning ratings received versus ratings others received (whom are perceived by the individual to have similar performance levels). For example, if Person A believes his performance levels to be identical to that of Person B, but received a lower rating, then the rating (and/or the system) would be perceived as unfair. This should be especially true if organizational rewards are dispensed on the basis of these ratings. In other words, if an employee was to compare across raters, then he/she might find that other workers are receiving higher ratings then him/herself, but whose job performance is not any better; in this case, different ratings are simply a function of rater

individual differences. To the extent that the organization does not take such differences into account when dispensing rewards (raises, promotions, etc.), perceived fairness of the system will suffer. This construct is also hypothesized to affect perceptions of system acceptability directly.

Acceptability. This construct was discussed at some length earlier. Lawler (1967) has stated that the acceptability of performance appraisal systems is a function of systemic, process, and contextual factors. Kavanagh and McAfee (1982) report that this appraisal criterion may be one of the most frequently ignored concepts in the literature; indeed, "...if users do not accept the appraisal system, they will either not use it or use it improperly, resulting in the failure of the system" (p. 11). Cummings (1976) has noted that an appraisal system with standard procedures for providing employees with feedback was found to be more acceptable than another system without the feedback component.

Items used to measure this construct are concerned with perceptions of whether the forms, ratings, and system all contribute to employee acceptance of the appraisal process as a whole.

Other Variables. Performance ratings were collected to test further the hypothesis that rating levels moderate reactions to appraisal systems (specifically, Fairness and Accuracy). Three one-item measures were also included in the model, one asking the frequency of evaluation, and two others

concerned with the perceived Use of the system. They are discussed below.

Formal System Devoted to Improving Performance. The two items used here are a conceptual extension of a measure in Landy et al. (1978)--"Existance of a Formal Appraisal System." In this context, however, it is the perceived use of the system, not the mere existance of one, that is hypothesized to impact processes in the feedback/goalsetting session. Appraisal systems developed for growth and development (Use - Growth) should causally impact fairness, accuracy and acceptability (through indirect paths) in a more favorable manner than will systems designed for administrative purposes (Use-Adm). That is, systems perceived to be used for growth and development will set a more positive environment in the performance review session than will systems perceived used for administrative purposes.

None of the available literature describing the differential effects of system use pertains to employee perceptions of either fairness/accuracy or system acceptability. Most of it is, in fact, related to the impact on accuracy of scores. In this vein, Landy and Farr (1980) state that the purpose component of performance appraisal systems is of central importance in that it differentially affects the rating process. They report that the intended use of the ratings has impact on various psychometric properties such that ratings done for administrative purposes tend to be more lenient (and therefore, less accurate).

DeCotiis and Petit (1978) report that this impact on accuracy can be explained in terms of the relationship between the purposes of appraisal and the likelihood of appropriateness of confidentiality (Bartlett & Sharon, 1969), rater role conflict (Dayal, 1969), and the possibility of negative impacts on future ratee outcomes (Dwyer & Dimitroff, 1976). Numerous other researchers have noted that appraisals conducted for employee development are likely to be more accurate than administrative appraisals, but less accurate than appraisals conducted for personnel research (DeCotiis & Petit, 1978; Maier, 1963; Meyer, Kay & French, 1965).

In the context of this research project, it is not expected that employee perceptions of rating accuracy will be directly affected by their perceptions of the intended use of the system. However, an indirect effect through outcomes of the performance review session is hypothesized for systems perceived used for growth and development. Note that no impact on any of the constructs included in this model are expected for systems perceived as used for administrative purposes.

Before discussing research methodology in the following chapter, it should be noted that the conceptual development of the constructs discussed above, as well as the hypothesized structure between them, has been guided by a desire to integrate the available literature and to increase our understanding of the influences and dynamics involved in perceptions of appraisal system acceptability. In no way is

it to be construed that these are the only constructs relevant to this issue, or that the causal structure proposed is the only plausible one. The constructs included are based on the author's belief that these are the more significant ones that impact perceptions of fairness, accuracy, and acceptability. No doubt there are others that should be included. Also, while the paths are both intuitively appealing and literature-based, the fact is, there is not much literature upon which to base these decisions. No doubt this state of affairs will change in the future as researchers tire of "flogging dead horses" (i.e., persuing research on rating formats, etc.). Therefore, this research is seen as partly exploratory in the sense of being the initial step in the development of constructs and models relevant to employee perceptions of appraisal systems.

CHAPTER THREE

Method

Subjects and Questionnaire Administration Procedure

Data for this research were collected from members of two organizations. A description as well as the data collection procedures for each organization are presented below.

Organization A. This organization was a police department located in a large metropolitan area in the Southeast. Questionnaires were sent to a stratified sample of 185 officers with the rank of police officer, corporal, or sergeant, and were returned by sealed envelope to a departmental contact. Of the 185 questionnaires distributed, 159 were returned, for a response rate of 87%. Of these, three were discarded as incomplete. In order to obtain performance ratings, respondents were given the option of providing their badge numbers, which could be used to identify their performance ratings from personnel files. Of the 159 questionnaires returned, 132 (83%) chose to provide this information. Although identifying information was requested, confidentiality of individual survey information was promised and maintained. These questionnaires were distributed within two weeks of each respondent's performance review session, and were returned within a two-week period following their distribution.

Organization B. This organization was a large university located in the same metropolitan area as the sample described above. In this organization, questionnaires were distributed to 700 classified employees located throughout all departments. These questionnaires were sent through interdepartmental mail to each employee; they returned the completed survey by way of interdepartmental mail to the researcher. Of the 700 questionnaires mailed, completed surveys were returned by 291 employees for a response rate of 42 percent. (Typical response rates for surveys in this organization average approximately 35 percent.) Of these 291 questionnaires, 31 were discarded as incomplete, for a within-sample total of 260 valid surveys, and a total between-sample size of 416. Because this organization would not allow questions which might potentially be used to identify respondents, performance ratings were obtained by asking each person to provide their latest rating, if they so desired. Of the 260 valid questionnaires returned, 194, or 75 percent, had this information. As in the previous sample, confidentiality was promised and maintained.

Demographic information for the two samples is presented in Table 3. As mentioned above, constraints were placed on survey design in the university sample in order to guarantee respondent anonymity. Therefore, any information which might be used to identify the respondents (i.e., age, sex, race) were not included in the questionnaire. The information presented in Table 3 pertains to the population of classified

Table 3

Percentage of Respondents in Each Sample by Race and Sex

	Sample	
Group	Police (n = 156)	University (n = 748)
Male		
Black	6%	10%
White	73%	23%
Female		
Black	0%	27%
White	4%	38%
Other or Unknown	17%	2%

Note. Percentages for the university sample are for the population of all classified employees from which the sample of respondents was drawn.

employees from which this sample was drawn. For these data, the figures are
provided for general information only; whether the values for the actual sample obtained match these figures is unknown and cannot be determined.

Questionnaires

Based on an inspection of questionnaires used in both published and unpublished studies, items were either adapted from these questionnaires or generated which measured each of the hypothesized constructs. Multiple items were generated to measure each of the constructs discussed in Chapter 2 (Supervisor's Knowledge of Performance; Trust in Supervisor; Performance Review Atmosphere; Feedback/Goalsetting; Perceived Accuracy; Perceived Fairness; and Acceptability). In addition, several one-item questions were included to measure frequency of evaluation and perceived use of the appraisal system (i.e., developmental and/or administrative). All items were rated on a seven-point scale; the low anchor was "strongly disagree," and the high anchor "strongly agree." These items were then randomly ordered, and for the police sample, were included within another survey that had previously been developed for their use. For the university sample, a list of items generated by the personnel department were added and the two lists combined as one larger questionnaire. (The questionnaires for the police and university sample can be found in Appendices B and C, respectively.)

Analysis Procedure

In order to test the proposed causal model of employee perceptions of appraisal systems, several steps were taken. First, the total sample of 416 respondents was randomly divided into a primary and hold-out sample. Second, confirmatory factor analyses of the items were undertaken in order to identify and delete items that measured more than one construct, and to validate (and replicate) the a-priori structure of the hypothesized constructs. The reliabilities for each of these scales were then determined, and factor scores were computed by summing items (i.e., using unit weight) within each construct. Finally, a correlation matrix was generated for these constructs, in addition to the one-item measures concerning system use, frequency of evaluation, and performance ratings. This correlation matrix was then used to test the hypothesized causal model. Each of these steps will be described in separate sections below.

Samples. The data from each organization were randomly assigned into a primary (n=208) or hold-out sample (n=208). The purpose of this step was to provide data with which to replicate the confirmatory factor analyses, thereby adding credence to the results obtained from them. This hold-out sample was also used as a cross-validation sample for the path model.

Confirmatory Factor Analyses. For the data from the primary sample, confirmatory factor analyses (see, for example Joreskog, 1969; Kim & Mueller, 1978; Mulaik, 1972;

Nunnally, 1978) were used to test the proposed structure of the interrelationships between the items developed for each construct. That is, items had been selected to measure seven hypothetically-distinct constructs (Knowledge; Trust; Atmosphere; FBGS; Accuracy; Fairness; and Acceptability). The major questions to be answered by this analysis were (a) are the items hypothesized to measure each construct factorially simple (i.e., contribute significant variance to only one construct)?, and (b) are these seven constructs conceptually distinct (factorial validity)?.

The typical approach taken to answer these questions is through the use of exploratory factor analysis (EFA). However, there are a number of shortcomings inherent in this approach. With EFA, there is an indeterminacy of the solution in terms of the number of factors to retain. That is, "A simple standard for 'when to stop factoring' has not been developed" (Harman, 1967, p. 23). In the common factor model, two solutions (with differing factor structures and numbers of factors) may be equally good at explaining the correlation matrix. Another shortcoming is called "indeterminacy through rotation." Factor patterns are not uniquely determined in that many such solutions (i.e., rotations) can be found that are equally successful at explaining the correlations between variables. Another significant problem with exploratory factor analyses is that this technique typically "underfactors." Hunter and Gerbing (1982) report:

> [Exploratory factor analysis] ... produces fewer factors than there are underlying variables in the data. This is particularly true for causally oriented studies because causal models deliberately include variables that are hypothesized to be highly correlated. [Exploratory] factor analysis tends to throw all highly correlated variables into the same factor. (p. 273)

Along this same vein, Joreskog (1978) writes, "The methods of exploratory factor analysis cannot take [a causal] ... structure into account and if applied to data having such a structure, it will usually give very misleading results (p. 444)."

Because of these problems, confirmatory factor analysis (CFA) was used to answer the two questions noted above. In exploratory factor analysis, the nature of the common factors is revealed by the analysis, rather than postulated in advance. However, in CFA, the researcher specifies a-priori a factor-analytic model concerning the latent variables which are hypothesized to generate the covariances between the observed variables. The parameters in this model can then be estimated, and the goodness of fit of this model to the data can be tested by large sample chi-square tests (if the maximum-liklihood method is used to estimate parameters).

In CFA, a factor model can be specified such that the two indeterminacies discussed above are solved. The "number of factors" indeterminacy is dealt with in that variables are assigned, according to objective features of either content or format, to measure a pre-specified number of factors. Therefore, the very nature of CFA requires the number of factors to be hypothesized. The "indeterminacy due to rotation" is solved if certain rules can be followed in specifying the fixed and free parameters in the model. These rules, discussed in Joreskog (1969), require there be at least k^2 fixed elements in the factor loading and factor covariance matrix in order for the solution to be unique, where k is equal to the number of factors in the model. Depending on the positions and values of these fixed parameters, the solution may also be what Joreskog refers to as "restricted," which imposes conditions on the whole factor space such that the obtained solution cannot be obtained by a rotation of an unrestricted solution. Joreskog (1979, p. 24) notes that these conditions are usually fulfilled in practice. (They are for the analyses presented in this research.)

