

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

**PERCEPTIONS OF SKILL IMPORTANCE RATINGS
FOR PERFORMANCE APPRAISAL: CONTEXTUAL EFFECTS
BY ORGANIZATION AND BY JOB CLASSIFICATION**

by

William Leverne Smith

A Dissertation Submitted to the Faculty of the
THE COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

**DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT**

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 5

UMI Number: 9531086

UMI Microform 9531086
Copyright 1995, by UMI Company. All rights reserved.

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI

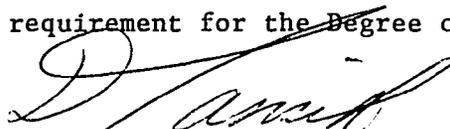
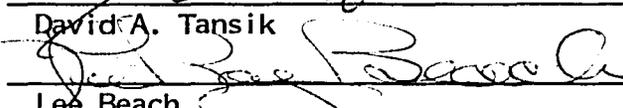
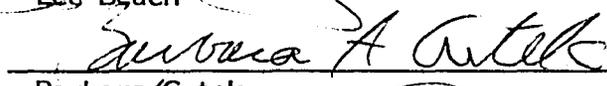
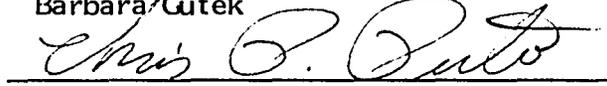
**300 North Zeeb Road
Ann Arbor, MI 48103**

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by William Leverne Smith

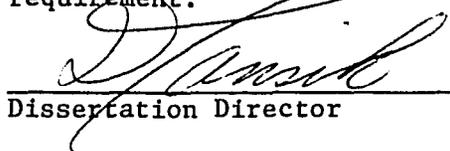
entitled Perceptions of Skill Importance Ratings for Performance
Appraisal: Contextual Effects by Organization and by
Job Classification

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

	<u>2-17-95</u>
David A. Tansik	Date
	<u>2-17-95</u>
Lee Beach	Date
	<u>2-17-95</u>
Barbara A. Gutek	Date
	<u>2-17-95</u>
Chris Puto	Date
_____	_____
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

	<u>2-17-95</u>
Dissertation Director	Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrower under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: *William Jerome Smith*

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the forbearance, guidance, encouragement, and support of many individuals and a variety of groups. The members of my committee (present and past) must be mentioned first: David Tansik, Lee Beach, Barbara Gutek, Bob Tindall, Chris Puto and Helen Anderson. You each provided a unique contribution without which this work would not be the same.

Thanks are due to Sister St. Joan Willert at Carondelet Health Care along with the managers and other staff members at both hospitals who completed my survey forms. Their participation was essential to my data collection and subsequent results.

I want to recognize the contributions of the College of Business and Public Administration at the University of Arizona, especially Vice-Dean Bill Barrett, and the entire faculty and staff of the Management and Policy Department. A special 'thank you' goes to the other students in the U of A Ph.D. program, while I was there, who provided countless critical bits of support from time to time, especially Kris Weatherly, who was there with me the whole way. My best wishes go to each of them.

An acknowledgement to the many clients of our family accounting business during the time of my academic pursuits is only appropriate for their support and encouragement, individually and collectively, over the years, perhaps best represented by Paul Buck.

Support and encouragement have also been received recently from my new university and colleagues at Emporia State University, especially Varkey Titus. It has been appreciated very much.

Finally, but by far the most important, I must thank my family. One younger brother, Barry, provided the early inspiration by earning his doctorate. My oldest daughter, Annette, after she completed her Ph.D., said "Dad, now it's your turn to finish your degree." Allison and Arrion have shared the sacrifices and yet offered continued support and encouragement. My wife, Nancy, has been the greatest. Her contribution to this effort cannot be put in words; it has been immeasurable.

DEDICATION

This dissertation is dedicated to our future; to my wife, Nancy, our three grown daughters and best friends, Arrion, Allison and Annette; and to our mothers: my mother, Eileen Smith Olson, and my mother-in-law, Ruth Bolger.

TABLE OF CONTENTS

LIST OF FIGURES	8
LIST OF TABLES	9
ABSTRACT	11
PERFORMANCE APPRAISAL REVIEW	
Brief History of Performance Appraisal	12
Performance Appraisal Functions	25
Prior Recent Reviews	27
Review of Last Five Years	29
Cognitive processing of information.	30
Rater/ratee personal characteristics.	31
Rating errors and accuracy.	34
Appraisal sources.	35
Performance appraisal formats.	37
Other research issues.	37
Contextual factors and situational characteristics.	39
Skill importance ratings.	43
Implications from the review	44
Implications for current studies	46
Research overview	50
STUDY 1	52
Method	52
Subjects.	52
Materials.	52
Results	54
Discussion	74
STUDY 2	78
Method	78
Subjects.	78
Materials.	78

	7
Results and Discussion	81
CONCLUSIONS	83
DISCUSSION	85
REFERENCES	89

LIST OF FIGURES

1	Factors Influencing the Measurement of Work Performance	22
2	Rating Instrument	52
3	Study 2 Rating Instrument	79
4	Satisfaction with scales	82

LIST OF TABLES

1	Sample Graphic Rating Scale	16
2	Sample Behaviorally Anchored Rating Scales (BARS)	18
3	Behavioral Observation Scales (BOS)	20
4	Hospital A - Mean responses of points assigned each skill category	55
5	Hospital B - Mean responses of points assigned each skill category	56
6	Hospital A - Job Knowledge	58
7	Hospital A - Quality	59
8	Hospital A - Teamwork	60
9	Hospital A - Communications/Interpersonal Skills	61
10	Hospital A - Policy and Procedures Adherence	62
11	Hospital A - Planning and Organizing	63
12	Hospital B - Job Knowledge	66
13	Hospital B - Quality	67
14	Hospital B - Teamwork	68
15	Hospital B - Communications/Interpersonal Skills	69
16	Hospital B - Policy and Procedures Adherence	70

LIST OF TABLES (Cont'd)

17	Hospital B - Planning and Organizing	71
18	T-test results comparing Hospital A to Hospital B for each skill category and each job classification	73

ABSTRACT

This dissertation reports the results of two studies conducted to determine whether perceptions of incumbents would vary across five job classifications in two hospitals on six skill importance categories. 172 subjects distributed 100 points across the importance categories in accordance with their perceptions of “what it takes to do a good job.” Significant differences were found across job classifications and organizations. A second study of 30 nurses found they were able to recognize their weightings on the skill importance categories when presented in a performance appraisal setting and rejected equal weighting and inverse weighting options. Future research implications are discussed.

PERCEPTIONS OF SKILL IMPORTANCE RATINGS FOR PERFORMANCE APPRAISAL: CONTEXTUAL EFFECTS BY ORGANIZATION AND BY JOB CLASSIFICATION

Performance appraisal takes many forms in different contexts. The empirical studies reported here demonstrate the need to tailor appraisal approaches to the specific organizational setting and the specific job classification. A review of the literature discloses an increased recent interest in contextual and situational variables as well as in skill importance ratings as they are applied to the evaluation of work performance. The present studies extend the research in these areas.

Performance Appraisal Review

Brief History of Performance Appraisal

Evaluation of work performance is as old as organizations. It is also as complex as any other human endeavor. Research throughout this century has taken many forms and provided many insights, but, adequate and accurate appraisal is still very elusive.

Formal performance appraisal probably began in the United States in 1813 when Army General Lewis Cass submitted to the War Department an evaluation of each of his men using terms such as “a good-natured man” or “knave despised by all.” This “man-to-man” rating

technique was based on trait psychology and adopted by industry following World War I. It evolved into the graphic rating scales still widely used today. Appraisal of industry employees became popular after World War I but appraisal of managers was not widely practiced until after World War II. In 1962, performance appraisal was conducted in 61% of the organizations surveyed (Spriegel, 1962). After the passage of the 1964 Civil Rights Act and the 1966 and 1970 Equal Employment Opportunity Commission (EEOC) guidelines for regulation of selection procedures, legal considerations exerted strong pressure on organizations to formalize their appraisal systems and to make the required efforts to prove they are valid (DeVries, Morrison, Shullman, & Gerlach, 1986). Federal legislation and the civil rights and women's movements of the 1960's and 1970's created the need for rapid improvement in organizational appraisal practices (Murphy & Cleveland, 1991). Performance appraisals were increasingly used for employee development and feedback, corporate planning, legal documentation, systems maintenance, and research.

Research efforts during the 1940's and 1950's included 1) work on the rating process, 2) search for the "ultimate criterion," and 3) classification frameworks. One of the most important contributions of the rating process research was the model developed by Wherry (1952). It

drew on psychometric and cognitive research and decomposed an observed rating into ratee performance, rater observation, rater bias, and error components. R.L. Thorndike's (1949) "ultimate criterion" stimulated intellectual debate about criteria and served as a statement of a goal that criterion specialists strive to attain (Austin & Villinova, 1992). Marx (1963) said the ultimate criterion is a construct used by the researcher as a theoretical "tool and goal." Classification frameworks were developed for partitioning sources of variance and covariance among predictors, actual criteria, and the ultimate criteria. The terms deficiency, contamination and relevance became organizing concepts. These developments converged on the idea that job performance is multidimensional and subject to various systematic and random fluctuations during measurement (Austin & Villinova, 1992).

During the 1960's and 1970's attention shifted from measuring outputs and examining personal traits to gathering behavioral measures as criteria. Smith and Kendall's (1963) development of the BARS (Behaviorally Anchored Rating Scales) system provided a method of measuring process aspects of performance. The BARS approach, unlike other ratings formats, was designed to use the stereotyping capabilities of raters to promote a common nomenclature among raters with respect to labeling observed incidents of performance. It was also unique in that it

was designed to standardize not only the rating process but also the observation process. The BARS methodology, indeed, took what had previously been viewed as an obstacle to rating agreement (i.e., the human tendency to stereotype and simplify) and applied it to facilitate rater agreement.