Joreskog and Sorbom's (1981) LISREL V program, which was used in this research, is very flexible in that is allows any number of parameters in the factor pattern, factor covariance, and/or error (uniqueness) matrices to be specified as either free to be estimated, fixed at any value (e.g., zero), or constrained equal to any other model

parameter. Also, significance tests are automatically made for all free parameters in the model, and modification indices for all fixed and constrained parameters are presented. These modification indices are defined as one-half the sample size multiplied by the ratio between the squared first order derivatives and the second order derivatives. This index is equal to the minimum decrease in chi-square if this parameter was set free, and therefore provides information on how the model could be modified to better fit the data. These modification indices were used to determine the complexity of each item. For example, if an item designed to measure the construct Atmosphere had a significant loading on that construct, but also had a large modification index on another construct, then that item in fact contributes a significant amount of variance to each of the two constructs. Therefore, the number of factors necessary to describe a variable, usually referred to as its "complexity," (see Harman, 1967), is greater than originally intended. For this study, items were retained only if they had a significant loading on the construct for which they were developed.

Computation of Construct Scores and Reliability Estimation. For each of the multi-item constructs in the model, the items retained following the completion of the confirmatory factor analyses were summed and averaged in order to compute construct scores for each individual. The internal consistency reliability (using Cronbach alpha) of

each of these scales was then determined. Construct scores and scale reliabilities were computed for each of the two samples.

Path Analysis. Path analysis was used to test the proposed causal model of employees' perceptions of the Fairness, Accuracy, and Acceptability of performance appraisal systems. Path analysis, also referred to as linear structural equation modeling, simultaneous equation modeling, covariance structure analysis, etc., was first developed by Sewall Wright (1921), and introduced into the social sciences by O. D. Duncan (1966), a sociologist. It is used as a method for studying both the direct and indirect effects of variables that are specified as causes on those variables that are specified as effects. As an analytical technique, it is becoming increasingly popular in the social sciences, as well as in the specific domain of Industrial/-Organizational Psychology (James, Mulaik, & Brett, 1982; Young, 1977), because it provides a systematic way of testing cause-effect relationships among constructs. It is therefore very useful in the process of theory-building. Excellent sources for information on these techniques are provided by Asher (1976), Bentler (1980), Blalock (1971), Goldberger and Duncan (1973), James et al. (1982), Kenny (1979), and Pedhazur (1982).

In order to provide a framework for the steps used in conducting the path analyses for this study, the sequence of operations typically involved in path analysis are presented below.

The three general steps involved in using path analysis to study cause-effect relationships involve model specification, parameter estimation, and model evaluation. First, a theoretical model depicting the causal relationships between constructs is specified. This model should indicate the direct "paths" of causal influence between constructs, such that a change in level in the causally-prior variable (Variable A) should produce a subsequent change in the "effect" variable (Variable B). For example, in the model presented in Figure 3, a change in the level of Supervisor's Knowledge of Performance is hypothesized to cause a subsequent change in Perceived Accuracy of Ratings. In addition, any hypothesized indirect effects (influences of Variable A on Variable B as moderated through a third variable - Variable C) should also be indicated. Again referring to Figure 3, a change in Supervisor's Knowledge is expected to produce change in Perceived Accuracy not only directly, but also through the indirect path from Supervisor's Knowledge to Feedback/Goalsetting, and from these to Perceived Accuracy. Causal influences are graphically denoted by lines drawn between constructs, with arrows used to depict the direction of causality. For the research discussed herein, the causal model was developed and discussed in Chapter 2, and its graphical representation can be seen in Figure 3.

The second step in the path-analytic technique involves the estimation of parameters, generally referred to as path

coefficients, reflecting the relative contribution of each predictor in explaining variability in the dependent measure (called exogenous and endogenous variables, respectively, in path-analytic terminology). Therefore, for a given endogenous variable, the exogenous variable with the largest path coefficient has the greatest influence on it; that is, if a unit of change was produced in all exogenous variables, this variable would generate (cause) the greatest change in the endogenous variable.

After the structural (path) coefficients have been estimated from the data, the model can then be tested. Each of the path coefficients should be tested for significance. Because linear multiple regression techniques are used to compute path coefficients, these coefficients can be tested for significance in the same manner as are beta weights. Paths found to be statistically insignificant might then be deleted from the model, although many writers (see for example McPherson, 1976) argue against this technique, their position being that paths should be deleted based on theoretical, not statistical, grounds. (This process of deleting insignificant paths is known as "theory-trimming"; see Duncan, 1975; Heise, 1969.)

Theory trimming is a method of testing the significance of each path, and therefore of "cleaning up" the model, but it is not a test of the ability of the entire model to accurately describe the data. Because path analysis decomposes correlations into direct and indirect effects,

path coefficients can be used to regenerate a correlation matrix that is implied by both the particular path model and the values of its paths. That is, each path model implies a specific correlation matrix (Pedhazur, 1982).

There are several criteria that can be used to evaluate the ability of a model to reproduce, or closely approximate, the original data matrix. Each of these three methods were used to assess the causal models tested in this research. They were also used to assess the fit of the confirmatory factor analysis models.

The first of these criteria is generally referred to as the "Root Mean Square Residual (RMR)." It is defined (Joreskog & Sorbom, 1981, p. I.41), as the average of the residual variances and covariances (or correlations, if this matrix is analyzed) between the original data matrix and the matrix generated by the model. Generally, RMR values at or below the .Ø5 level (for correlation matrices) are considered as acceptable evidence of the accuracy of the model (Kerlinger & Pedhazur, 1973, p. 318).

Another criterion for assessing the statistical adequacy of a particular path model is referred to as the "Goodness-of-Fit Index (GFI)." This is a measure of the relative amount of variances and covariances that are accounted for by the model, with values ranging from zero to one. Although the statistical properties of its distribution are unknown, GFI is independent of sample size, and is relatively robust against departures from normality (Joreskog

& Sorbom, 1981). Both RMR and GFI can be used to compare the fit of two different models for the same data. In addition, GFI can be used to compare the fit of models to different data.

The final criterion that will be used to assess the model proposed in Chapter 2 is the chi-square measure generated when using the maximum-liklihood estimation technique for determining parameter values. This measure is used to test the hypothesis that the input matrix is of the form of the specified model, against an alternative hypothesis that the matrix is unconstrained (Joreskog, 1977; Joreskog & Sorbom, 1981). This test uses properties of the observed and reproduced matrices to generate a chi-square value with degrees of freedom equal to the number of overidentifying restrictions placed on the model (Joreskog, 1979). A significant chi-square value leads to the rejection of the null hypothesis of no differences, thus concluding the model does not adequately describe the data. Furthermore, the larger the probability level associated with the chi-square value and its degree of freedom, the better the model is said to fit the data.

Unfortunately, this maximum-liklihood statistic has several serious shortcomings. The chi-square value is extremely sensitive to sample size; it is biased in an upward direction as sample size increases (Bentler & Bonett, 1980; James et al., 1982; Joreskog & Sorbom, 1981; Kim & Mueller, 1978). Most writers on this subject (e.g., Bentler, 1980;

Joreskog, 1979; Pedhazur, 1982) indicate that if sample size is too large, then even trivial differences between matrices will be considered statistically significant. Therefore, the probability of rejecting any model increases as the sample size increases, even when the model is minimally false (Bentler & Bonett, 1980, p. 591).

Another shortcoming of this statistic in evaluating models is that departures from normality also increase chi-square values over and above what can be expected due to specification error in the model (Joreskog & Sorbom, 1981). Bentler and Bonett (1980) report that little is known about the robustness of violations of multivariate normality in maximum-liklihood estimators, and although these are techniques for robustifying correlations and covariances, "their use cannot be statistically rationalized in causal modeling (p. 448)."

The general consensus in the literature is that the chi-square test can be used to indicate marked departures from the data, which then lead the investigator to assess this lack of fit by examining other information (such as modification indices, residuals, etc.). Also, Bentler and Bonett (1980) discuss the use of a chi-square <u>difference</u> test for fitting nested models with different numbers of parameters. This difference in chi-square values for the two models, when compared to their differences in degrees of freedom, is itself asymptotically distributed as a chi-square statistic which can be used to test the overidentifying

restrictions that differentiate the two models. Finally, Bentler and Bonett also suggest the use of an "incremental fit index" (IFI), which represents the improvement in fit obtained in evaluating any hierarchial step-up comparison of two nested models (p. 599) (see Bentler & Bonett, 1980, for an excellent discussion of these methodologies).

Summary of Analytical Procedures

To summarize, the analytical procedures used in this research are as follows. First, the data from both organizations were randomly assigned into primary and hold-out samples. Confirmatory factor analyses were then computed for the primary data set to eliminate complex items and to demonstrate construct validity. This solution was replicated on the hold-out sample. The reliability of the summated scales was determined for each sample, and the model was tested for the primary data set using path analytic techniques. This solution was then cross-validated on the hold-out sample.

CHAPTER FOUR

Results and Discussion

In order to present the results of analyses used to test the hypotheses proposed in this research project more clearly, they are organized and presented according to the temporal order in which they occurred. The first section discusses the confirmatory factor analyses of the primary sample and the replication of the final factor model on the hold-out sample. Here, item statistics are also presented. In the following section, the reliabilities and summary statistics of the construct scores are presented for each sample. In the third section, tests of and modifications to the proposed causal model are presented, as are the results of the cross-validation on the hold-out sample.

Confirmatory Factor Analyses

Primary Sample. In the first phase of analyses a series of confirmatory factor analyses of the primary sample were used in order to (a) identify and delete items that were factorially complex, and (b) test and compare the proposed factor model with alternative factor models. Bentler and Bonett (1980) recommend testing the relative effectiveness of alternative models using a chi-square difference test as the solution to interpretive problems associated with the biased nature of the chi-square statistic due to sample size.

To test the factor model proposed for the constructs in this study, the factor pattern matrix was established such that the 47 original items were free to measure only the 7 constructs they were designed to measure. (These items grouped according to construct can be found in Appendix A.) That is, the Knowledge items were specified to load on only the Knowledge factor and no others, the Trust items were specified to load on only the Trust factor, and so on. Therefore, a seven factor solution was hypothesized (Knowledge, Trust, Atmosphere, FBGS, Accuracy, Fairness, and Acceptability). The factor covariance matrix was specified as an oblique solution, and the error/uniqueness matrix was set such that errors were uncorrelated (a diagonal matrix).

In an initial series of factor analyses, the 47 items that were originally included as indicators of the constructs in this study were reduced to a total of 26 items, based on the requirement that each item have a significant loading on only the pre-specified construct. Therefore, 21 items were identified that either had insignificant loadings on the constructs they were designed to measure, or that had significantly high loadings on other constructs. These 26 items that were retained are identified with an asterisk in Appendix A. In addition, these initial runs indicated that the factor covariance matrix as specified was not positive definite; that is, the number of factors inherent in the data was less than specified by the model. An inspection of a correlation matrix of construct scores, computed by summing

all items within each construct, indicated that the constructs Trust and Knowledge correlated at .83, while the constructs Atmosphere and FBGS correlated .91. Therefore, these constructs correlated near unity, thus making the rank of the matrix less than seven. For this reason, the items designed to measure both Knowledge and Trust were combined into a single construct that will be referred to as "Supervisor," so named because both sets of items measure an employee's relations with and perceptions of her supervisor. Likewise, the items measuring Atmosphere and FBGS were combined into a single construct called "Review Session," reflecting the common element of these items sets. Therefore, the five factor model tested below includes the constructs Supervisor, Review Session, Accuracy, Fairness, and Acceptability.

After deleting complex items and combining the two sets of items discussed above, the factor model was tested on the remaining 26 items. Results indicate that the model was only moderately adequate at describing the data from this sample. (Table 4 presents values for the criteria used to assess the adequacy of this and other factor analyses discussed in this section.)

Inspection of Table 4 shows that the five-factor solution had a GFI value of .597, indicating that this factor model accounts for 60 percent of the variances and covariances used as input data. The RMR criterion shows that a sizeable amount of information has been extracted from the

Table 4

Summary Statistics and Tests of Hypotheses for Factor Analyses (Primary Sample)

Summary of Analyses

Hypothesis (Number of Factors)	χ^2	df	GFI	RMR
5	666.96	289	.597	.268
Ø	5495.51	325	.127	1.821
1	2266.3Ø	299	.346	.452
6	49Ø.1Ø	27Ø	.751	.178

Tests of Hypotheses

Hypothesis[a]	χ^2_{diff}	df	p	Decision
5 vs Ø	4828.55	36	<.ØØØ	Accept 5-Factor Model
Ø vs 1	3229.21	26	<.ØØØ	Accept 1-Factor Model
6 vs 1	1776.2Ø	29	<.ØØØ	Accept 6-Factor Model
6 vs 5	176.86	19	<.ØØØ	Accept 6-Factor Model

Note. n for all factor analyses was 2Ø8.

[a] The hypothesis is the test of a factor model with the first-listed number of factors versus the second-listed number of factors.

data matrix; its value was .268. When a correlation matrix is analyzed, the average residual is easy to interpret due to our familiarity with the correlation metric, and because of the simple fact that this metric is invariant. However, when a covariance matrix is analyzed, as was done in all models reported herein, this RMR value can only be interpreted relative to the size of the elements in the particular covariance matrix which generated it (Joreskog & Sorbom, 1981). The average value of the 351 elements in the input covariance matrix used was 1.844. Therefore, an index of the average size of the residuals obtained when analyzing a covariance matrix, transformed into a metric analogous to correlation coefficients, is approximated by the formula $RMR_T = RMR_{cov}/\bar{X}_{cov}$, where the numerator is the root mean residual value obtained by the LISREL output, and the denominator is the average value of the input covariance matrix. In the case of this five-factor model, RMR_T = .268/1.844, for an average residual matrix value (in a metric approximating a correlational metric) of .145. This value, when compared to the .05 value discussed for a correlational metric in Chapter 3, indicates that there might be more information that could be extracted from the input data than this five-factor model can account for. This is corroborated by the GFI value presented above. While factor analytic solutions that account for 60 percent of variance are not poor solutions, these two criteria indicate that this fit might be improved.

Finally, the chi-square value of 666.96 (289, $\underline{n}$ = 208) indicates that the covariance matrix generated by the five factor model was significantly ($\underline{p} < .001$) different than the input matrix. However, for reasons discussed in Chapter 3, this chi-square goodness-of-fit index should not be emphasized because it is strongly influenced by sample size. Bentler and Bonett (1980) suggest using chi-square difference tests to compare hypothesized models with alternative nested models, and Joreskog (1971) reports that the results obtained through confirmatory factor analysis should be used to determine and test alternative factor models. Therefore, several such alternatives to the hypothesized factor model were generated and tested in the manner recommended by Bentler and Bonett (1980).

A model proposed by Bentler and Bonett as the appropriate baseline with which to compare models is called the "null model." This is a restricted model that tests the hypothesis of mutually independent variables and no factor structure. For this primary sample, the null model was run in order to provide information which could be used to test the improvement of fit between the five-factor model and this null model. The chi-square value for the null model was 5495.51 (325, $\underline{n}$ = 208). In order to compare the two models, Bentler and Bonett (1980) suggest the chi-square difference test $[(\chi_1^2 - \chi_2^2)/(df_1 - df_2)]$, where χ_1^2 and df_1 are the chi-square and degrees of freedom for the least restrictive (i.e., null) model and χ_2^2 and df_2 are the values for the most

restrictive (i.e., five-factor) model. Note that this formula is simply the difference in chi-square values compared to the difference in degrees of freedom between models. The difference value comparing the five-factor solution with the null model was equal to 5495.51 - 666.96 with degrees of freedom equal to 325 - 289, or χ^2_{diff} = 4828.55 (36, $\underline{n}$ = 208). Therefore, the five factor solution is significantly ($\underline{p}$ < .001) better at reproducing the original data than is the hypothesis of mutually independent variables. (An inspection of AGF and RMR in Table 4 support this interpretation.)

Another plausible alternative model for the data in this study is a one-factor solution. That is, responses to questionnaire items could result from a generalized impression of the appraisal system, to response bias, etc. The results of a one-factor solution produced a chi-square value of 2266.30 (299, $\underline{n}$ = 208). A comparison of the null and the one-factor models using the difference test indicate the one-factor solution to be significantly better at describing the data (χ^2_{diff} = 3229.21, 26, $\underline{n}$ = 208, $\underline{p}$ < .001). In addition, both GFI and RMR support the one-factor model. It should be noted that the five- and one-factor models cannot be evaluated using the difference test because they are not nested.

Because both the five- and one-factor models were found to be significantly better than the null model, and it was demonstrated earlier that the five-factor model could

potentially be improved, it is possible that a combination of these two models would be needed to best describe the data. In order to test this hypothesis, a general factor was added to the five-factor model. This was specified such that each item was allowed to load on both its own construct and the general factor. Results of this analysis indicate this six-factor solution had a GFI of .751, thus accounting for more of the input covariance than does five factors (GFI = .597). Also, for this solution, the RMR value was much lower (.178) than for the five-factor solution (.268).

Since both the one- and five-factor models are nested within the six-factor model, their relative effectiveness at explaining the data can be statistically compared using the chi-square difference test. Comparing the one-factor model with the six-factor model resulted in a chi-square difference value of 1776.20 with 29 degrees of freedom ($p < .001$). Therefore, this six-factor model is considerably better than the general factor model. The same test was used to compare the five- and six-factor models, with the six factor model being significantly more representative of the data (χ^2 = 176.86, 19, n = 208, $p < .001$). Finally, Bentler and Bonett's (1980) incremental fit index (IFI) was computed to test the improvement of both the five and six factor models when compared to the null model. They suggest that values less than .90 indicate that the model can usually be improved substantially. Their formula for this index is IFI = $(\chi_o^2 - \chi_h^2)/\chi_o^2$, where χ_o^2 is equal to the value obtained from the null

model, and χ^2_h is the value for the hypothesis to be tested. For the five factor model, IFI was equal to .88, indicating a fairly satisfactory model, but one which might be improved. For the six-factor model, the IFI value was .91, thus indicating an improvement going from the five- to the six-factor models. The factor loadings for this six-factor solution are presented in Appendix D.

To summarize the steps taken to test and fit the factor model for the seven constructs proposed in this study, initial analyses found 21 of the original 47 items to be complex; they were deleted from further analyses. In addition, two pairs of the original constructs were found to be highly correlated: items measuring Supervisor's Knowledge and Trust in Supervisor were combined into a single constuct called "Perceptions of Supervisor," and items measuring Atmosphere of Performance Review Session and Perceptions of Feedback/Goalsetting (FBGS) were combined into a construct entitled "Perceptions of Performance Review Session."

A series of nested alternative models to this five-factor model were run; results support a six-factor model which includes the five constructs plus a general factor.

Hold-Out Sample. The six-factor solution discussed in the previous section was replicated using the data from the hold-out sample. The hypothesis was that the six-factor model obtained from the primary sample would adequately describe the data from this sample. Results of these

analyses can be found in Table 5. As was found in the primary sample, the six-factor solution was the most appropriate, although the fit was not quite as good as in the previous sample (e.g., GFI = .670; RMR = .377). The incremental fit index was also computed and indicated that the five-factor model had a value of .85, which the six-factor model improved to .88. The factor model generated by the primary sample had a decrease in IFI across samples of only .03 (from .91 to .88 in the primary and hold-out samples, respectively). Therefore, the six-factor solution is tenable in both samples, thus lending support for construct validity of the scales used in this research. (Descriptive statistics of the data used in these factor analyses can be found in Table 6 for both the primary and hold-out samples.)

Construct Score Statistics and Reliabilities

This section reports summary statistics and intercorrelations for the construct scores and one-item measures used in the path analyses discussed below. In addition, the reliability of the five construct scales are reported.

Construct scores were computed for each individual by summing the items retained within each of the five constructs and dividing by the number of items summed. Table 7 reports the intercorrelations, means, standard deviations, number of items, and reliability for the construct scales within the two samples. In addition, the one-item measures included in

Table 5

Summary Statistics and Tests of Hypotheses for Factor Analyses (Hold-Out Sample)

Summary of Analyses

Hypothesis (Number of Factors)	χ^2	df	GFI	RMR
5	870.02	289	.581	.269
0	5720.57	325	.133	1.834
1	2681.91	299	.335	.565
6	689.27	270	.670	.377

Tests of Hypotheses

Hypothesis[a]	χ^2_{diff}	df	p	Decision
5 vs 0	4850.55	36	<.000	Accept 5-Factor Model
0 vs 1	3038.66	26	<.000	Accept 1-Factor Model
6 vs 1	1992.64	29	<.000	Accept 6-Factor Model
6 vs 5	180.75	19	<.000	Accept 6-Factor Model

Note. n for all factor analyses was 208.

[a] The hypothesis is the test of a factor model with the first-listed number of factors versus the second-listed number of factors.

Table 6

Summary Statistics for Items in Both Primary and Hold-Out Samples

	SAMPLE			
	PRIMARY		HOLD-OUT	
Variable	M	SD	M	SD
TRUS1	5.45	1.80	5.52	1.75
TRUS2	5.47	1.60	5.56	1.64
TRUS3	5.01	1.74	5.12	1.86
TRUS4	5.01	1.58	5.02	1.72
TRUS5	5.24	1.75	5.15	1.90
SUPR1	5.18	1.88	5.30	1.73
SUPR3	4.89	1.94	5.13	1.84
SUPR4	5.11	1.63	5.23	1.75
SUPR6	5.17	1.65	5.26	1.74
ATMO2	5.01	2.09	5.12	2.06
ATMO4	4.57	2.17	4.49	2.21
ATMO5	4.66	2.18	4.66	2.19
FBGS2	3.83	2.01	3.88	2.04
FBGS4	4.38	1.96	4.42	2.05
FBGS5	4.39	2.05	4.34	2.20
FBGS6	4.65	2.14	4.65	2.22
FBGS7	4.50	2.16	4.46	2.21
ACUR1	5.21	1.77	5.14	1.79
ACUR4	4.79	1.94	4.78	1.88
ACUR5	4.99	1.83	4.89	1.82
FAIR1	5.17	1.76	5.10	1.80
FAIR2	4.50	2.03	4.78	1.93
ACPT2	3.95	2.00	4.11	1.93
ACPT3	3.69	1.83	4.05	1.87
ACPT7	3.69	1.87	3.83	1.81
ACPT9	4.41	1.88	4.50	1.92
FREQ	4.19	0.99	4.09	1.08
USE (Adm)	3.78	1.93	3.80	2.08
USE (Growth)	3.09	1.67	3.06	1.75
PERF	-0.02	1.00	0.02	1.00

NOTE. All items except performance ratings evaluated using a 7-point scale. Performance ratings were converted to standardized scores within the university and police department samples.