Prior to the introduction of the BARS approach, graphic ratings scales had been extensively used, tested and researched since first introduced in 1922. These graphic ratings scales begin with a list of traits or characteristics. Raters are asked to rate each employee on each item of the list by checking a number or box that represents how much of each trait or characteristic is present. Research attempted to determine anchors, response categories, etc. to make the graphic ratings scales useful in more situations. Table 1 illustrates a graphic rating scale for a factory worker. On each Rating Factor, the worker gets a check under either Unsatisfactory (US), Conditional (C), Satisfactory (S), Above Satisfactory (AS) or Outstanding (OS). Sample check marks are included in the table.

Table 1
Sample Graphic Rating Scale

<u>Rating Factors</u>	<u>US</u>	<u>C</u>	<u>S</u>	<u>AS</u>	<u>OS</u>
Attendance				x	
Appearance			x		
Dependability			x		
Quality of Work				x	
Quantity of Work				x	
Relationship with People		x			
Job Knowledge			x		

Source: Cascio, 1992

Behaviorally Anchored Rating Scales (BARS) represented a systematic attempt to enhance the graphic rating scale (cf. Smith and Kendall, 1963). Actual work behaviors are represented by a number of behavioral statements arranged along a continuum from poor work performance to the descriptive statement of the best performance on that work element. Raters circle a number representing the statement which

best describes the employee's behavior. A portion of a BARS for a Resident Adviser, University Housing is reproduced in Table 2.

Table 2**Sample Behaviorally Anchored Rating Scales (BARS)****Performance Dimension**

Concern for individual dorm residents: Attempts to get to know individual residents and responds to their individual needs with genuine interest. This resident advisor could be expected to:

Rating scale:

(1) Recognize when a floor member appears depressed and ask if person has problem he/she wants to discuss.

(2) Offer floor members "tips" on how to study for a course he/she has already taken.

(3) See person and recognized him/her as a floor member and say "hi."

(4) Be friendly with a floor member; get into discussion on problems, but fail to follow up later on with student.

(5) Criticize a floor member for no being able to solve his/her own problems.

Source: Milkovich & Newman, 1987

Development of BARS set off a flurry of activity by practitioners and researchers alike. No less than five different behavior-based scales were introduced shortly thereafter including mixed standard scales (Blanz & Ghiselli, 1972), behavioral expectation scales (BES; Smith & Kendall, 1963), behavioral observation scales (BOS; Latham & Wexley, 1977), behavioral discrimination scales (BDS; Kane & Lawler, 1979), and behavioral standard scales (Borman, 1986).

The Behavioral Observation Scales (BOS) developed by Latham and Wexley (1977), for example, attempted to improve BARS by including the frequency of specific behaviors and critical incidents observed. Table 3 shows an excerpt from a BOS used in a service organization which demonstrates the use of these scales.

Table 3**Behavioral Observation Scales (BOS)**

Each are rated on a five point scale anchored by “Almost never” and “Almost always.”

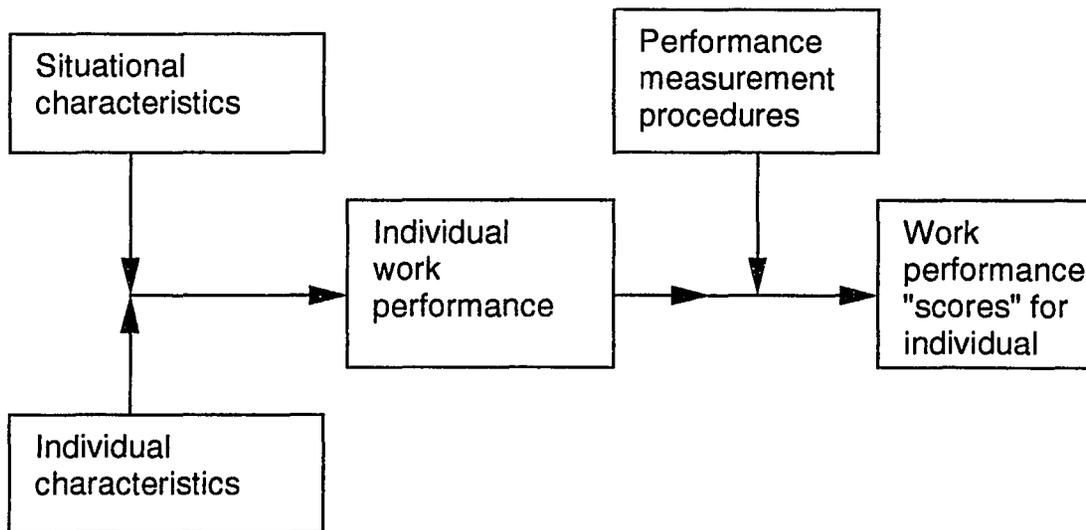
For waiter/waitress:

1. Comes to work on time.
2. Uses the words “please” and “thank you” when talking to fellow customers.
3. Stops talking to a fellow employee as soon as a customer approaches the counter.
4. Asks the customer if everything is satisfactory.

Source: Latham & Wexley, 1977

The development of these new scales was followed closely by several studies in support of training of raters as a means of reducing rater errors (cf., Bernardin, 1978; Borman, 1975; Brown, 1968; Latham, Wexley, & Pursell, 1975). Several tactics were reported which modeled judgments of performance including policy capturing being applied to performance appraisals (Zedeck & Kafry, 1977). This period saw rating format research reach its zenith in popularity along with research on rater training and rater individual differences because all the effort employed had failed to identify one rating format that was superior (Austin & Villanova, 1992).

Landy and Farr (1980) were so disappointed with the format research after wading through it for a review that they called for a moratorium on such efforts. They were disappointed because they found that while rating scales were well-developed and possessed certain critical characteristics (e.g., defined performance domain, behavioral anchors, and 3-9 scale points), the differences in the quality of ratings across different formats were negligible. Their contribution was a model (See Figure 1) with orientation toward the cognitive processes of raters (Austin & Villanova, 1992).

Figure 1**Factors Influencing the Measurement of Work Performance.**

Source: Adapted from Landy and Farr (1983), Figure 1.2, p.8.

Landy and Farr (1983) noted that the difficulty of measuring work performance is one of the most vexing problems facing industrial-organizational psychologists. Reliable, valid and accurate performance measures are sought which are also practical and useful. In addition, performance measures should provide sufficient information for their cost of attainment, should be meaningful to those whose performance is being assessed, and should be related to personal and organizational goals.

Factors influencing the measurement of work performance are shown in Figure 1. Both individual characteristics and situational characteristics are noted as factors influencing the individual's work performance. Individual characteristics include ability (such as cognitive, physical, social, and emotional factors; past work experience; education; training), motivation (level of effort expenditure), and role perceptions (the individual's beliefs about what constitutes effective performance of his or her job; cf. Porter & Lawler, 1968). Situational characteristics are broadly defined to include the entire work setting other than the individual whose performance is being measured, including managers/supervisors, peers, work design, reward system, organizational structure, and other environmental characteristics.

Figure 1 also shows that performance measurement procedures are

applied to individual work performance to generate work performance “scores” for individuals. Note especially the distinction between the actual work performance of the individual and the work performance “score” of the individual as obtained from some performance measure. Whatever measure is developed and used is an imperfect translator of behavior into some quantified index of work performance (Landy and Farr, 1983).

Wexley and Klimoski (1984), in their review and update of performance appraisal literature, state that Landy and Farr (1980) “took a needed and critical look at many years of research in the area and pointed out in detail several directions for future research...A fair number of recent papers seem to have been stimulated by this review.” The many directions of future research that are still being pursued can be seen from the following list of elements from the Landy and Farr (1980) process model (an earlier version than that shown in Figure 1) of performance rating (most are the same used in the most recent reviews noted below):

1. Purpose of rating
2. Rating process
3. Cognitive processing
4. Rating instrument
5. Scale development
6. Rater characteristics

7. Ratee characteristics
8. Organization characteristics
9. Position characteristics.

Performance appraisal functions

Performance appraisal information serves at least two functions. It provides feedback to the individual about the adequacy of past behavior, and it serves as a basis for personnel decisions linked to individual rewards (Ilgen & Feldman, 1983). A critical decision for management is whether or not these two functions should, or can, be separated. Ilgen and Feldman state that most people follow the implications of the General Electric research (Meyer, Kay, and French, 1965) that these two functions should be separated. That is, performance appraisals intended to provide performance feedback, and counseling of employees should not occur at the same time that the reward implications of performance are discussed. Ilgen and Feldman make a strong case that the functions should not be separated. They say this advice seems faulty because recipients of feedback often cannot avoid considering reward implications even when they are not explicitly discussed. The appraiser is often in a position of authority, thereby controlling many of the individual's rewards. Attempting to separate the two promotes a benign fiction at best,

hypocrisy at worst (Ilgen & Feldman, 1983).

In their followup to the General Electric study, Lawler, Mohrman, and Resnick (1984) offer the following insight:

Performance appraisal is both a personal event between two people who have an ongoing relationship and a bureaucratic event that is needed to maintain an organization's human resource management system. Therefore, it is a major mechanism for integrating the individual and the organization. As such, it will always be subject to contradictory purposes, misperceptions, miscommunications, and some ineffectiveness. On the other hand, our data suggest that there are some ways to make it go better and that it is worth investing time and effort to do it well. At best, it's two people sharing their perceptions of each other, their relationships, their work, and their organization - sharing that results in better performance, better feelings, and a more effective organization. At its worst, it is one person in the name of the organization trying to force his or her

will on another with the result of miscommunication, misperception, disappointment, and alienation. The best is achievable, but only with considerable effort, careful design, constant attention to process, and support by top management (p. 35).

Prior Recent Reviews

Three reviews of the evaluation of work performance have been conducted in recent years. (Note: “evaluation of work performance” and “performance appraisal” are used interchangeably in this dissertation.)