Table 7

Intercorrelations and Summary Statistics for Primary and Hold-out Samples

Primary Sample

Variable	1	2	3	4	5	6	7	8	9
1. Supervisor	(.94)	.60**	.74**	.76**	.46**	.09	.13	.01	.18**
2. Review		(.97)	.54**	.54**	.39**	.30**	.13	.00	.09
3. Accuracy			(.90)	.72**	.48**	.11	.05	.03	.39**
4. Fairness				(.71)	.55**	.02	.15	.01	.29**
5. Acceptability					(.84)	-.10	.36**	.17*	.20**
6. Frequency						(--)	-.12	-.17*	.01
7. Use (Growth)							(--)	.12	-.01
8. Use (Adm)								(--)	.02
9. Performance									(--)
M	5.17	4.50	5.00	4.83	3.94	4.19	3.09	3.78	-.02
SD	1.44	1.89	1.69	1.67	1.59	.99	1.67	1.93	1.00
#Items	9	8	3	2	4	1	1	1	1

Hold-Out Sample

Variable	1	2	3	4	5	6	7	8	9
1. Supervisor	(.95)	.53**	.65**	.76**	.56**	.07	.18*	.05	.27**
2. Review		(.97)	.48**	.50**	.39**	.31**	.22**	.02	.17*
3. Accuracy			(.90)	.68**	.50**	.11	.14	.07	.40**
4. Fairness				(.83)	.56**	.13	.19*	.04	.32**
5. Acceptability					(.85)	.01	.35**	.10	.09
6. Frequency						(--)	-.03	-.15	.00
7. Use (Growth)							(--)	.26**	.00
8. Use (Adm)								(--)	.04
9. Performance									(--)
M	5.25	4.50	4.93	4.94	4.12	4.09	3.06	3.80	.02
SD	1.48	1.94	1.67	1.72	1.60	1.08	1.75	2.08	1.00
# Items	9	8	3	2	4	1	1	1	1

NOTE. n for each sample = 208. Coefficient alpha reliabilities are in parentheses.

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

the path analyses reported below are also included here. An inspection of this table indicates that the reliabilities (Cronbach alpha) of these summated scales are very high (ranging from .71 to .97), both within and across the primary and hold-out samples used in this research.

Path Analyses

LISREL V was used to test the causal model proposed in this research. In this section, modifications to the proposed model required because of combining the two sets of construct discussed above, as well as the results for the path analyses using both samples, are presented.

Model Modification. Because two pairs of constructs were combined due to their unidimensionality, the path model had to be modified to reflect this change. Inspection of the original model in Figure 3 indicates that the conceptual model proposed for employee perceptions of appraisal system acceptability would not be altered if these two pairs of constructs were combined.

For example, the first pair of constructs (Knowledge and Trust) were conceptually very similar to each other in terms of their hypothesized effects. That is, every construct which Knowledge was hypothesized to influence was also hypothesized to be influenced by Trust. Therefore, combining these two constructs would not affect the model in that no conceptual influences between constructs (i.e., paths) would be added to the model, nor would any be deleted. By reducing the number of constructs within the model without reducing

any conceptual linkages, a more parsimonious model was produced.

The same results obtained by combining Knowledge and Trust into a single construct entitled Supervision are obtained by combining the other pair of constructs found to be unidimensional -- Atmosphere and FBGS. Atmosphere was originally hypothesized to influence Accuracy, Fairness, and Acceptability, which are the same effect variables for FBGS. Therefore, their combined construct (Review Session) also does not alter the model conceptually. The revised model that was tested below can be seen in Figure 4.

Path Analyses Results - Primary Sample. The revised model reported in Figure 4 was tested by using the LISREL program for the data from the primary sample. (Results of analyses reported in this section are summarized in Table 8.) For the proposed model, results indicate support, as indicated by a GFI value of .943 and a RMR value of .056. The chi-square goodness-of-fit index was 44.49 (13, n = 208, p = .0001). This chi-square value will be used to compare alternative nested models using Bentler and Bonett's (1980) chi-square difference test, in a manner similar to that in the confirmatory factor analysis section discussed earlier.

An inspection of the residual correlation matrix produced by the LISREL program indicated that the correlation between Use (Growth) and Acceptability was not sufficiently reproduced by the proposed model (i.e., the residual

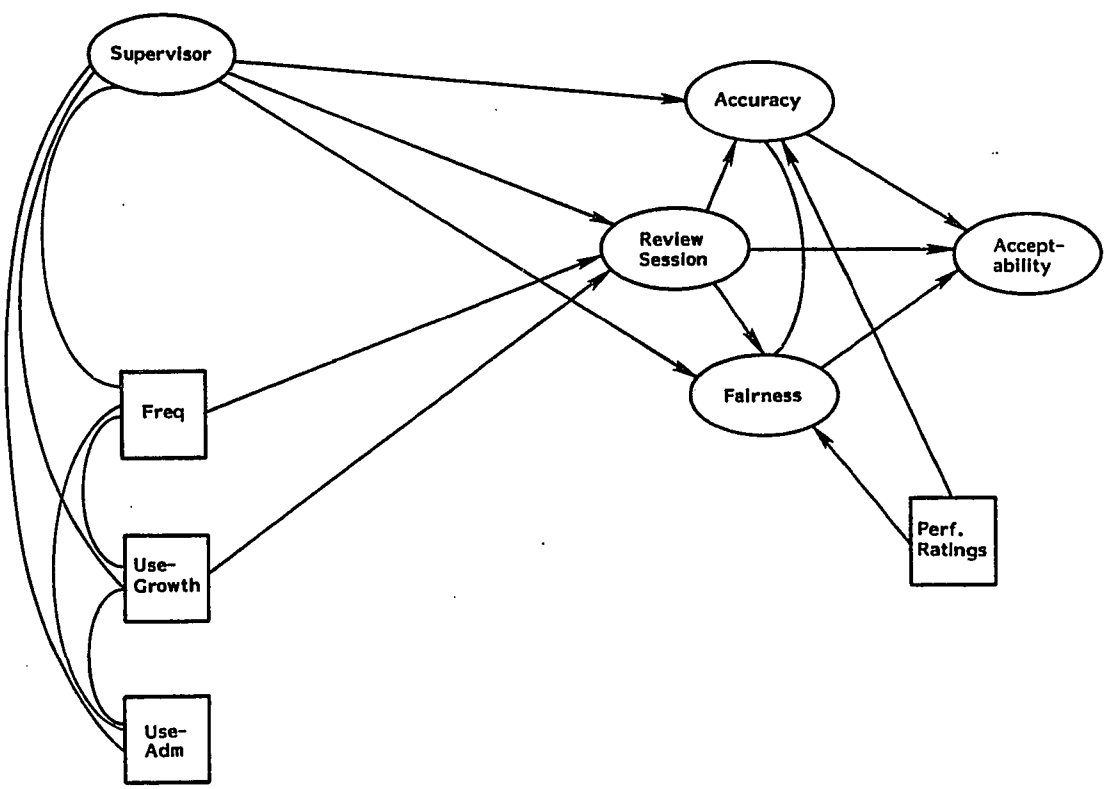

Figure 4. Revised model of causal determinants of employees' perceptions of appraisal system fairness, accuracy, and acceptability.

Table 8

Summary Statistics and Tests of Hypotheses for Path Models

Summary of Analyses

Hypothesis	χ^2	df	p	GFI	RMR
1 (original model)	44.49	13	.000	.943	.056
2 (add Use--Accept)	18.26	12	.110	.967	.031
3 (Delete Review--Accept)	19.22	13	.120	.945	.030
4 (Delete Use--Review)	21.92	14	.080	.960	.034

Tests of Hypotheses

Hypothesis[a]	χ^2_{diff}	df	p	Decision
1 vs 2	26.23	1	.000	Retain Use--Accept
2 vs 3	.96	1	.671	Delete Review--Accept
3 vs 4	2.70	1	.090	Delete Use--Review

Note. n = 208 for all models.

[a] These hypotheses compare the models presented in the Summary of Analyses section of this table.

correlation value, or difference between the original and reproduced correlation, was equal to .293). Therefore, this path model was tested again, with a direct path specified between Use (Growth) and Acceptability. These two models are nested, and can be compared using the difference test. The addition of this path is conceptually justified in that the extent to which employees accept the appraisal system is very likely a function of the purpose of that system. Systems designed to assist employees in identifying areas for growth and improvement are likely to be perceived less threatening and evaluative than systems designed for administrative use.

The RMR for the revised model was .031, indicating an improvement over the previous model based on this criterion. The same conclusion is reached by inspection of the GFI index, which increased after adding this path from .943 to .967. The chi-square value for this second model was equal to 18.26 (12, n = 208, p = .11). The non-significant chi-square value adds support to the improvement of the model by the addition of the path between Use (Growth) and Acceptability. Further confirmation is provided by a significant chi-square difference test (χ^2_{diff} = 26.23, 1, n = 208, p < .001).

An inspection of the ratio of parameter values to their standard errors, what Joreskog & Sorbom (1981) refer to as "T-Values," indicated that that path from Review Session to Acceptability was not significant. It is conceptually possible that the total effect of Review Session on

Acceptability is sufficiently described by its indirect impact through Accuracy and Fairness. Because these indirect paths were significant, the hypothesis of no direct effect was tested. A chi-square value of 19.22 (13, $\underline{n}$ = 208, $\underline{p}$ = .117) was produced by the model, with GFI and RMR values of .945 and .030, respectively. The chi-square difference test produced a non-significant difference (χ^2_{diff} = .96, 1, $\underline{n}$ = 208, $\underline{p}$ = .671), which indicates that by freeing the direct path from Review to Acceptability, a non-significant improvement in the model is made. Support for this constrained, more parsimonious model is added by the decrease in the RMR value. Although the GFI value actually decreased slightly from .967 to .945 between the two models, this is to be expected since any additional paths (regardless of significance) will increase the ability of a model to reconstruct the input matrix. This small decrease in GFI is not considered to be meaningfully significant.

Finally, the path from Use (Growth) to Review had a near significant T-Value of 1.63. To test the improvement of the model if this path were deleted, another LISREL run was made without this path (i.e., set to zero). The chi-square generated by this model was 21.92 (14, $\underline{n}$ = 208, $\underline{p}$ = .08), with GFI equal to .960 and RMR equal to .034. The chi-square difference test produced a value (χ^2_{diff} = 2.70, 1, $\underline{n}$ = 208) that was significant at the .09 level. Therefore, the improvement in the model (i.e., the decrease in chi-square) produced by freeing this one additional parameter was not greater than expected by chance.

This final causal model of employee perceptions of Fairness, Accuracy, and Acceptability of performance appraisal systems can be found in Figure 5. The difference between this final model and the model originally specified (Figure 4) is minimal. These differences include the deletion of two hypothesized causal influences (from Use - Growth to Review, and from Review to Acceptability), and an additional causal influence was added (from Use - Growth to Acceptability) that significantly improved the fit of the model to the data.

This model suggests that employee perceptions of performance appraisal system acceptability are caused primarily (as judged by an inspection of the relative sizes of the standardized path coefficients) by their belief that the system is fair in terms of the ratings they received versus the ratings others receive. The causal process with the next largest influence is the extent to which employees perceive their system as being used for their own developmental purposes. This not only has a direct effect on acceptability, but also has indirect effects through perceptions of the performance review session, which influence the perceived accuracy of ratings and fairness of the system, both of which affect the acceptance of the system. Therefore, the perceived use of the system has many affects, both direct and indirect, on whether appraisal systems will be accepted.