Two of these reviews are of special note because they were published under the banner of APA CENTENNIAL FEATURE. The first (Katzell & Austin, 1992) reviews the development of industrial/organizational psychology in the United States from the turn of the century and provides an excellent background and overview. Evaluation of work performance is a subfield of industrial/organizational psychology and included in this review. They concluded that industrial/organizational psychology is a dynamic field that has been making contributions both to the science of behavior and to industrial society. Among the problems still

facing the field, they found, is the disjunction between science and practice. The second review (Austin & Villanova, 1992) is even more relevant to this paper as it reviews "The Criterion Problem" from 1917 to 1992. For each of the periods 1917-1939, 1940-1959, 1960-1979, 1980-1992 the review provides an overview, dimensions, methods, categorizing frameworks, and a summary with extensive citations. They found a shift away from brute prediction toward a balanced treatment of both empirical and conceptual issues highlighted by calls for the validation of criteria and by increased attention to modeling performance, as well as a recognition of multiple perspectives and competing values from which to view performance and criterion measurements. Rather than seeking an "ultimate criterion" they suggest multiple perspectives and competing values from which to view performance and criterion measurement.

The most recent review of performance appraisal research and practice is Bretz, Milkovich, and Read (1992). They examined literature published in both academic and practitioner outlets between 1985 and 1990. They found that performance appraisal research and practice converge on many issues and diverge on others. Divergence on some issues, they suggest, is not necessarily a problem because relevancy for decision makers is not the purpose for all research efforts. However, since performance evaluation is an applied subject, research should eventually

lead to improvements in practice. Continued reliance on student samples and laboratory settings, they note, is not facilitating the transfer of research into application. They found that the collection of studies on information processing to be the most serious, concentrated attempt to date to better understand the rating process. While they urge continued research in that framework to better understand the rating process, they also state that attention must be paid to the potential effects of situational or contextual variables. They note that examining appraisal issues in sterile environments may isolate the effects many researchers wish to investigate, but it also limits the generalizability of the results and removes the issues from the attention and interests of human decision makers. In addition, they reported on three surveys of current performance appraisal practices (First, the Wyatt Company survey of 3052 organizations (Wyatt, 1989), second, a survey by the Conference Board of its 435 member organizations (Milkovich & Wigdor, 1991), and, third, their own survey of the Fortune Industrial 100), and stated:

Most organizations do not systematically collect data to determine either the managers' or the employees' perceptions of fairness of the appraisal process or the results obtained....

Managers identified fairness as the most important performance appraisal issue organizations face. They also

tend to be very concerned that the appraisal system be an effective tool to manage future performance, not just one that reflects past performance. Managers indicated that they are most likely to use performance information for improving future performance, making pay distribution decisions, and communicating expectations regarding future performance (p. 332).

They conclude that if research is to inform practice, interaction between researchers and managers and application of research results are important.

The next section is a review of published articles from 1990 to the middle of 1994 along with comments comparing and contrasting this period with the earlier period (1990 to 1994 compared to 1985 to 1990) reviewed by Bretz, Milkovich, and Read (1992).

Review of Last Five Years

As an organizing structure for this review, the topic headings adopted by Bretz, Milkovich, and Read (1992) for their review will be used in this section. These topic headings are: Cognitive processing of information, rater/ratee personal characteristics, rating errors and accuracy, appraisal sources, performance appraisal formats, and other

research issues. Because of the increased frequency of articles appearing during the current review period related to two additional topics, the following headings are also used: contextual factors and situational characteristics; and, skill importance ratings.

Cognitive processing of information.

Woehr and Feldman (1993) extended earlier work on information processing in performance appraisal judgments by examining the relation between memory and judgment processes. Performance appraisal researchers have traditionally assumed that performance judgments are based on memory for specific behaviors assuming that as memory for specific behaviors improves, judgment accuracy should also improve. Their results of testing three structural models of the relationship between memory and judgment found that the causal relation between memory and judgment is driven by contextual factors at the time ratings are required as well as at the time information is encoded. More specifically, they found that the relationship between memory and judgment is much more flexible than previously proposed which implies that the causal priority between memory and judgement can be influenced through contextual manipulation both before observation and at the time the judgment is required. For example, when no evaluation has been formed before some

required judgment, memory serves as the basis for judgment. However, when some prior evaluation has been formed, memory may still influence judgment, especially when memory for specific behaviors is highly accessible. In addition, the actual rating format, another contextual cue, can influence the way in which individuals process performance information.

Rater/ratee personal characteristics.

Miller, Kaspin, and Schuster (1990) examined the impact of personal and organizational factors and performance appraisal methods on 53 Age Discrimination in Employment Act (ADEA) federal courts cases and found that appraisal system characteristics which in previous studies were found to be significantly related to case outcome were not reported in these case decisions. Previous research (e.g., Feild & Holley, 1982; Study of Courts Decisions, 1984) concluded, for instance, that successful employer-defendants tended to have formal appraisal systems based on job analysis and characterized by behavior-oriented questions, specific written instructions to raters, and feedback of results to employees. These studies concluded that the courts follow the strict standards from the EEOC employee selection guidelines. Miller et al. (1990) found that the Uniform

Guidelines on Employee Selection Procedures were not being applied to performance appraisals in ADEA litigation. Furthermore, formal performance evaluation procedures were not required for an employer-defendant to mount a successful defense. For example, the credibility of the supervisor's testimony appeared to play an important role in case outcome.

One study (Hartel, 1993) investigated the proposition that differences in rating accuracy associated with different ratings formats are contingent on rater characteristics. It compared field-dependent and field-independent raters on four different performance measures. Field dependence refers to cognitive dependence on the external organization of information, whereas field independence refers to the ability to impose organization on information independent of the form in which it is perceived. The results of the study were that field-independent raters were more accurate raters than field-dependents when performance formats are holistic rather than decomposed (i.e., each of four dimensions used in the format consisted of three components, treated individually, that is, decomposed, or treated holistically, that is, combined). Field-independents were also more confident in their ratings and less frustrated and confused with the rating task than were field-dependents.

While not specific to appraisals only, Cleveland and Shore (1992)

examined person- and context-oriented definitions of age as they relate to work attitudes, performance ratings, and reports of developmental practices. The five age measures used in the study were employee chronological age, employee subjective age (i.e., self-perceptions of age), social ages (i.e., others' perceptions of age), as well as self- and supervisors' perception of the employee's relative age (i.e., compared with the employee's work group). The results of the study emphasized the importance of conceptualizing age more broadly. In particular, both perceptual and contextual age measures (including the person by context interactions) provided better predictions of a variety of work criteria than did chronological age alone. They concluded that there is clearly a need to include measures of the perceived age context in order to more fully understand the work attitudes and judgments made by and about older workers. Performance ratings included both self ratings and manager ratings of performance, promotability and managerial potential. The findings suggest that chronological age may not be directly associated with organizational criteria such as performance ratings but, rather, chronological age appears to have the greatest predictive power when combined with perceptions of the ages of other employees in the work setting.

Rating errors and accuracy.

Nathan and Tippins (1990) reported a positive relationship between halo and accuracy. Their study of the performance ratings of 294 clerical workers, measuring halo as the standard deviation across dimensions, found that greater halo was related to higher performance ratings. A critical analysis of the many studies on the nature and consequences of halo error (Murphy, Jako, & Anhalt, 1993) reported that all of the four major elements of the conception of halo research reviewed (i.e., (a) halo error was common; (b) it was a rater error, with true and illusory components; (c) it led to inflated correlations among rating dimensions and was due to the influence of a general evaluation on specific judgments; and, (d) it had negative consequences and should be avoided or removed) were either wrong or problematic. An additional finding was that overall ratings of performance, rather than being highly contaminated by halo “error,” appeared to be as good if not better (i.e., more predictable) than the more specific dimension ratings. They went on to suggest specific directions for future research on halo that take into account the context in which judgments are formed and ratings are obtained and that more clearly distinguish between halo errors and the apparent halo effect. These results, they concluded, suggest:

Not only “that halo error is not nearly as serious a problem as

has generally been presumed” (Kozlowski & Kirsch, 1987, p. 259), but that halo may serve as a valuable role in ensuring that raters consider the “employee as a whole” when conducting performance ratings rather than attending too much to specific, but unrepresentative, critical incidents. In the future, researchers may want to examine how and why halo increases rather than decreases rating accuracy (p. 296).

Maurer, Palmer, & Ashe (1993) reported on contrast effects using diaries and checklists as moderators. When the contrast effect occurs, ratings of target behavior are contrasted away from the level of behavior observed in the same or a preceding context. For example, an average interviewee may get a poor rating if he or she is interviewed after a good interviewee. Their results were consistent with the traditional conception of contrast effects: upward in a poor context and downward in a good context.

Appraisal sources.

The role that groups might play as performance raters was examined in a study which compared the behavioral rating accuracy of groups versus individuals (Martell & Borg, 1993). Using a behavioral checklist in a laboratory study, 191 subjects rated the behavior of a police

officer, as presented in a vignette of effective and ineffective work behaviors, individually or in 4-person groups. Ratings were made immediately or after a 5-day delay. Measures of memory sensitivity and decision criteria indicated that in the delayed rating condition, groups remembered the behaviors more accurately than did individuals, whereas in the immediate rating condition, groups and individuals did not differ. The results suggest that groups can be a help, but they are not a panacea for the problems of rating accuracy, e.g., difficulty in accurate recall of critical incidents, or recency and representativeness biases.

Williams and Levy (1992) reported a study to examine the effect of perceived system knowledge (PSK) on the discrepancy between self-and supervisor ratings. The results suggested that the level of knowledge did in fact influence the agreement among subordinates and supervisors. That is, self-ratings were more congruent with supervisors' ratings when the subordinates reported high levels of perceived system knowledge.

Anticipated feedback sharing and knowledge of subordinates' self-assessments were investigated as two factors felt to influence the quality of ratings (Klimoski & Inks, 1990). It was found that despite their potential to influence ratings, neither factor was necessarily detrimental, but rather care should be taken to minimize their potential to reduce rating quality.

Performance appraisal formats.

One study (Steiner, Rain, & Smalley, 1993) compared a distributed rating format with a Behavioral Observation Scale (BOS) rating form which found that the distributional ratings were more sensitive to variability in performance.

Other research issues.

Borman, White, Pulakos, and Oppler (1991) reported a set of causal models which partially confirmed Hunter's (1983) earlier model of supervisory job performance ratings and presented expanded models incorporating additional variables that accounted for more than twice the variance in ratings than Hunter's model. Hunter's model included cognitive ability, job knowledge, task proficiency, and ratings. The Borman, et al. (1991) expanded model used these variables from Hunter's model plus technical proficiency and ratee problem behavior which had direct effects on supervisory ratings and ratee ability and dependability which played strong indirect roles in the new model. DuBois, Sackett, Zedeck and Fogli (1993) reported additional results from their study of typical and maximum performance criteria. First, they examined the growing recognition of the importance of distinguishing between typical

and maximum performance criteria. Second, they provided clarification of definitional and measurement issues. Finally, they reported additional results of an empirical study first reported in the prior period (Sackett, Zedeck, & Fogli, 1988). The relationship of appraisal satisfaction to several appraisal characteristics was explored by Dobbins, Cardy, and Platz-Vieno (1990). Their results provided support for conceptualizing appraisal as a contingent function of both appraisal characteristics (e.g., action plans, frequency, and rater training) and organizational variables (e.g., span of control, role conflict, and role ambiguity). Jordan and Jordan (1993) found moderate positive correlations between ratings of performance appraisal and satisfaction with those ratings by head nurses. Four questions regarding satisfaction with ratings were asked shortly after the nurses had received ratings on a 138-item mixed-standard scale. Each of the four "satisfaction with ratings" questions was answered on a 5-point scale in Likert-format anchored by 1 as "highly dissatisfied" and 5 as "highly satisfied." Pearson correlations of ratings with satisfaction responses ran from 0.37 to 0.52.

Contextual factors and situational characteristics.

The terms contextual factors and situational characteristics are each used by different sets of researchers. For purposes of this dissertation

they may be used interchangeably or together with the general meaning attached by Landy and Farr (1980; 1983) noted earlier:

Situational characteristics are broadly defined to include the entire work setting other than the individual whose performance is being measured, including managers/supervisors, peers, work design, reward system, organizational structure, and other environmental characteristics.

Giles and Mossholder (1990) sought to extend the development of measures that assess system contextual aspects of performance appraisal, to investigate the relationship between contextual variables and employees' reactions to performance appraisal, and to assess the extent to which system contextual variables (e.g., complexity, implementation, and follow-up), as compared with supervisory session variables (e.g., participation, goal setting, and criticism), were related to employee satisfaction with appraisal. The results suggest that additional attention be given to these and other characteristics of the appraisal system in future research on contextual influence in the performance appraisal process.

Zalesny (1990) investigated how rater confidence in evaluation affects the outcome of the appraisal process when information and

potential influence are provided by multiple sources. Rater confidence was defined as a contextual factor based on prior findings that expert, competent, or high-status individuals are more likely than nonexpert, incompetent or low status individuals to successfully exert social influence, especially where that expertise, competence or status is recognized by the ratee, for instance (cf., Bottger, 1984; Yetton & Bottger, 1982). Evaluation information supplied by a highly confident, extreme rater significantly influenced evaluations from noninteracting novice raters but only minimally influenced evaluations from interacting novice raters.

Kravitz and Balzer (1992) reported a methodological critique and empirical study on context effects in performance appraisals. They demonstrated that the design of a study of context effects can preclude unambiguous interpretation of results and called into question the results of all previous studies of context effects.

Judge and Ferris (1993) employed a set of key social and situational variables to investigate the overall relationship between social processes and performance ratings. The model tested demonstrated significant relationships between job performance ratings and the following social and situational variables: supervisor-subordinate demographic similarity and work relationship, supervisor's affect toward subordinate,

supervisor's opportunity to observe subordinate's job performance and supervisor's inference of subordinate's self-ratings of job performance. The support for the model of the performance-rating process demonstrated that performance rating is a process with multiple social and situational facets that need to be considered simultaneously.

Kellogg and Chase (1994) measured twenty-seven contextual variables as they empirically derived a measurement model for customer contact, a widely used construct in service management. Multidimensional scaling showed that the construct of customer contact is multidimensional and complex. An interval scale was developed using the paired comparison methodology, and a measurement model was developed using this contact scale. The central finding was that the degree or level of contact can be measured at the episode level by averaging just three of the contextual variables examined: normalized values of communication time, the information richness, and the level of intimacy. They suggest that many service system design issues (e.g., employee selection and appraisal, training, and empowerment) are dependent upon customer contact. The customer contact construct can also be used to develop contingency models in various situations.

Skill importance ratings.

Skill importance ratings are commonly used as part of job analysis procedures to determine the skills, behaviors, activities, and/or tasks required of a job incumbent. Such ratings often require judgments of the relative importance of a skill with regard to successful performance by a job incumbent.

Waldman, Yammarino, and Avolio (1990) used a multiple levels of analysis approach to assess whether job performance and skill importance ratings were based on individual differences, incumbent/supervisor dyadic agreement, group membership, or functional department. Results indicated that (a) incumbent ratings of both their own performance and their own skill importance perceptions were characterized by individual differences, (b) supervisory ratings of both their subordinate's performance and skill importance perceptions were primarily characterized by between-groups (supervisors) differences, (c) matched incumbent-supervisory performance ratings displayed individual differences, and (d) matched incumbent-supervisory importance ratings displayed dyadic agreement (i.e., between-dyads differences). A subsequent analysis and report on the same database (Yammarino & Waldman, 1993) investigated relationships between the performance

levels of incumbents on job skills and the importance attributed to those job skills. Supervisors and job incumbents rated both the performance of skill dimensions and the importance of those skills for respective incumbents' positions. Results indicated that the ratings of performance and importance on matched skill areas correlated positively, although relationships were somewhat stronger for incumbent data as compared with data provided by supervisors.

Implications from the review

Through the years of this century scientists have primarily focused their attention on finding the universal meanings and the universal tools for measuring and understanding the events of human activity. It has only been in the last thirty years or so that serious consideration has been given to the importance of differences. So-called "contingency theories," for instance, have begun to appear in just about every field of study. Rather than seeking the "universal" answer, contingency theories suggest that any workable solution to a problem "is contingent upon" the context or the situational characteristics of the problem. In my study of the reviews and many of the individual papers cited above, this theme (the need to examine the context or situational characteristics) with respect to the study of work performance evaluation and performance appraisal was

recurrent.

Landy and Farr (1980) considered organization characteristics and position characteristics to be crucial elements of their work performance appraisal model. Situational and contextual factors represented four of the nine elements, nearly half, of their model. Bailey (1983) asserted that dimensions of criteria are context dependent. She noted that use of the same measure to represent a conceptual criterion in one settings might prove unsuitable in a different context. Bretz et al. (1992) called for multiple perspectives from multiple organizations to better understand performance appraisal systems. They also said that research is needed to examine the variety of situational variables (e.g., environmental, organizational, and dispositional factors) that affect appraisal design and administrative choices. Austin and Villanova (1992) noted from their review that work performance “may be more situationally specific than KSAOs (knowledge, skills, abilities and other factors) used as predictors...searching for general performance constructs may be futile and more refined taxonomies will be required to match criterion measures or composites to specific situations (p. 862).” Kellogg and Chase (1994) have shown that the customer contact construct may represent an important set of contextual variables that should be considered when examining situational characteristics.

Implications for current studies

One possible operationalization for examining differences in contextual or situational variable effects on performance criteria would be to study a series of different jobs within a large organization and then do the same study across similar jobs in another comparable but different organization. Significant differences between jobs and/or between organizations would indicate that situational differences do exist even where many similarities are assumed. Measures need to be included of a number of situational characteristics in order to determine their impact on the differences observed and measured. For example, it is not unusual for an organization to use an identical performance appraisal instrument for all employees evaluated by the appraisal system even though their job situations vary a great deal. What elements of the jobs or the job contexts are important for accurate measurements?

A major regional hospital in a large southwestern city (visited by the author prior to the start of the studies reported here) used the same instrument consisting of the same six items to do performance evaluations for all staff personnel from janitors to nurses. Each of the six items was weighted the same whether being used for food service workers or for radiology technicians. Quality, as a criterion, was given the same weight

as Planning and Organizing, for instance. Job Knowledge was weighted equally with Teamwork. But aren't these jobs different? Shouldn't, perhaps, Job Knowledge be more important for nurses than for housekeepers? The customer contact dimension, for instance, was shown by Kellogg and Chase (1994) to vary by job classification. For example, the contact score was shown to be 5.10 for nurses and 2.13 for housekeepers in a hospital setting. Might this complex contextual variable provide some insight into how jobs vary?

What would the worker's perceptions be? Would a nurse believe that Job Knowledge was more important in his/her job than Policy and Procedure Adherence? These kinds of measures of perceptions of incumbents in jobs are called skill importance ratings. Skill importance ratings are commonly used as part of job analysis procedures to determine the skills, behavior, activities, and tasks required of an incumbent. Also, in training needs analysis, for instance, importance ratings are collected to gain a better understanding of the skills necessary for potential trainees to perform effectively in their jobs (e.g., McGehee & Thayer, 1961; Moore & Dutton, 1978; Waldman, Yammarino, & Avolio, 1990). Job analytic research has shown that importance ratings do vary systematically depending on work group membership, functional area, and/or department (Dowell & Wexley, 1978; Ford & Noe, 1987; Pravett & Lau,

1983; Tornow & Pinto, 1976; Waldman, Yammarino, & Avolio, 1990). The results of this previous research, then, suggest that skill importance ratings will vary by job classification within an organization. Also, as noted earlier in the review section, skill importance ratings and work performance ratings have been used together in studies to seek a better understanding of performance appraisal systems.

The findings of Zammuto, London, and Rowland (1982) suggest that organizational differences may act to circumscribe further the generality of the findings of appraisal research across organizational settings; the validity of research findings based on appraisal participants in one applied setting may not generalize to other people and settings. This observation harkens back to Bailey's (1983) contention that criteria are context-sensitive (Austin & Villanova, 1992). Therefore, it is expected that skill importance ratings gathered from incumbents of two different organizations will vary by organization/institution even though the two organizations are rather similar.

The present set of studies seek to extend these findings in a service organization context. The set of studies reported below examines skill importance ratings in a range of jobs in two hospital settings.