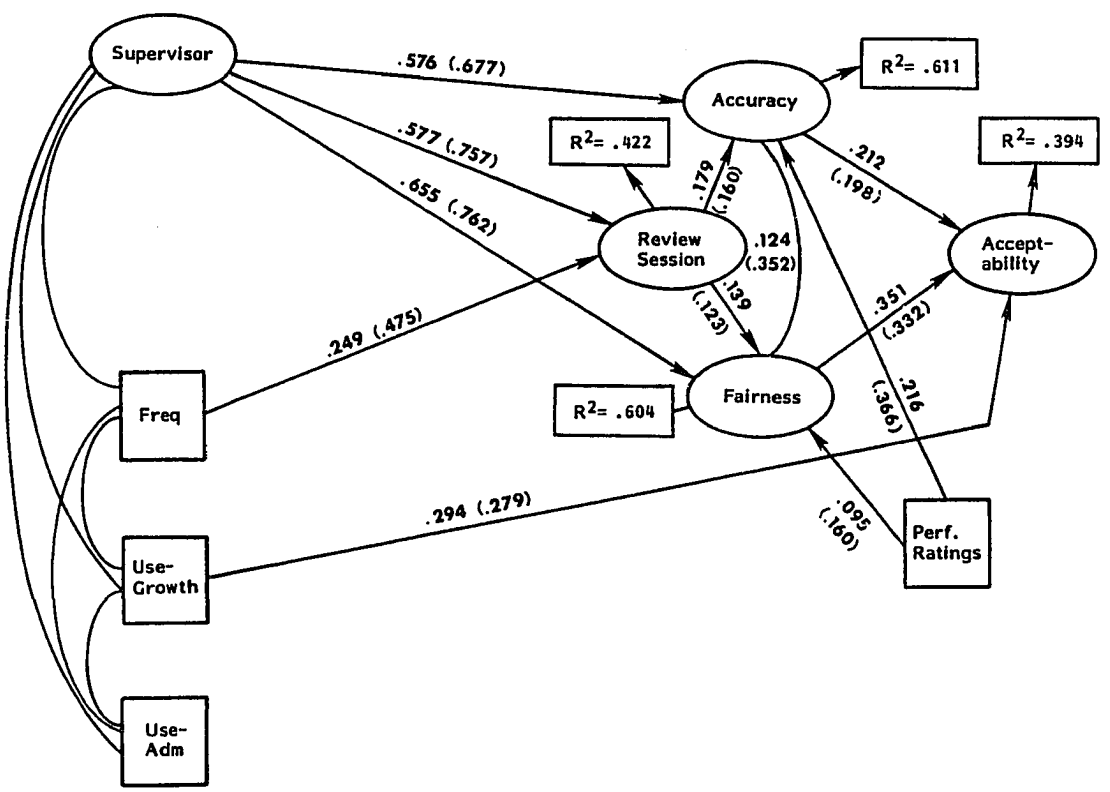

Figure 5. Final model of causal determinants of employees' perceptions of appraisal system fairness, accuracy, and acceptability (unstandardized coefficients are in parentheses).

Another factor which causally affects system acceptance is the extent to which employees believe they received the rating they deserved. Clearly, if an employee felt her level of performance was better than that reflected in the performance rating her supervisor reported, then a negative effect on perceptions of the rating process and outcomes could be expected.

This model also indicates that performance ratings do not moderate perceived acceptability of appraisal systems, an extension of the previous findings that perceptions of fairness/accuracy are not affected by level of rating received (Landy et al., 1980; Vance et al., 1982). Inspection of the revised model reveals that an employer's perception of the accuracy of ratings is largely a function of outcomes relating to the supervisor/subordinate relationship. The construct Supervisor had a large impact on perceived accuracy, indicating that positive perceptions of the supervisor (in terms of believing both that he knows how well you are performing, and that a good interpersonal relationship exists within the dyad) translate into a confidence that the supervisor will rate your performance accurately. Clearly, if a subordinate did not believe his supervisor knew how well he was performing, he could hardly expect the subsequent performance rating to be accurate. Also, if poor interpersonal relations existed, the subordinate might not have confidence that the supervisor would take appropriate actions to gather information with which to make a more accurate rating.

Also, performance ratings were found to have an impact on the perceived accuracy of ratings. That is, respondents with high performance ratings perceive the ratings as more accurate than do respondents with low ratings. While this finding should be taken at face value, in the organizations studied in the research, it might be true that people who received low ratings actually performed at a higher level. Therefore, the perception that their rating was not accurate would in fact be correct.

Another possible explanation involves the concept of cognitive dissonance (Festinger, 1957). There are two types of individuals with high performance ratings--those who truly believe they deserved them (i.e., they may have performed at high levels), and those who did not perform at a level comparable to the rating received from their supervisor. The dissonance created by this difference may have been reduced by altering beliefs concerning their performance level to conform to that of their supervisor. That is, one might rationalize, "If my supervisor thinks I am doing such a good job, then I must be doing better than I thought." By altering these self-perceptions, high performance ratings do "cause" perceptions of accuracy of these ratings.

The effect of ratings on perceived accuracy neither contradict nor support the findings of Landy et al. (1980) or Vance et al. (1982), because the criterion used in their studies (perceived fairness/accuracy of ratings) cannot be equated with perceived accuracy, the criterion used in this

study. (Inspection of Figure 5 indicates that performance ratings also had a significant, but small, effect on perceived fairness.) The impact of performance ratings on accuracy indicate that to increase favorable attributes towards the appraisal system, ratings could be intentionally biased upward, thus making them more acceptable to the employee (at the expense of being less accurate)! In future research, if a relationship can be demonstrated between acceptability and other desired outcomes (e.g., improved organizational effectiveness, job satisfaction, reduced turnover), this might be a viable, albeit a conceptual and procedurally unsettling, alternative. Additional research is needed to more thoroughly understand this relationship between rating levels and perceived accuracy.

The final causal model in Figure 5 also indicates that perceived fairness of system can be accurately determined (i.e., 61 percent of its variance accounted for). The three constructs found to have an impact on fairness were those originally proposed--Supervisor, Review, and performance ratings. Perceptions of the supervisor had the largest impact on fairness. If an employee believes her supervisor to have a clear understanding of his performance, knows the requirements of performing her job, and has a good working relationship with her supervisor (all indicators of this construct), then perceptions of fairness will increase.

Perceptions that occur of the Performance Review Session were also found to impact Perceived Fairness, as was

hypothesized. Apparently, by discussing ratings and goals in a positive, relaxed manner, the employee can better understand the rationale and criteria which was used by his supervisor in his rating process.

Although performance ratings were found to relate to fairness, the effect was small. The discussion on the effects of ratings on accuracy are relevant here, also. Perhaps low performance ratings did not reflect actual (higher) levels for some respondents. These subordinates (relative to other persons with higher ratings) would then not receive their "fair share" of benefits which accrue from higher ratings (i.e., recognition, merit pay, promotions, etc.).

Finally, the construct Perceptions of the Performance Review Session was predicted by two of the three hypothesized constructs. Again, Perceptions of Supervisor had a large impact on favorable impressions of this session. If the supervisor demonstrates an interest in the employee, has a good understanding of his performance, and has good interpersonal relations with her subordinate, then the process of providing feedback and setting goals within this session should be less threatening and more constructive.

The frequency of evaluation also impacts perceptions of the Review Session. Certainly, the more opportunities that supervisors and subordinates have to discuss performance and goals, the fewer chances there are for perceptual differences to occur that could lead to problems when discussing past performance and setting future goals.

Perceived use (Growth) of the system did not enter as a significant predictor of the perceptions of the Review Session at the .05 level, although it was significant at the .09 level. It could be that there was insufficient power to detect this effect (Cohen, 1973). For example, had the sample size been n = 250 intead of n = 208, with the correlations remaining constant, this path would have been statistically significant.

Summary of Path Analysis--Primary Sample

The results of the path analysis indicate support for the hypothesized causal model, concerning employees' perceptions of the determinants of appraisal system fairness, accuracy, and acceptability. Three modifications to the original model (two path deletions and one addition) resulted in an improved, more parsimonious model.

Cross-Validation of Path Model

In order to confirm the path model discussed above, a cross-validation strategy was used. Cross-validation of a path model is somewhat different from, but conceptually related to, the standard cross-validation procedures used for regression equations. In order to cross-validate a regression equation, weights derived from one sample are used to generate predicted criterion scores for a second sample of subjects. The correlation between the predicted and actual criterion scores, when compared to the multiple correlation obtained from the first sample, serves as an index of the stability of beta weight across samples.

Cross-validation of a path model involves the constraining of path coefficients for the cross-validation model to be equal to those derived from the primary sample. The difference between the correlation matrix generated by these path coefficients (i.e., the predicted correlation matrix), compared to the actual correlation matrix for the second sample, serves as an index of the stability of path coefficients across samples. That is, if these path coefficients generate a predicted correlation matrix nearly identical to that of the hold-out sample (as judged by a residual matrix between the two), then these paths are stable across the primary and hold-out groups.

The results of this analysis indicate substantial support for the model derived from the primary sample. After setting all paths equal to those presented in Figure 5, results indicate a non-significant difference betwen the two correlation matrices (χ^2 = 35.02, 25, n = 208, p > .05). The GFI value was .943, indicating that the use of path coefficients resulted in a decrease in the correlations accounted for from .960 for the primary sample to only .943 for this sample. In addition, the RMR score, or the average squared value in the residual matrix, was only .047. Therefore, the path model was successfully cross-validated.

Finally, an inspection of the residual correlation matrix indicated that the correlation between Use (Growth) and Review Session was not adequately reproduced as judged by a residual correlation of .121. This path was set free, and

the model rerun. A chi-square difference test ($\chi^2_{diff} = 4.74$, 1, $\underline{n} = 208$, $\underline{p} = .028$) indicated that the addition of this path (its value equal to .121) significantly improved the fit of this model to the data. Note that it was this path that was almost significant in the primary sample.

(The values for the standardized path coefficients were presented in Figure 5. Unstandardized coefficients are presented in parentheses in this same figure.)

CHAPTER FIVE

Conclusions

The purpose of this study was to develop and operationalize constructs that could adequately describe employee perceptions of performance appraisal processes and systems. Then, a proposed model consistent with available research integrated both these constructs and other measures into a causal structure that was developed and tested.

Summary of Causal Effects

In the following section, a summary of the causal processes related to each of the endogenous variables (perceptions of fairness, accuracy, review session, and acceptability) are presented. Next, an explication of causal relationships from each construct to its various effects are discussed. Limitations concerning the interpretation of the findings are then presented, as are implications for future research.

As proposed, perceptions of both the fairness and accuracy of ratings were shown to be predicted by the quality of the relationship with the supervisor, by events that occur in the performance review session, and by performance ratings. The hypothesis that impressions of the performance review session would be causally subsequent to perceptions of the supervisor, the frequency of evaluation, and the perception that the appraisal system was used for their growth and development purposes was partially supported, with

the latter path significant in only one of the two samples. The hypothesis that appraisal system acceptability would be causally subsequent to the perceptions of accuracy, fairness, and of satisfaction with the performance review session was also supported; however, the effect of the review session on acceptability was found to be indirect, not both indirect and direct, as hypothesized. Finally, perceptions that the system was used for administrative purposes was not related to any of the constructs, as expected.

Perceptions of the Supervisor. Positive relationships with the supervisor were hypothesized and found to have a powerful effect in many ways on the perceptions of their subordinates. According to the results obtained in this study, these positive relations led to a belief that ratings were accurate. This construct (Supervisor) measured perceptions related to both the supervisor's knowledge of performance, as well as the quality of the dyadic relationship between the supervisor and subordinate. However, it is not known whether both of these components are necessary in order for this causal relationship to hold. For example, supervisors could be perceived as knowledgeable about performance levels, but if the quality of the dyadic relationship was poor, subordinates might not have confidence (or trust) that her supervisor would rate accurately. On the other hand, if a subordinate did not believe his supervisor to be knowledgeable concerning his performance levels, if the relationship was one characterized by trust, the subordinate

might believe that such a supervisor would not rate until he gathered relevant information with which to make a more accurate decision.

Relationships with and perceptions of the supervisor also had a large impact on perceived fairness of the appraisal system. Ratings can be unfair for two reasons: (a) a supervisor's ratings may be perceived as biased in that he gives certain employees higher ratings not reflective of their performance (i.e., intra-rater differences), and/or (b) different raters are perceived as having different implicit standards which they use to compare their impressions of subordinates' performance (inter-rater differences). If the subordinate has a good relationship with and trusts her supervisor, then perhaps she might perceive the probability of intentional bias occurring as being low. The conceptual link between the quality of the dyadic relationship and inter-rater levels of fairness is not as easy to explain. However, the items retained in this study measured only the intra-individual aspects of fairness. Future research is needed to investigate the relationships with and causal determinants of perceptions of inter-rater fairness.

Positive perceptions of and relationships with the supervisor also had an impact relating to events that took place in the performance review session. Clearly, if a supervisor has a good understanding of the performance levels of her subordinates, then the discussion concerning performance feedback in this review session can be expected

to proceed more smoothly in that the supervisor may be more likely to be able to justify, perhaps even to the subordinate's satisfaction, the ratings received. If the subordinate does not perceive his supervisor as knowledgeable, and discrepant ratings are received, then the probability of reaching an accord is diminished unless the supervisor is willing to alter the ratings.

The positive perceptions of the relationships with the supervisor also can be expected to improve the quality of the communication and atmosphere within this session. If the dyadic relationship is generally good (in this study, perceptions of the supervisor are seen as relatively long-term in nature), then this can be expected to carry over into the review session as well.

In light of the above findings, the relationship with and impressions of the supervisor clearly have a large impact on perceptions of the appraisal process. These findings support and extend the results obtained by Landy et al. (1978), Barr et al. (1981), Dipboye and de Pontbriand (1981), and Vance et al. (1982), all of whom reported that an item measuring one aspect of this construct ("supervisor's knowledge of performance") was a significant predictor of the fairness/accuracy of ratings. This research indicates that this larger construct not only predicts perceptions of both fairness and accuracy as independent concepts, but also has a large impact in producing satisfactory outcomes concerning the feedback session.

Frequency of Evaluation. The frequency of evaluation was found to have a positive impact on perceptions of the performance review session. Apparently, the more often supervisors and subordinates have an opportunity to discuss performance and set goals, the more favorable this process is viewed by the subordinates.

Perceptions of Performance Review Session. Employee perceptions of the performance review session were also found to have a significant impact on perceptions of both fairness and accuracy of ratings. Within the review session, the subordinate has an opportunity to discuss ratings (he received) with his supervisor. If the supervisor has the appropriate skills necessary to establish a proper climate (i.e., atmosphere) within this review session, as well as the communication skills necessary to effectively discuss and justify the basis upon which the specific ratings were given, then the employee might come to a better understanding of his performance as perceived by others. This could lead him to alter self-perceptions regarding his performance, thus making the supervisor's ratings perceived as more accurate.

Processes within the performance review session were also found to impact perceived fairness of ratings. Positive interactions, hopefully in which the supervisor presents the basis upon which she made the rating, might lead subordinates to believe that ratings are not made in some arbitrary, and often-times biased manner.

Perceptions of this feedback session were expected to have direct impact on system acceptability, but this hypothesis was not confirmed. It was originally thought that a positive perception of the quality of the interpersonal contact within this session, partially independent of considerations of fairness and accuracy would be related to employee satisfaction with the appraisal system in general; that is, a "halo effect" would occur if the subordinate was satisfied with the processes and outcomes of this session. This hypothesis was not substantiated. Instead, the review session had indirect effects on the acceptability of the system through perceptions of both the fairness and the accuracy of the ratings received, as hypothesized.

Acceptability of Appraisal Systems. Finally, the acceptability of appraisal systems is a causal effect of the perceived accuracy and fairness of ratings, as originally hypothesized, as well as an effect of subordinates' beliefs that their system is used for growth and development. The finding that both fairness and accuracy are determinants of system acceptability lends support for efforts to improve supervisory skills that are related to the processes of providing and discussing performance feedback and setting of future goals. This might involve the training of counseling-type skills, interpersonal (e.g., communication) skills, and observational skills. Interpersonal skills are seen as particularly important, as they were shown to have a large impact on all major aspects of the appraisal process.

The significant effect that employees' perceptions of the developmental use of the appraisal system has on acceptability of the appraisal process is not difficult to understand. It is generally known that administrative systems are often considered threatening by the users, because such systems are judgmental in nature. These systems attempt to delineate and accentuate inter-individual differences in performance, so the organization can differentially provide rewards to "better" performers. However, by emphasizing these differences, problems can occur within the workplace when employees with essentially the same performance are rewarded differently. Developmental systems, on the other hand, emphasize intra-individual differences for the purpose of determining, for a given employee, her relative strengths and weaknesses. By emphasizing intra-individual differences, employees may come to believe that their supervisors, as well as the organization, care for their personal development, thus leading to the acceptability of the appraisal process and system.

Performance Ratings. Performance ratings were expected and found to have impact on both the fairness and accuracy of ratings. It is possible that these effects are the result of dissonant cognition as discussed earlier. In addition, in light of the negative findings presented by Landy et al. (1980) and Vance et al. (1982), it might be that these results are sample-specific. Future research is needed to determine if these relationships are stable across

populations, and to delineate more clearly the conceptual and cognitive basis for these relationships if substantiated.

Limitations of Research Findings

The results obtained from this study should be viewed in light of certain limitations. The small sample size may have produced results that would not have been consistent with a larger sample. The lack of significance of the causal path from Use (Growth) to Performance Review is one possible consequence of the relatively small sample size in that this path was significant in the hold-out sample, but not (p = .Ø9) in the primary sample. This model should be replicated with a larger sample to investigate such differences.

Another shortcoming of this research is related to organizational constraints imposed concerning the collection of demographic information. It is generally a good research strategy to rule out any effects these individual-difference variables might have on results; however, one of the participating organizations would not allow the collection of these data.

Because all of the measures collected in this research were designed to measure employee perceptions, by necessity they were subjective in nature. It would have been desirable to collect objective data, especially organizationally-relevant data such as tenure, time in position, etc., or to obtain objective performance measures and compare them to self-reports; however, this was not possible. The subjective nature of these data raise the

possibility of response bias, of which the general factor added to the factor analysis model may have been a part. However, methods that could be used to measure employee perceptions, without asking questions to which they respond, are difficult, if often impossible, to develop.

Finally, another limitation of this research is the cross-sectional design used to collect data, and the inherent problems in attributing causality from such information. In the future, this and other related models should be investigated with longitudinal data, where possible.

Implications for Future Research and Practice

This study presented here has direct implications for future research and applications in the personnel and organizational psychology areas. Clearly, the important role that supervisors play in determining employees' perceptions of the fairness, accuracy, and acceptability of appraisals and systems indicates the need for more training and a higher level of interpersonal skills then has typically been assumed. Lip-service is often given attesting to the importance of such skills; however, if appraisal acceptability is a relevant organizational goal, then more attention must be paid to assessing and training for them.

The impact of the review session on both perceived fairness and accuracy of ratings also indicates that the processes within these sessions should be investigated that produce such results. By improving our understanding of these processes, supervisors could be taught to incorporate them as they provide feedback and set goals.

While it is intuitively reasonable to assume that appraisal system acceptability may be related to a decrease in employee grievances and/or legal complaints, future research should empirically investigate this potential relationship by collecting hard criterion measures. In addition, the relationship between acceptability and other production and personnel-related measures (i.e., satisfaction, turnover, committment) should be investigated.

The research reported in this study indicates substantial support for a revised model of employee perceptions of the causal determinants of performance appraisal system fairness, accuracy, and acceptability. Prior research in this area has not produced conceptual models which could be used to drive research, and therefore may have hindered the development of a better understanding of the perceptual processes involving performance appraisal. This research has extended the available literature in terms of providing an initial model which could be used as a point of departure. No doubt there are other relevant factors influencing employee perceptions of appraisal systems that were not included in this study. Also, research is needed to develop and test the determinants of appraisal system acceptability in terms of other levels in the organization (e.g., the determinants of system acceptability in relation to supervisors, or acceptability in terms of the overall organization, as briefly mentioned in Figure 1). For example, Lawler (1967) suggested that appraisal systems which

are acceptable at the supervisory level may lead to more accurate appraisals. Since performance appraisal data is often used as criteria for other human resource systems (such as training and/or selection programs), such improved data might lead to more effective organizational processes.

It is hoped that future research will examine, elaborate, and refine the present model, as well as investigate the relationship between employee acceptance of their appraisal systems and other organizationally-relevant variables. It is not until the appraisal system is accepted at all organizational levels will there be both an improvement in the accuracy of measurement as well as satisfaction with the evaluation process. Only then will an increase in the utility of the appraisal process from both the organizational and individual perspectives be fully realized.

References

Asher, H. (1976). Causal modeling. Beverly Hills, CA: Sage Publications.

Barr, S. H., Brief, A. P., & Fulk, J. L. (1981). Correlates of perceived fairness and accuracy of performance. Academy of Management Proceedings, 156-160.

Bartlett, C. J., & Sharon, A. T. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22, 251-263.

Bass, B. M., & Barrett, G. V. (1981). People, work, and organizations (2nd Ed.). Boston: Allyn and Bacon.

Beer, M., Ruh, R., Dawson, J. A., McCaa, B. B., & Kavanagh, M. J. (1978). A performance management system: Research, design, introduction, and evaluation. Personnel Psychology, 31, 505-535.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bernardin, H. J., & Boetcher, R. (August 1978). The effects of rater training and cognitive complexity on psychometric error in ratings. Paper presented at the meeting of the American Psychological Association, Toronto.

Bernardin, H. J., & Cardy, R. L. (1981). Cognitive complexity in performance appraisal: It makes no nevermind. Academy of Management Proceedings, 306-310.

Blalock, H. M., Jr. (1971). Causal models in the social sciences. Chicago: Aldine-Atherton.

Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 238-252.

Borman, W. C. (1979). Individual difference correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.

Borman, W. C. (1980). Performance judgments: The quest for accuracy in ratings of performance effectiveness. Paper presented at the First Annual Scientist-Practitioner Conference, Old Dominion University, Norfolk, VA.

Cascio, W. F. (1978). Applied psychology in personnel management. Reston, VA: Reston Publishing.

Cohen, J. (1973). Statistical power analysis and research results. American Educational Research Journal, 10, 225-230.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Cummings, L. L. (September 1976). Appraisal purpose and the nature, amount, and frequency of feedback. Paper presented at 84th Annual convention of American Psychological Association, Washington, D.C.

Dayal, I. (1969). Some issues in performance appraisal. Personnel Administration, 32, 27-30.

DeCotiis, T. A. (1977). An analysis of the external validity and applied relevance of three rating formats. Organizational Behavior and Human Performance, 19, 247-266.

DeCotiis, T. A., & Petit, A. (July 1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 635-646.

Dipboye, R. L., & de Pontbriand, R. (1981). Correlates of employee reactions to performance apprisal and appraisal systems. Journal of Applied Psychology, 66(2), 248-251.

Drasgow, F., & Miller, H. (1982). Psychometric and substantive issues in scale construction and validation. Journal of Applied Psychology, 67, 268-279.

Duncan, O. D. (1966). Path analysis: Sociological examples. American Journal of Sociology, 72(1), 1-16.

Duncan, O. D. (1975). Introduction to structural equation models. New York: Academic.

Dwyer, J. C., & Dimitroff, N. J. (1976). The bottoms up/tops down approach to performance appraisal. Personnel Journal, 56, 349-353.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Goldberger, A. S., & Duncan, O. D. (Eds.). (1973). Structural equation models in the social sciences. New York: Seminar Press.

Graen, G., & Schiemann, W. (1978). Leader-member agreement: A vertical dyad linkage approach. Journal of Applied Psychology, 63(2), 206-212.

Harman, H. H. (1967). Modern factor analyses (2nd Ed.). Chicago: University of Chicago Press.

Hedge, J. W. (1982). Improving the accuracy of performance evaluations: A comparison of three methods of performance appraiser training. Unpublished doctoral dissertation, Old Dominion University, Norfolk, VA.

Heise, D. R. (1969). Problems in path analysis and causal inference. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), Sociological methodology 1969. San Francisco: Jossey-Bass.

Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second order factor analysis, and causal models. Research in Organizational Behavior, 4, 267-320.

Huse, E. F. (1967). Performance appraisal--a new look. Personnel Administration, Mar.-Apr., 3-5, 16-18.

Ilgen, D. R. (September 1981). Performance rating scales for feedback purposes (Technical Report No. 81-3). Washington: Army Research Institute.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64(4), 349-371.

Ilgen, D. R., Peterson, R. B., Martin, B. A., & Boescher, D. A. (1981). Supervisor and subordinate reactions to performance appraisal sessions. Organizational Behavior and Human Performance, 28, 311-330.

Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. Personnel Psychology, 33, 595-640.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills, CA: Sage Publications.

Jenkins, G. D., & Taber, T. A. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. Journal of Applied Psychology, 62, 392-398.

Joreskog, K. G. (1969). A general approach to confirmatory factor analysis. Psychometrika, 34, 183-202.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Joreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), Applications of statistics. Amsterdam: North-Holland Publishing Co., 265-287.

Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Joreskog, K. G. (1979). Basic ideas of factor and component analysis. In K. G. Joreskog and D. Sorbom, Advances in factor analysis and structural equation models. Cambridge, Mass: AGT Books, 5-20.

Joreskog, K. G., & Sorbom, D. (1981). LISREL: Analysis of linear structural relationships by the method of maximum liklihood, version V. Chicago: National Education Resources, Inc.

Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. Psychological Bulletin, 85, 555-586.

Kane, J. S., & Lawler, E. E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In. B. Staw (Ed.), Research in organizational behavior (Vol. 1). Greenwich, Conn: JAI Press.

Kavanagh, M. J. (1981). Evaluating performance. In K. Rowland & G. Ferris, Personnel Management: New Perspectives. Boston: Allyn and Bacon.

Kavanagh, M. J., & McAfee, B. (1982). Performance appraisal acceptability working papers.

Kavanagh, M. J., Hedge, J., & DeBiasi, G. (May 1983). A closer look at the correlates of performance appraisal system acceptability. Paper presented at the meetings of the Eastern Academy of Management, Pittsburgh, PA.

Kenny, D. A. (1979). Correlation and causality. New York: John Wiley & Sons, Inc.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Kim, J., & Mueller, C. W. (1978). Introduction to factor analysis. Beverly Hills: Sage Publications.

Landy, F. J., Barnes, J. L., & Murphy, K. R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. Journal of Applied Psychology, 63(6), 751-754.

Landy, F. J., Barnes-Farrell, J., & Cleveland, J. N. (1980). Perceived fairness and accuracy of performance evaluation: A follow-up. Journal of Applied Psychology, 65(3), 355-356.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychology Bulletin, 87(1), 72-107.

Larson, J. R., Jr. (1981). Some hypotheses about the cause of supervisory performance feedback behavior. Academy of Management Proceedings, 198-202.

Lawler, E. E., III. (1967). The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 51(5), 369-381.

Maier, N. R. F. (1963). Problem-solving discussions and conferences. New York: McGraw-Hill.

McPherson, J. M. (1976). Theory trimming. Social Science Research, 5, 95-105.

Meyer, H. H., Kay, E., & French, J. (1965). Split roles in performance appraisal. Harvard Business Review, 43, 123-129.

Mulaik, S. A. (1972). The foundations of factor analysis. New York: McGraw Hill, Inc.

Nunnally, J. C. (1978). Psychometric theory (2nd Ed.). New York: McGraw-Hill.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd Ed.). New York: Holt, Rinehart, & Winston.

Roberts, K. H., & O'Reilly, C. A. (1974). Measuring organizational communication. Journal of Applied Psychology, 59, 321-326.

Sauser, W. I., & Pond, S. B. (1981). Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. Personnel Psychology, 34, 609-626.

Schneier, C. E. (1977). The operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 62, 541-548.

Sullivan, J. L., & Feldman, S. (1979). Multiple indicators: An introduction. Beverly Hills: Sage Publications.

Vance, R. J., Winne, P. S., & Wright, E. S. (1982). Correlates of rater and ratee reactions to a performance appraisal system. Unpublished manuscript, Old Dominion University.

Wright, S. (1921). Correlation and causation. Journal of Agricultural Research, 20, 557-585.

Young, J. W. (1977). The function of theory in a delemma of path analysis. Journal of Applied Psychology, 62, 108-110.

APPENDICES

Appendix A

Abbreviated Items for Each of the Seven Constructs

Abbreviated Items for Each of the Seven Constructs
(Those retained following the initial factor analyses
are marked with an asterisk)

Trust in Supervisor:

* TRUS1 Able to have frank and open communications with my supervisor.
* TRUS2 Supervisor would use power to help you.
* TRUS3 Supervisor interested in my well being.
* TRUS4 Supervisor supports my decisions.
* TRUS5 Supervisor honest in dealings with me.

Feedback and Goalsetting Session:

FBGS1 My supervisor and I set specific performance goals.
* FBGS2 The goals were related to performance weaknesses.
* FBGS3 I am satisfied with amount of information from supervisor.
* FBGS4 Goals were reasonable.
* FBGS5 I knew what action to take to improve performance.
* FBGS6 We spent enough time discussing results.
* FBGS7 I knew what my supervisor expected in the future.

Supervisor's Knowledge of Performance:

* SUPR1 Supervisor observes enough performance to rate me.
SUPR2 Supervisor has good understanding of what I do.
* SUPR3 Supervisor knows how much work I do.
* SUPR4 Supervisor knows how well I am performing my job.
SUPR5 My immediate supervisor should rate me.
* SUPR6 My supervisor knows the requirements of my job.

Atmosphere of the Performance Review Session:

ATMO1 I feel free to disagree with my supervisor.
* ATMO2 I had opportunities to express feelings in review session.
ATMO3 The interview had a relaxed atmosphere.
* ATMO4 I was encouraged to present my point of view.
* ATMO5 My supervisor exhibited a helping attitude.

Accuracy of Ratings:

* ACUR1 I received the rating I expected.
  ACUR2 My supervisor understands constraints beyond my control.
  ACUR3 I accept the judgment of my rater.
* ACUR4 My rating represented my true performance.
* ACUR5 My supervisor's evaluation was accurate.
  ACUR6 Other supervisors are accurate in their evaluations.
  ACUR7 My supervisor's evaluation was accurate.

Fairness of Rating System:

* FAIR1 My supervisor is as fair as other supervisors when rating.
* FAIR2 I am confident my ratings are fair in relation to others.
  FAIR3 My supervisor is a lenient rater.
  FAIR4 System distinguishes good and poor performers.
  FAIR5 Performance ratings provide a fair basis for decisions.
  FAIR6 My supervisor rates harder than other supervisors.
  FAIR7 I feel the evaluation system is fair.

Acceptability of Appraisal System:

  ACPT1 I understand the evaluation system.
* ACPT2 The appraisal program is an acceptable way to evaluate performance.
* ACPT3 Ratings in general reflect performance.
  ACPT4 Overall, I was satisfied with last evaluation.
  ACPT5 Employees were involved in the development of the system.
  ACPT6 Appraisal system is constructive.
* ACPT7 Appraisal form is acceptable.
  ACPT8 I had a clear idea of standards.
* ACPT9 I benefit from my performance evaluation.

Other Items Included in the Path Model:

  USE1 System is used primarily for making administrative decisions.
  USE2 System is used for growth and development of employees.
  FREQ How frequently were you evaluated?
  PERF Performance ratings (standardized within university and police samples)

Appendix B

Police Sample Questionnaire

___ PERFORMANCE RATING AND FEEDBACK SURVEY

During the past two years we have been administering surveys to gain additional information about the ___ performance appraisal system as well as on the feedback session. The information has been used to modify the system according to your needs and to fulfill the overall purpose of the ___ performance evaluation system. Consequently, the purpose of this survey is to pursue further the refinement and involvement of the system. Your participation, of course, is completely voluntary. However, we urge you to complete and return this survey. Remember, this is a good opportunity to provide input for making the rating system more useful to you.

Listed on the following pages are a number of statements about the performance ratings and the recent feedback sessions that were conducted with your supervisor. The instructions are as follows:

(1) Each set of statements is preceded by a rating scale to use in responding.
(2) Please read each item carefully and write the number corresponding to your choice in the blank to the left of the item.
(3) Answer all items please.
(4) After you have completed the survey, please place it into a sealed envelope and return it to _____.

All of your responses will remain strictly confidential. For research purposes, it would be very helpful if you would supply the following information. Thank you for your cooperation.

Rank:

Control No.:

Amount of time in the Department (in months):

Consultants to the Center for Applied
Psychological Studies
Old Dominion University

SECTION I: PERFORMANCE EVALUATIONS

INSTRUCTIONS: This section contains a number of statements about topics related to the ___ Performance Rating System. Read each statement and decide to what degree you agree or disagree with it, according to the scale below. Write the number corresponding to your choice in the blank to the left of each item. IF YOU WERE NOT RATED, DO NOT COMPLETE THE QUESTIONNAIRE.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 1. My supervisor observes enough of my performance to rate me.

_____ 2. I am able to have frank and open communications with my supervisor.

_____ 3. I have a good understanding of the performance rating system used to rate me.

_____ 4. I would feel free to disagree with my supervisor if I felt his/her evaluation was not fair or did not take all the facts into consideration.

_____ 5. In my last performance evaluation, I received the ratings I expected.

_____ 6. I feel that the performance rating system is an acceptable way to evaluate job performance.

_____ 7. I feel my supervisor is as fair as other supervisors in providing ratings of job performance.

_____ 8. My supervisor has a good understanding of what I do on my job.

_____ 9. I feel that performance ratings, in general, accurately reflect how well an employee has performed his or her job.

_____ 10. I would prefer a performance evaluation system in which I would be rated more than two times a year rather than on a semi-annual basis.

_____ 11. Overall, I was satisfied with my last semi-annual evaluation.

_____ 12. My supervisor understands that many things which affect how well I do my job are beyond my control, and rates accordingly.

_____ 13. I feel my supervisor knows how much work I actually do.

_____ 14. I have confidence that my ratings are made on a fair and equal basis with others being rated.

_____ 15. Employees of the _____ _____ _____ had an important part in the development of the performance rating system (for example, developing job requirements, standards, scales).

_____ 16. I feel that my supervisor would be personally inclined to use his/her power to help me solve a problem in my work.

_____ 17. I think that ___'s rating system serves a useful purpose (i.e., helps identify training needs; helps my supervisor and me agree on job duties; helps improve job performance, etc.)

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 18. When compared to other supervisors, my supervisor is an extremely lenient (very easy) rater when he/she rates my performance.

_____ 19. The current performance rating form is acceptable to me.

_____ 20. I accept the judgment of the person who rates me regarding the strengths and weaknesses of my job performance.

_____ 21. My supervisor knows how well I am performing my job.

_____ 22. I feel my last performance rating represented my true performance.

_____ 23. I believe the present rating system accurately distinguishes between good workers and poor workers.

_____ 24. I have a clear idea of the standards used to judge my job performance.

_____ 25. I believe my immediate supervisor (rather than someone else) should rate my performance.

_____ 26. I feel that performance ratings provide a fair basis for transferring, promoting, or demoting employees.

_____ 27. My supervisor is very interested in my well-being.

_____ 28. I benefit from my performance appraisal.

_____ 29. I feel that my supervisor rates harder than other supervisors.

_____ 30. My supervisor supports my decisions.

_____ 31. My supervisor's evaluation of my performance was accurate.

_____ 32. I believe that OTHER supervisors are very accurate when providing performance ratings of their subordinates.

_____ 33. My supervisor knows what it takes to perform my job well.

_____ 34. I feel that the current rating system is fair.

_____ 35. My supervisor is very honest in his/her dealings with me.

_____ 36. The present performance rating form enables my supervisor to evaluate my performance fairly and accurately.

_____ 37. I believe ___'s performance appraisal program is used for only administrative actions (pay raises, special assignments, promotions, etc.).