The hospital setting is especially useful for a study of this kind because of the diversity of job classifications within each hospital as well

as the similarity of job classifications between hospitals. Performance appraisals are conducted regularly. Service workers have regular contact, in varying degrees, with customers and are constantly seeking ways in which to improve customer satisfaction. Surveys are a normal part of staff meetings and research studies are a regular part of hospital life.

Previous studies (e.g., Kellogg, 1991; Kellogg & Chase, 1994) of contextual factors in a hospital setting provide potentially useful comparative information. Their findings suggest, for instance, that the degree of contact that service workers have with patients in the hospital, measured by a combination of three behavioral scales, may be a useful contextual characteristic measure for further study.

Finding significant differences between job classifications and between the institutions in the present set of studies would suggest that serious consideration be given to differential weighting of the criteria for use in performance appraisal of the incumbents of these jobs as well as between institutions.

Failure to recognize the need for differential weightings between jobs and between institutions brings the accuracy of the performance appraisal system into question.

Research overview

The research studies reported here draw on employees' perceptions of a set of skill importance categories as they might be used in a performance appraisal system. The context of the studies is a set of two hospitals located in a large city in the southwestern part of the United States.

The two hospitals chosen are physically located in the same city approximately eight miles from each other, are owned by the same not-for-profit organization, serve similar constituencies and offer generally similar services. For these reasons, any significant differences noted in perceptions of the employees between the organizations would be of special note to management.

Within each hospital, five job classifications were chosen for study which represented an array of service workers along the customer contact dimension (Kellogg & Chase, 1994). While the job classifications used are distinguishable by other attributes as well, of course, the customer contact dimension was one possible contextual variable examined in these studies. Age group and gender were collected for the purpose of ascertaining any variance that might be attributable to these demographic differences.

The skill importance categories, which each employee was asked to rate relative to the importance each item held for his or her job, were

adapted from the actual elements used by another hospital in the same city. These elements were also chosen for these studies because they were very similar to the elements derived empirically by Waldman, Yammarino, and their colleagues (Waldman, Yammarino, & Avolio, 1990; Yammarino & Waldman, 1993) in their studies in another industry. The intent of this research was to examine relationships not to test any specific skill importance categories.

If people are to be appraised, and they frequently are, either for improving future performance or for making pay distribution decisions, then organizations should make, I believe, reasonable efforts to, first, gather manager and employee perceptions about the appraisal process, and, second, use that information to design and implement appraisal systems that are seen as possibly more fair by reflecting the perceptions of those managers and employees.

Study 1

Method

Subjects.

Participants were 172 service workers in five job classifications of two hospitals located in a large city in the southwestern part of the United States. Job incumbents in similar departments in the two hospitals typically had the same job title and common job description. For instance, in the two nursing departments, the incumbents were “nurses” and in the two radiology departments, the incumbents were “radiology technicians.”

The other job classifications used, in addition to nurses and radiology technicians, were food service workers, housekeepers, and admissions clerks. These five classifications were chosen to broadly represent the array of occupations within the hospital settings.

Data were collected at regularly scheduled staff meetings over a two month period with no duplication of shifts or personnel. Rating instruments were administered, received and reviewed by the researcher at each meeting to assure correct completion and appropriate category of each respondent.

Materials.

The rating instrument listed six skill importance categories (a copy of the rating instrument is included as Figure 2) along with a category of

Figure 2 Rating Instrument

Performance Factor Weightings

The following seven factors are frequently used for performance evaluations. Please assign 100 points across these factors (0 to 100 points per factor to total 100 in all) based on how they apply to your job - what it takes to do a "good job."

___% **Job Knowledge** Demonstrates ability to apply knowledge to assignments and maintain expertise in his/her field by keeping current with new developments.

___% **Quality** Is thorough, accurate and carefully attends to detail in performing work; completes tasks according to established standards; participates in setting work standards and in collecting data to review actual quality outcomes.

___% **Teamwork** Achieves results with and through others; interacts with other individuals/departments to accomplish tasks.

___% **Communications/Interpersonal Skills** Expresses ideas, both oral and written, that are clear, concise, and accurate and interacts with others in a caring and effective manner.

___% **Policies and Procedures Adherence** Follows organizational and departmental policies and procedures, such as policies on attendance, punctuality, appropriate dress, and safety.

___% **Planning/Organizing** Anticipates, establishes priorities, schedules activities and allocates resources to achieve results within set time frames and according to plan.

___% **Other** (Please specify, briefly) _____

100% Total Points

B02

A. Job Title

B. Gender (circle one)

Female Male

C. Age (circle one)

Under 20 20s
30s 40s 50s Over

“Other.” Participants were asked to distribute 100 points across the seven categories according to their perception of “...how they apply to your job - what it takes to do a good job.” The six importance categories were representative of categories often found on performance appraisal forms and are similar to those derived empirically in earlier research (cf., Waldman, Yammarino, & Avolio, 1990; Yammarino & Waldman, 1993). Each participant was asked to write in his/her job title on the rating instrument form as a way to check that only persons in the desired job classifications were included in the study. Demographic information on gender and age group was also collected.

Results

Descriptive statistics for the workers’ ratings of the six importance categories in each of the five job classifications are presented in Table 4 for Hospital A and in Table 5 for Hospital B (The “Other” category is not included in further analysis because it was used by less than fifteen percent of the respondents whereas each of the other six importance categories were used extensively; the least used was on 98.5% of the responses). Analysis by age group and gender disclosed no significant results.

Table 4**Hospital A - Mean responses of points assigned each skill category**

<u>Job Classifications: Skill Categories</u>	<u>F</u>	<u>H</u>	<u>R</u>	<u>A</u>	<u>N</u>
Job Knowledge	19.2	21.5	26.4	21.4	24.5
Quality	17.5	17.5	16.4	18.2	18.7
Teamwork	16.7	16.7	20.0	21.8	17.3
Communication/ Interpersonal	17.9	15.6	14.3	24.5	16.2
Policies and Procedure Adherence	15.4	15.4	12.1	8.6	9.5
Planning and Organizing	11.6	12.3	9.3	5.5	13.4
		<u>Subjects</u>			
F = Food workers		13			
H = Housekeepers		26			
R = Radiology Technicians		7			
A = Admissions		11			
N = Nurses		<u>26</u>			
	Total	83			

Table 5**Hospital B - Mean responses of points assigned each skill category**

<u>Job Classifications:</u> <u>Skill Categories</u>	<u>F</u>	<u>H</u>	<u>R</u>	<u>A</u>	<u>N</u>
Job Knowledge	14.2	25.6	35.0	28.6	26.0
Quality	29.3	22.4	23.0	18.9	16.1
Teamwork	23.9	15.5	15.0	14.4	17.9
Communication/ Interpersonal	10.2	10.3	13.0	15.6	14.0
Policies and Procedure Adherence	10.7	11.6	6.0	10.3	10.4
Planning and Organizing	9.7	12.1	8.0	9.3	15.5
		<u>Subjects</u>			
F = Food workers		10			
H = Housekeepers		18			
R = Radiology Technicians		5			
A = Admissions		16			
N = Nurses		<u>40</u>			
	Total	89			

For each of the six skill importance categories, t-tests were performed between each set of job classifications (nurses vs. foodworkers, nurses vs. admissions, nurses vs. housekeepers, etc.) to determine where differences existed between job classifications. In addition, t-tests were performed for each of the six categories in each of the five job classifications between the participant's responses in the two hospitals to determine institutional differences.

The results of the t-tests between each of the five job classifications for each of the six skill categories for Hospital A are summarized in Tables 6 through 11. The results for Hospital B are summarized in Tables 12 through 17. T-test results comparing importance ratings by incumbents in Hospital A with incumbents in Hospital B for each job classification on each skill category are summarized in Table 18.

Table 6 shows that radiology technicians and nurses in Hospital A rated Job Knowledge significantly higher than did food workers. In Table 9, admissions clerks in Hospital A rated Communications/Interpersonal Skills significantly higher than did any of the other job classifications. Table 10 shows that both food workers and housekeepers in Hospital A rated Policy and Procedure Adherence significantly more important than did either nurses or admissions clerks.

Table 6
Hospital A - Job Knowledge

	F	H	R	A	N
F	-				
H	-.93	-			
R	-2.46*	-1.46	-		
A	-1.00	.07	1.64	-	
N	-2.29*	-1.05	.60	-1.25	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 7
Hospital A - Quality

	F	H	R	A	N
F	-				
H	.02	-			
R	.30	.26	-		
A	-.38	-.28	-.46	-	
N	-.74	-.52	-.62	-.33	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 8**Hospital A - Teamwork**

	F	H	R	A	N
F	-				
H	- .02	-			
R	-1.06	-1.02	-		
A	-1.69	-1.63	- .45	-	
N	- .30	- .26	.84	1.42	-

F = Food workers

* .05

H = Housekeepers

** .01

R = Radiology Technicians

*** .001

A = Admissions

N = Nurses

Table 9**Hospital A - Communications/Interpersonal Skills**

	F	H	R	A	N
F	-				
H	1.01	-			
R	1.09	.37	-		
A	-2.30*	-2.94*	-2.65*	-	
N	.90	-.31	-.61	3.05*	-

F = Food workers

* .05

H = Housekeepers

** .01

R = Radiology Technicians

*** .001

A = Admissions

N = Nurses

Table 10
Hospital A - Policy and Procedures Adherence

	F	H	R	A	N
F	-				
H	.00	-			
R	1.28	1.24	-		
A	3.16**	3.02**	1.29	-	
N	3.85***	3.54***	1.15	- .50	-

F = Food workers

* .05

H = Housekeepers

** .01

R = Radiology Technicians

*** .001

A = Admissions

N = Nurses

Table 11
Hospital A - Planning and Organizing

	F	H	R	A	N
F	-				
H	- .37	-			
R	1.31	2.65*	-		
A	3.39**	5.69***	3.54**	-	
N	- .81	- .62	2.50**	-4.70***	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 11 shows that there were six different pairings of job classifications in Hospital A with significant differences for the Planning and Organizing importance category.

For Hospital B, Table 12 shows that Job Knowledge is significantly more important for housekeepers, radiology technicians, admissions clerks and for nurses than it is for food workers.

Table 13 shows that food workers, housekeepers and radiology technicians all rate Quality as significantly more important to their jobs than do nurses.