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 38. I believe ___'s performance appraisal program is used for only research purposes (unit and/or department effectiveness, program evaluations, etc.).

_____ 39. I believe ___'s performance appraisal program is used for only the growth and development of employees.

_____ 40. I believe ___'s performance appraisal program is used for more than one purpose (growth and development/administrative actions/research).

_____ 41. I believe my supervisor is accurate when he/she rates my job performance.

SECTION II: FEEDBACK SESSIONS

INSTRUCTIONS: This section contains a number of statements about topics related to the ___ Performance Rating Feedback Sessions which your supervisor conducts with you. Read each statement and decide to what degree you agree or disagree with it, according to the scale below. Write the number corresponding to your choice in the blank to the left of each item. Check here (_____) if you did not have a formal feedback session with your supervisor.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 42. I had ample opportunity to express my opinions during the feedback session.

_____ 43. My supervisor and I set specific performance goals or objectives for improvement during the feedback session.

_____ 44. The performance goals or objectives my supervisor and I set during the feedback session were related to my performance weaknesses on specific work dimensions.

_____ 45. The feedback session was conducted in a relaxed atmosphere, without interruptions or distractions.

_____ 46. I am satisfied with the amount of information I get from the person who evaluates me about how well I am performing my job.

_____ 47. I feel that the performance goals and development plan completed during the feedback session are reasonable (i.e., not too easy or too difficult).

_____ 48. After discussing the results of my evaluation with my supervisor, I felt I knew what action I could take to improve my performance.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 49. My supervisor spent enough time discussing the results of my rating with me.

_____ 50. At the conclusion of my most recent feedback session, I knew what my supervisor expected in the way of specific future improvements.

_____ 51. In my feedback session, I was encouraged to present my point of view.

_____ 52. My supervisor exhibited a helping attitude during the feedback session.

SECTION III: GENERAL INFORMATION

INSTRUCTIONS: This section contains several questions related to the ___ performance rating system. Read each question and write your answer in the blank to the left of each question.

_____ 53. How frequently is your performance evaluated?

A. I have never been formally evaluated.
B. Less than once every 2 years.
C. Every 18 months.
D. Every 12 months.
E. Every 6 months.
F. More than twice per year.

_____ 54. During the past 12 months, how many times have you and your supervisor discussed plans for achieving or maintaining high levels of performance?

A. No discussion in the past year.
B. Only once in the last year.
C. About once every 9 months.
D. Once every 6 months.
E. Once every 3 months.
F. Once or more per month.

Thank you very much for your cooperation.

Please make sure you have included the information requested on the cover sheet.

Appendix C

University Sample Questionnaire

Date: November 30, 1982
To: All Classified Employees
From: _____ _____, Employee Relations Manager
Michael Secunda, Performance Evaluation Specialist
Subj: ___ Performance Evaluation Survey

We would greatly appreciate your taking a few minutes to complete the attached form. The questionnaire has been designed to gain information about the Performance Evaluation System used here at the university. The Performance Review sessions give employees the opportunity to present views concerning their performance and encourage communication between supervisors and employees on job-related issues.

Your participation in this survey is strictly voluntary. However, we urge you to complete and return it since this is your opportunity to help us make the evaluation process more useful to you. Please note that all of your responses will remain strictly confidential. Do not put your name on this survey. If you have questions, call _____ _____ at ex. ____ or Michael Secunda at ex. 4747. Please answer all items in the survey and send it in the enclosed envelope through inter-departmental mail by December 10 to Michael Secunda in the Psychology Department, Life Sciences Building.

Thank you very much for your time and cooperation.

Which division of the University do you work for?

_____ Operations & Finance
_____ Academics
_____ University Advancement
_____ Educational Services (computer activities, student affairs, athletics, planning & budget, etc.)
_____ Other--please specify:
_____

SECTION I: PERFORMANCE EVALUATIONS

INSTRUCTIONS: This section contains a number of statements about topics related to the ___ Performance Evaluation System. Read each statement and decide to what degree you agree or disagree with it, according to the scale below. Write the number corresponding to your choice in the blank to the left of each item.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 1. My supervisor observes enough of my performance to rate me.

_____ 2. I am able to have frank and open communications with my supervisor.

_____ 3. I have a good understanding of the performance evaluation system used to rate me.

_____ 4. I would feel free to disagree with my supervisor if I felt his/her evaluation was not fair or did not take all the facts into consideration.

_____ 5. In my last performance evaluation, I received the ratings I expected.

_____ 6. I feel that the performance evaluation system is an acceptable way to evaluate job performance.

_____ 7. I feel my supervisor is as fair as other supervisors in providing ratings of job performance.

_____ 8. My supervisor has a good understanding of what I do on my job.

_____ 9. I feel that performance ratings, in general, accurately reflect how well an employee has performed his or her job.

_____ 10. I would like a performance evaluation system in which I would be rated at least two or more times a year rather than on an annual basis.

_____ 11. Overall, I was satisfied with my last performance evaluation.

_____ 12. My supervisor understands that many things which affect how well I do my job are beyond my control, and rates accordingly.

_____ 13. I feel my supervisor knows how much work I actually do.

_____ 14. I have confidence that my ratings are made on a fair and equal basis with others being rated.

_____ 15. Employees of ___ _____ _____ had an important part in the development of the performance evaluation system (for example, developing job requirements, standards, scales).

_____ 16. I feel that my supervisor would be personally inclined to use his/her power to help me solve a problem in my work.

_____ 17. I think that ___'s rating system serves a useful purpose (i.e., helps identify training needs; helps my supervisor and me agree on job duties; helps improve job performance, etc.)

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 18. When compared to other supervisors, my supervisor is an extremely lenient (very easy) rater when he/she rates my performance.

_____ 19. The current performance evaluation form is acceptable to me.

_____ 20. I accept the judgment of the person who rates me regarding the strengths and weaknesses of my job performance.

_____ 21. My supervisor knows how well I am performing my job.

_____ 22. I feel my last performance rating represented my true performance.

_____ 23. I believe the present performance evaluation system accurately distinguishes between good workers and poor workers.

_____ 24. I have a clear idea of the standards used to judge my job performance.

_____ 25. I believe my immediate supervisor (rather than someone else) should rate my performance.

_____ 26. I feel that performance ratings provide a fair basis for transferring, promoting, or demoting employees.

_____ 27. My supervisor is very interested in my well-being.

_____ 28. I benefit from my performance evaluation.

_____ 29. I feel that my supervisor rates harder than other supervisors.

_____ 30. My supervisor supports my decisions.

_____ 31. My supervisor's evaluation of my performance was accurate.

_____ 32. I believe that OTHER supervisors are very accurate when providing performance ratings of their subordinates.

_____ 33. My supervisor knows what it takes to perform my job well.

_____ 34. I feel that the current performance evaluation system is fair.

_____ 35. My supervisor is very honest in his/her dealings with me.

_____ 36. The present performance evaluation form enables my supervisor to evaluate my performance fairly and accurately.

_____ 37. I believe ___'s performance appraisal program is used for only administrative actions (pay raises, special assignments, promotions, etc.).

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 38. I believe ___'s performance appraisal program is used for only research purposes (unit and/or department effectiveness, program evaluations, etc.).

_____ 39. I believe ___'s performance appraisal program is used for only the growth and development of employees.

_____ 40. I believe ___'s performance appraisal program is used for more than one purpose (growth and development/administrative actions/research).

_____ 41. I believe my supervisor is accurate when he/she rates my job performance.

SECTION II: PERFORMANCE REVIEW AND DISCUSSION SESSIONS

INSTRUCTIONS: This section contains a number of statements about topics related to the ___ performance review and discussion sessions which your supervisor conducts with you. Read each statement and decide to what degree you agree or disagree with it, according to the scale below. Write the number corresponding to your choice in the blank to the left of each item. Check here (_____) if you did not have a formal performance review and discussion session with your supervisor.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 42. I had ample opportunity to express my opinions during my review session with my supervisor.

_____ 43. My supervisor and I set specific performance goals or objectives for improvement during the performance review and discussion session.

_____ 44. The performance goals or objectives my supervisor and I set during the review session were related to my performance weaknesses on specific work dimensions.

_____ 45. The performance review and discussion session was conducted in a relaxed atmosphere, without interruptions or distractions.

_____ 46. I am satisfied with the amount of information I get from the person who evaluates me about how well I am performing my job.

_____ 47. I feel that the performance goals and development plan completed during the review session are reasonable (i.e., not too easy or too difficult).

_____ 48. After discussing the results of my evaluation with my supervisor, I felt I knew what action I could take to improve my performance.

Use this scale to respond to ALL statements:

1	2	3	4	5	6	7
Disagree Strongly	Disagree	Disagree Slightly	Neutral	Agree Slightly	Agree	Agree Strongly

_____ 49. My supervisor spent enough time discussing the results of my rating with me.

_____ 50. At the conclusion of my most recent performance review and discussion session, I knew what my supervisor expected in the way of specific future improvements.

_____ 51. In my performance review and discussion session, I was encouraged to present my point of view.

_____ 52. My supervisor exhibited a helping attitude during the performance review and discussion session.

SECTION III: GENERAL INFORMATION

INSTRUCTIONS: This section contains several questions related to the ___ performance evaluation system. Read each question and write your answer in the blank to the left of each question.

_____ 53. How frequently is your performance evaluated?
- A. I have never been formally evaluated.
- B. Less than once every 2 years.
- C. Every 18 months.
- D. Every 12 months.
- E. Every 6 months.
- F. More than twice per year.

_____ 54. During the past year, how many times have you and your supervisor discussed plans for achieving or maintaining high levels of performance?
- A. No discussion in the past year.
- B. Only once in the last year.
- C. About once every 9 months.
- D. Once every 6 months.
- E. Once every 3 months.
- F. Once or more per month.

_____ 55. How many months have you been employed by ___

_____ 56. How many months have you worked with your current supervisor?

_____ 57. How many months has it been since your last performance evaluation?

_____ 58. Did your supervisor attend a training seminar on how to be a more accurate rater?
- 1 = Yes
- 2 = No
- 3 = Don't Know

_____ 59. For research purposes, it would be helpful if you could provide us with your last overall average performance rating. As with all the previous items, this information will remain strictly confidential.

Thank you very much for your cooperation. Please send this survey through interdepartmental mail to Michael Secunda, Psychology Department, Life Sciences Building.

Appendix D

Factor Loadings for the Six-Factor Model

Factor Loadings for the Six-Factor Model

	Constructs					
Item	Super	Review	Accur	Fair	Accept	General
TRUS1	1.274					1.438
TRUS2	Ø.965					1.267
TRUS3	1.2Ø2					1.575
TRUS4	Ø.9Ø3					1.433
TRUS5	1.134					1.579
SUPR1	1.533					1.111
SUPR3	1.733					1.367
SUPR4	1.387					1.188
SUPR6	1.2Ø6					1.2Ø5
ATMO2		1.787				Ø.582
ATMO4		1.9Ø2				Ø.79Ø
ATMO5		1,85Ø				1.ØØ7
FBGS2		1.555				Ø.277
FBGS4		1.621				Ø.343
FBGS5		1.783				Ø.453
FBGS6		1.975				Ø.516
FBGS7		2.ØØ4				Ø.526
ACUR1			1.464			1.3Ø1
ACUR4			1.771			1.517
ACUR5			1.563			1.61Ø
FAIR1				1.173		Ø.685
FAIR2				1.392		Ø.333
ACPT2					2.Ø56	Ø.514
ACPT3					1.6Ø6	Ø.546
ACPT7					1.652	Ø.176
ACPT9					1.234	1.369

Note. All other loadings were fixed at Ø.ØØ.