Food workers in Hospital B, as shown in Table 14, rated Teamwork significantly more important than do any of the other four job classification participants in the study.

For Hospital B, nurses, as shown in Table 15, rated Communications/Interpersonal Skills significantly higher than did either food workers or housekeepers.

Housekeepers in Hospital B, as shown in Table 16, as well as admissions clerks and nurses, rated Policy and Procedure Adherence significantly higher than did radiology technicians.

Table 17 shows that nurses in Hospital B rated Planning and Organizing significantly higher than did food workers, radiology technicians and admissions clerks, while housekeepers rated Planning and Organizing significantly higher than did radiology technicians.

Table 12
Hospital B - Job Knowledge

	F	H	R	A	N
F	-				
H	-3.09**	-			
R	-3.34**	-1.72	-		
A	-2.88**	-.75	.99	-	
N	-3.12**	-.17	1.63	.64	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 13
Hospital B - Quality

	F	H	R	A	N
F	-				
H	1.30	-			
R	1.14	- .19	-		
A	2.02	1.35	1.36	-	
N	2.63*	2.82**	2.51*	1.20	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 14
Hospital B - Teamwork

	F	H	R	A	N
F	-				
H	2.40*	-			
R	2.64*	.13	-		
A	3.66**	.32	.18	-	
N	2.56*	-.75	-.95	-1.62	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

Table 15
Hospital B - Communications/Interpersonal Skills

	F	H	R	A	N
F	-				
H	- .04	-			
R	-1.49	-1.70	-		
A	-1.69	-1.76	- .83	-	
N	-2.30*	-2.83**	- .68	.53	-

F = Food workers

* .05

H = Housekeepers

** .01

R = Radiology Technicians

*** .001

A = Admissions

N = Nurses

Table 16
Hospital B - Policy and Procedures Adherence

	F	H	R	A	N
F	-				
H	-0.33	-			
R	2.01	3.15**	-		
A	.19	.70	-2.74*	-	
N	.15	.73	-3.54**	-0.09	-

F = Food workers

* .05

H = Housekeepers

** .01

R = Radiology Technicians

*** .001

A = Admissions

N = Nurses

Table 17
Hospital B - Planning and Organizing

	F	H	R	A	N
F	-				
H	-1.46	-			
R	1.03	2.35*	-		
A	.26	1.73	-.82	-	
N	-3.51***	-1.87	-4.31***	-3.83***	-

F = Food workers

H = Housekeepers

R = Radiology Technicians

A = Admissions

N = Nurses

* .05

** .01

*** .001

To compare the results between the two hospitals for each of the five job classifications on the six skill importance categories an additional set of t-tests was conducted. Table 18 shows that both food workers and admissions clerks in Hospital A and Hospital B rated three of the six skill categories significantly different. Both job classifications showed differences on Teamwork and Communications/Interpersonal Skills while food workers also differed on Quality and admissions clerks differed on Planning and Organizing. Housekeepers showed significant differences on Communications/Interpersonal Skills and radiology technicians showed significant differences on Policies and Procedures Adherence. Nurses showed no significant differences between Hospital A and Hospital B.

Table 18

T-test results comparing Hospital A to Hospital B for each skill category and each job classification

<u>Job Classifications:</u> <u>Skill Categories</u>	<u>F</u>	<u>H</u>	<u>R</u>	<u>A</u>	<u>N</u>
Job Knowledge	-1.40	1.55	1.46	1.78	.58
Quality	2.33*	1.70	1.50	.22	-1.73
Teamwork	3.14**	- .39	-1.26	-2.26*	.30
Communication/ Interpersonal	-3.75***	-2.58*	- .40	-2.39*	-1.57
Policies and Procedure Adherence	-1.87	-1.84	-2.60*	.79	.82
Planning and Organizing	- .98	- .09	- .91	2.93**	1.10

Subjects

	<u>A</u>	<u>B</u>
F = Food workers	13	10
H = Housekeepers	26	18
R = Radiology Technicians	7	5
A = Admissions	11	16
N = Nurses	26	40

Discussion

Some trends can be identified among the several combinations of comparisons identified and presented as results.

Table 18 shows that Admission Clerks in Hospital A perceive Teamwork and Communication/Interpersonal Skills to be significantly more important than do the Admissions Clerks in Hospital B while those in Hospital B perceive Planning and Organizing to be significantly more important than do their counterparts in Hospital A. Why would this be? Could this, perhaps, reflect a difference in emphasis (even a difference of personality) of the managers of the respective units? The social context studies of Judge and Ferris (1993) would suggest it might. Or, could it be that the clerks in Hospital A just had some training classes on the importance of teamwork, communications and interpersonal skills? This view, or something similar, might be enhanced by looking at Table 9 to note that Admissions Clerks in Hospital A perceived Communications/Interpersonal skills to be significantly more important than any of the workers in the other five job classifications in Hospital A. Further, Table 11 shows that Admissions Clerks in Hospital A perceived Planning and Organizing as not only significantly less than did workers in each of the other job classifications, but, per Tables 4 and 5, they rated Planning and Organizing lower than any other skill category in any job classification in

either organization.

Food Workers in Hospital B (see Table 18) perceived both Quality and Teamwork to be much more important than did the Food Workers in Hospital A. Conversely, Food Workers in Hospital A perceived that communications and interpersonal skills were significantly more important than did the Food Workers in Hospital B. This finding is especially interesting since there is one manager who is responsible for both of these food operations even though they are located in two separate facilities across town from each other. There are, however, separate assistant managers assigned to the individual facilities who may present different leadership styles, for instance, in the two departments. This set of relationships appears to be worthy of further study.

For Hospital B, it is interesting to note the strong significance (.01 level) for Job Knowledge (Table 12) in all four other job classifications as compared to Food Workers (lower). Although the direction is similar in Hospital A (Table 6), the strength of the comparison is not near the level of that in Hospital B. Further interpretation of these results is difficult without the availability of measurements of additional variables to use for comparison. For instance, these results might relate to education level of the participants in the study or they might relate to some training efforts underway or other contextual variables.

Nurses in Hospital B perceived Planning and Organizing to be significantly more important than did the workers in the other job classifications (see Table 17) but also perceived Quality as less important than did the others (see Table 13). One explanation might be that Nurses assume quality will be consistently high. As such, they may not think of Quality as a category of distinguishing characteristics for Nurses. On the other hand, they may perceive that differences do exist among Nurses in the Planning and Organizing skill area. Nursing assignments may also vary on this variable, so it is a category that is important to measure.

Also in Hospital B, Teamwork was perceived to be of significantly higher importance by Food Workers than by workers in any other Job Classification (see Table 14). Food Workers frequently do work in teams or groups to meet their goals. Teamwork, therefore, would be a very important skill category for Food Workers. In future research, it might be useful, if possible, to collect some teamwork or groupwork contextual variables, such as group size, for comparison.

Radiation Technicians in Hospital B perceived Policy and Procedures Adherence to be significantly less important than did workers in the other job classifications (see Table 16). Radiation Technicians tend to work alone and are professionally trained to do so. Adherence to policies and procedures is what they do each day in each phase of their

operations. It is reasonable to assume that they would not expect this to be a distinguishing feature.

With differences as prevalent as these, should performance appraisal formats continue to use equally weighted skill categories in this context? Would greater satisfaction be achieved by using varied weights corresponding to the perceptions of the incumbents of these jobs?

Would the incumbents recognize their own relative weights if presented in a proposed weighted appraisal situation?

Study 2 examines these questions.

Study 2

Method

Subjects.

Participants were 30 nurses from Hospital B. Data were collected at two regularly scheduled staff meetings approximately six months after Study 1. The rating instruments for Study 2 were administered, received and reviewed by the researcher at each meeting to assure correct completion and appropriate category of each respondent. Participants were asked to write their job title on the rating instrument to assure that only nurses were included in the study.

Materials.

The Study 2 rating instrument (a copy of the Study 2 rating instrument is shown in Figure 3) consisted of three sets of weights for the six skill categories. Detailed descriptions of each of the six categories were printed on the back of the single page instrument (using the same wording as used for Study 1). Each participant was asked to rate each of the three sets of weights, labeled A, B, and C, along a seven point scale with respect to how they would feel if their job performance were to be evaluated on the six categories with the respective weights assigned.

Figure 3
Study 2 Rating Instrument - Page 1
Follow-up Questionnaire

The following six factors are frequently used for performance evaluations.

JK - Job Knowledge
QU - Quality
TW - Teamwork
CI - Communications/Interpersonal Skills
PP - Policies and Procedures Adherence
PO - Planning/Organizing

(Full definitions are printed on the back of the form)

Weighting these factors differently based on your opinions of their relative importance is one possible way of making performance ratings more fair. Listed below are three possible such weightings. I would like you to give me your impression of each of the three by marking an "x" on the line below for each, that you would be "very satisfied" to "not at all satisfied" if you were to be rated on each weighted scale:

<u>Scale A</u>	<u>Scale B</u>	<u>Scale C</u>
26 x JK	17 x JK	9 x JK
16 x QU	17 x QU	18 x QU
18 x TW	17 x TW	17 x TW
14 x CI	17 x CI	15 x CI
10 x PP	16 x PP	25 x PP
16 x PO	16 x PO	16 x PO

	Very Satisfied						Not at all Satisfied
Scale A	!_____!	!_____!	!_____!	!_____!	!_____!	!_____!	
Scale B	!_____!	!_____!	!_____!	!_____!	!_____!	!_____!	
Scale C	!_____!	!_____!	!_____!	!_____!	!_____!	!_____!	

Your Job Title _____

Figure 2 (Continued)**Study 2 Rating Instrument - Page 2**

Job Knowledge Demonstrates ability to apply knowledge to assignments and maintain expertise in his/her field by keeping current with new developments.

Quality Is thorough, accurate and carefully attends to detail in performing work; completes tasks according to established standards; participates in setting work standards and in collecting data to review actual quality outcomes.

Teamwork Achieves results with and through others; interacts with other individuals/departments to accomplish tasks.

Communications/Interpersonal Skills Expresses ideas, both oral and written, that are clear, concise, and accurate and interacts with others in a caring and effective manner.

Policies and Procedures Adherence Follows organizational and departmental policies and procedures, such as policies on attendance, punctuality, appropriate dress, and safety.

Planning/Organizing Anticipates, establishes priorities, schedules activities and allocates resources to achieve results within set time frames and according to plan.

The scale used was anchored by “Very satisfied” and “Not at all satisfied.” Half of the Study 2 rating instruments (interleaved every other one) reversed the order of the A, B, and C. No presentation order effect was found.

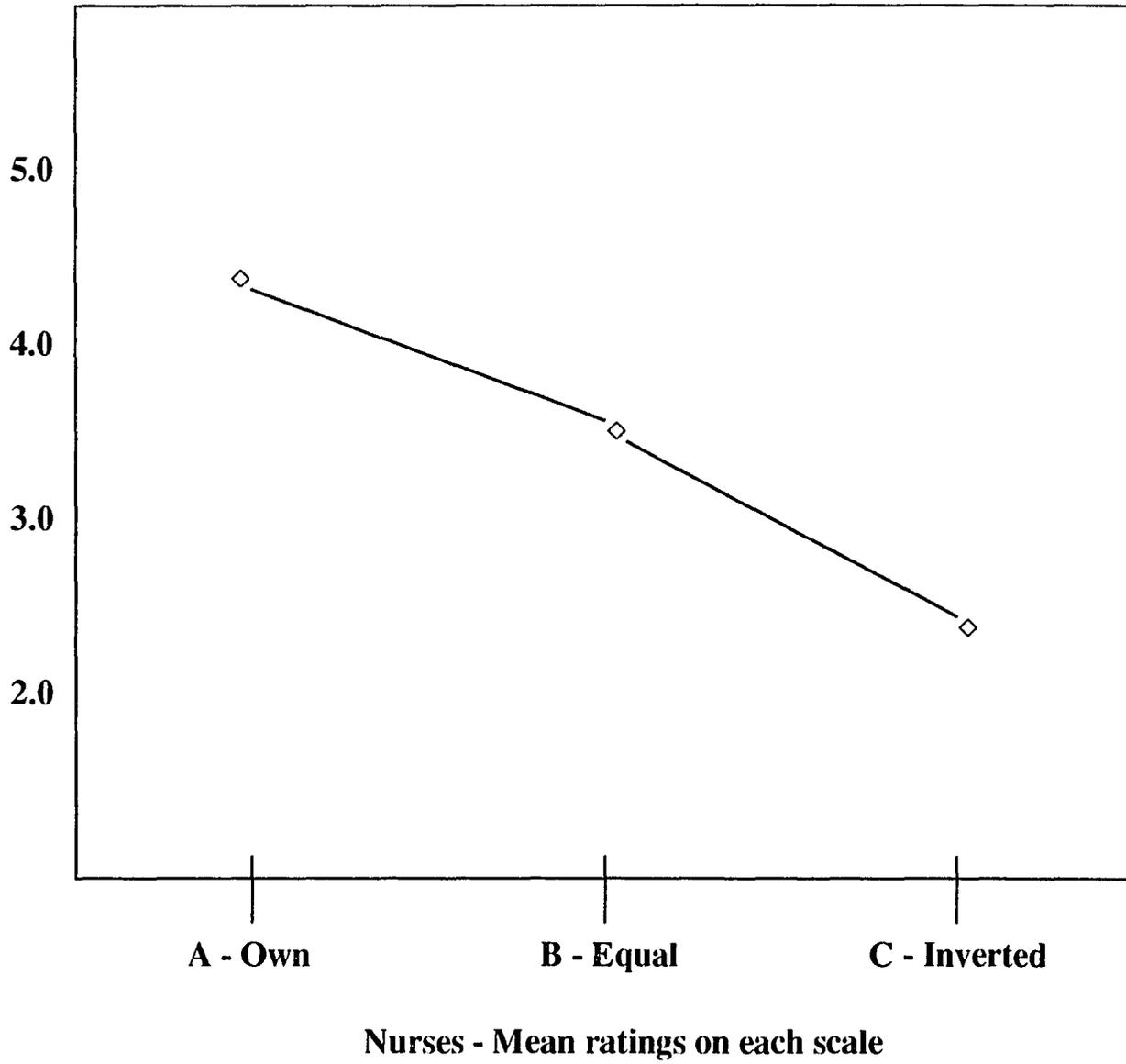
The three sets of scales (A, B, and C) represented, respectively, their own weights (from Study 1), an equal weighting, and an approximate inverse ranking (i.e., the Job Knowledge weight of 26 for scale A became 9 for scale C and Policy and Procedures weights of 10 for scale A became 25 on scale C, for the two extreme categories).

Results and discussion

Figure 4 summarizes the results of the survey. The participants indicated slightly higher satisfaction for A (their own weights) over B (equal weights) (4.4 to 3.6, $p = .083$). A compared to C (inverted weights) was significantly rejected (4.4 to 2.5, $p = .000$) as was B compared to C (3.6 to 2.5, $p = .007$).

Participants were able to select their own perceived weightings for the skill categories and indicated a significant preference for their own perceived weightings compared to either an equal weighting or the inverse weighting offered.

Figure 4
Satisfaction with Scales



Conclusions

The two studies reported here have shown that incumbents of a variety of jobs in two similar organizations in the same city reported significantly different perceptions of the same importance categories used frequently for the evaluation of work performance. These results support a contingency approach to performance appraisal taking into account contextual variables such as organization, location and job classification.

Results were based on a sample of the possible population of subjects, job classifications, and organizational settings. However, these results, obtained from the generally homogeneous sample population used, suggest that even more striking results might be found in a broader sample population.

These results provide evidence that consideration should be given to differentially weighting elements of performance appraisal formats in response to contextual effects related to different job classifications in different organizations and different locations. Performance appraisal formats are frequently equally weighted across job classifications and even across organizations. These studies bring this practice into question as a meaningful procedure. It appears from the results obtained here that

some importance ratings categories are more important than others in determining “what it takes to do a good job” from the perception of the job incumbent. If employee perceptions are to be considered in performance appraisal, then weighting the categories should be considered.

Discussion

These studies have shown that differences exist in the perceptions of workers across job classifications and across organizations with respect to skill importance ratings such as those used in performance appraisals. These findings are consistent with earlier work on skill importance ratings by Waldman, Yammarino, and Avolio (1990) and Yammarino and Waldman (1993) as they compared ratings of performance and ratings of importance for both supervisors and job incumbents. They called for additional tests of importance ratings in a variety of jobs and organizations. These studies have done that.

The two organizations included in these studies had similar job classifications. Perceptions of the job incumbents in comparable job classifications across these two organizations were shown to be different on the importance ratings tested. This raises two new questions: (1) Why were there differences between the organizations in similar job classifications? and (2) Would not differences between the incumbents of similar job classifications in unlike organizations be even more different? Consideration of the first question raises a weakness in the current studies that needs to be corrected in future research. Due to the nature of the

field settings where the data were collected, only age and gender were collected as demographic information. No significant differences were found based on these two factors. Other contextual data, such as educational level, years of service (both in job classification and in current position), time worked under current supervisor, span of control of supervisor, for example, should be collected to see how they might contribute to the differences observed and measured. Judge and Ferris (1993), for instance, suggest that social and situational elements, especially relational elements between the supervisor and subordinate, such as supervisor-subordinate demographic similarity, supervisor-subordinate work relationship, supervisor's affect toward subordinate, and supervisor's experience, may affect the rating process. A contextual factor included in the current studies was the recently developed customer contact construct (Kellogg, 1991; Kellogg & Chase, 1994). While the customer contact construct did appear to explain some portion of the differences observed, without other factor measurements available, no attribution can be affirmed. Continued investigation is required.

The second question raised calls for examination of each of these issues in additional organizations using the same job classification. For instance, nurses also work in nursing homes, private medical practices, and in other organizations. How would the perceptions of importance

ratings differ in these additional contexts? Only future research can determine the answer.

In addition, only five job classifications were included in the results of the current studies. Will the differences found extend to additional job classifications? Further research is needed to determine this.

For practitioners and decision makers in organizations, the implications of the results of this study are to suggest that systematic inaccuracies may be included in many, if not most, performance appraisal systems that were not previously suspected. Other recent research results have also suggested this concern. Giles and Mossholder (1990) studied the contribution of three contextual variables (complexity, implementation, and follow-up) which they found influenced the performance appraisal process. Judge and Ferris (1993) found, using a seven factor model of social influence in the performance evaluation process, that the social context does affect the performance-rating process. They called for continued study of the many social and situational influences on the performance appraisal process.

Practitioners, of course, are faced with the dilemma of the already high cost of performance appraisal systems using currently acceptable procedures compared with the relative importance of the results obtained. This is one reason many organizations do use performance appraisal

formats with common elements through the entire organization across many job classifications. This is done with full knowledge that the results obtained may be flawed despite prior research confirmation. The current research suggests one way that one flaw might be neutralized. By weighting the importance elements used by job classification, it appears that performance appraisal information would be both more accurate, and in the eyes of the workers, more fair.

References

- Austin, J.T., & Villanova, P. (1992). The criterion problem: 1917-1992. Journal of Applied Psychology, 77(6), 836-874.
- Bailey, C.T. (1983). The measurement of job performance. Aldershot, England: Gower Press.
- Bedeian, A.G. (1989). Totems and taboos: Undercurrents in the management discipline. The Academy of Management News, 19(4), 2-6.
- Bernardin, H.J. (1978). Effects of rater training on leniency and halo errors in student ratings of instruction. Journal of Applied Psychology, 63, 301-308.
- Bernardin, H.J., & Villanova, P. (1986). Performance appraisal. In E.A. Locke (Ed.), Generalizing from laboratory to field settings, 43-62. Lexington, MA: Lexington.
- Blanz, F., & Ghiselli, E.E. (1972). The mixed standard scale: A new rating system. Personnel Psychology, 25, 185-199.
- Borman, W.C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.
- Borman, W.C. (1986). Behavior-based rating scales. In R. A. Berk (Ed.), Performance assessment: Methods & applications (pp. 110-120). Baltimore: Johns Hopkins University Press.
- Borman, W.C., Hanson, M.A., Oppler, S.H., Pulakos, E.D., & White, L.A. (1993). Role of early supervisory experience in supervisor performance. Journal of Applied Psychology, 78(3), 443-449.

- Borman, W.C., White, L.A., Pulakos, E.D., & Oppler, S.H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76(6), 863-872.
- Bottger, P.C. (1984). Expertise and air time as bases for actual and perceived influence in problem-solving groups. Journal of Applied Psychology, 69, 214-221.
- Bretz, R.D., Jr., Milkovich, G.T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. Journal of Management, 18, 321-352.
- Cascio, W.F. (1992). Managing Human Resources: Productivity, Quality of Work Life, Profits, Third Edition. New York: McGraw-Hill, Inc.
- Cleveland, J.N., Murphy, K.R., & Williams, R.E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74(1), 130-135.
- Cleveland, J.N., & Shore, L.M. (1992). Self- and supervisory perspectives on age and work attitudes and performance. Journal of Applied Psychology, 77(4), 469-484.
- Dansereau, F., & Markham, S.E. (1987). Levels of analysis in personnel and human resources management. Research in Personnel and Human Resources Management, 5, 1-50.
- DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.
- DeVries, D.L., Morrison, A.M., Sullman, S.L., & Gerlach, M.L. (1986). Performance Appraisal on the Line. Greensboro, NC: Center for Creative Leadership.

- Dobbins, G.H., Cardy, R.L., & Platz-Vieno, S.J. (1990). A contingency approach to appraisal satisfaction: An initial investigation of the joint effects of organizational variables and appraisal characteristics. Journal of Management, 16, 619-632.
- Dowell, B.E., & Wexley, K.N. (1978). Development of a work behavior taxonomy for first-line supervisors. Journal of Applied Psychology, 63(5), 563-572.
- DuBois, C.L.Z., Sackett, P.R., Zedeck, S., & Fogli, L. (1993). Further exploration of typical and maximum performance criteria: Definitional issues, prediction, and white-black differences. Journal of Applied Psychology, 78(2), 205-211.
- Farh, J. & Dobbins, G.H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. Personnel Psychology, 42, 835-850.
- Feild, H.S., & Holley, W.H. (1982). The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases. Academy of Management Journal, 25, 392-406.
- Ford, J.K., & Noe, R.A. (1987). Self-assessed training needs: The effects of attitudes toward training, managerial level, and function. Personnel Psychology, 40, 39-53.
- Giles, W.F., & Mossholder, K.W. (1990). Employee reactions to contextual and session components of performance appraisal. Journal of Applied Psychology, 75(4), 371-377.
- Graves, L.M., & Karren, R.J. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. Personnel Psychology, 45, 313-340.

- Harris, C. (1988). A comparison of employee attitudes toward two performance appraisal systems. Public Personnel Management, 17(4), 443-456.
- Hartel, C.E.J. (1993). Rating format research revisited: Format effectiveness and acceptability depend on rate characteristics. Journal of Applied Psychology, 78(2), 212-217.
- Harvey, R.J. (1986). Quantitative approaches to job classification: A review and critique. Personnel Psychology, 39, 267-289.
- Hunter, J.E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In F.J.Landy, S. Zedeck, & J. Cleveland (Eds.), Performance measurement and theory (pp. 257-266). Hillsdale, NJ: Erlbaum.
- Ilggen, D.R., & Feldman, J.M. (1983). Performance appraisal: A process focus. Research in Organizational Behavior, 5, 141-197.
- Jordan, J.L., & Jordan, D.N. (1993). Satisfaction with performance appraisal ratings. Psychological Reports, 72, 1222.
- Judge, T.A., & Ferris, G.R. (1993). Social context of performance evaluation decisions. Academy of Management Journal, 36(1), 80-105.
- Kane, J.S., & Lawler, E.E., Jr. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B.M. Staw (Ed.), Research in organizational behavior (Vol. 1, pp. 425-478). Greenwich, CT: JAI Press.
- Katzell, R.A., & Austin, J.T. (1992). From then to now: The development of industrial-organizational psychology in the United States. Journal of Applied Psychology, 77(6), 803-835.

- Kellogg, D.L. (1991). Customer contact: Measurement and operations management implications. Dissertation, University of Southern California.
- Kellogg, D.L., & Chase, R.B. (1994). Constructing an empirically derived measure for customer contact. Forthcoming in Management Science.
- Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. Organizational Behavior and Human Decision Processes, 45, 194-208.
- Kozlowski, S.W., Jr., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. Journal of Applied Psychology, 72, 252-261.
- Kravitz, D.A., & Balzer, W.K. (1992). Context effects in performance appraisal: A methodological critique and empirical study. Journal of Applied Psychology, 77(1), 24-31.
- Landy, F.J., & Farr, J.L. (1980). Performance ratings. Psychological Bulletin, 87, 72-107.
- Latham, G.P., & Wexley, K.N. (1977). Behavioral observation scales. Personnel Psychology, 30, 255-268.
- Latham, G.P., Wexley, K.N., & Pursill, E. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- Lawler, E.E., III, Mohrman, A.M., Jr., & Resnick, S.M. (1984). Performance appraisal revisited. Organizational Dynamics, 13(1), 20-35.
- Martell, R.F., & Borg, M.R. (1993). A comparison of the behavioral rating accuracy of groups and individuals. Journal of Applied Psychology, 78(1), 43-50.

- Marx, M.H. (Ed.). (1963). Theories in contemporary psychology. New York: Macmillan.
- Maurer, T.J., Palmer, J.K., & Ashe, D.K. (1993). Diaries, checklists, evaluations, and contrast effects in measurement of behavior. Journal of Applied Psychology, 78(2), 226-231.
- McGehee, W., & Thayer, P.W. (1961). Training in business and industry. New York: Wiley.
- Meyer, H.H., Kay, E., & French, J.R.P., Jr. (1965). Split roles in performance appraisal. Harvard Business Review, 43, 123-129.
- Milkovich, G.T., & Newman, J.M. (1987). Compensation, Second Edition. Plano, TX: Business Publications, Inc.
- Miller, C.S., Kaspin, J.A., & Schuster, M.H. (1990). The impact of performance appraisal methods on age discrimination in employment act cases. Personnel Psychology, 43, 555-578.
- Mitchell, T.R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F.J. Landy, S. Zedeck, & J. Cleveland (Eds.), Performance Measurement and Theory (pp. 29-59). Hillsdale, NJ: Erlbaum.
- Moore, M.L., & Dutton, P. (1978). Training needs analysis: Review and critique. Academy of Management Review, 3, 532-545.
- Mount, M.K. (1983). Comparisons of managerial and employee satisfaction with a performance appraisal system. Personnel Psychology, 36, 99-110.
- Murphy, K.R., & Cleveland, J.N. (1991). Performance Appraisal: An Organizational Perspective. Boston: Allyn and Bacon.

- Murphy, K.R., Jako, R.A., & Anhalt, R.L. (1993). Nature and consequences of halo effect: A critical analysis. Journal of Applied Psychology, 78(2), 218-225.
- Nathan, B.R., & Tippins, N. (1990). The consequences of halo "error" in performance ratings: A field study of the moderating effect of halo on test validation results. Journal of Applied Psychology, 75(3), 290-296.
- Pavett, C.M., & Lau, A.W. (1983). Managerial work: The influence of hierarchical level and functional specialty. Academy of Management Journal, 26(1), 170-177.
- Pooyan, A., & Eberhardt, B.J. (1989). Correlates of performance appraisal satisfaction among supervisory and nonsupervisory employees. Journal of Business Research, 19, 215-226.
- Sackett, P.R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum performance. Journal of Applied Psychology, 73, 482-486.
- Schmitt, N., Noe, R.A., & Gottschalk, R. (1986). Using the lens model to magnify raters' consistency, matching, and shared bias. Academy of Management Journal, 29(1), 130-139.
- Smith, D.E. (1986). Training programs for performance appraisal: A review. Academy of Management Review, 11(1), 22-40.
- Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.
- Spiegel, W. R. (1962). Company practice in appraisal of managerial performance. Personnel, 39, 77.

- Steiner, D.D., Rain, J.S., & Smalley, M.M. (1993). Distributional ratings of performance: Further examination of a new rating format. Journal of Applied Psychology, 78(3), 438-442.
- Study of court decisions in cases involving employee performance appraisal systems. (1984, December 26). Daily Labor Report, 248, E-1-5.
- Thorndike, R.L. (1949). Personnel selection. New York: Wiley.
- Tornow, W.W., & Pinto, P.R. (1976). The development of a managerial job taxonomy: A system for describing, classifying, and evaluating executive positions. Journal of Applied Psychology, 61(4), 410-418.
- Waldman, D.A., Yammarino, F.J., & Avolio, B.J. (1990). A multiple level investigation of personnel ratings. Personnel Psychology, 43, 811-835.
- Wallace, M.J., Jr. (1983). Methodology, research practice, and progress in personnel and industrial relations. Academy of Management Review, 8(1), 6-13.
- Wexley, K.N., & Klimoski, R. (1984). Performance appraisal: An update. Research in Personnel and Human Resources Management, 2, 35-79
- Wherry, R.J., Sr. (1952). The control of bias in rating: A theory of rating (Personnel Research Board Report 922). Washington, D.C.: Department of the Army, Adjutant General's Office, Personnel Research Section.
- Williams, J.R., & Levy, P.E. (1992). The effects of perceived system knowledge on the agreement between self-ratings and supervisor ratings. Personnel Psychology, 45, 835-847.

- Woehr, D.J., & Feldman, J. (1993). Processing objective and question order effects on the causal relation between memory and judgment in performance appraisal: The tip of the iceberg. Journal of Applied Psychology, 78(2), 232-241.
- Yammarino, F.J., & Waldman, D.A. (1993). Performance in relation to job skill importance: A consideration of rater source. Journal of Applied Psychology, 78(2), 242-249.
- Yetton, P.W., & Bottger, P.C. (1982). Individual versus group problem solving: An empirical test of a best-member strategy. Organizational Behavior and Human Performance, 29, 3070321.
- Zalesny, M.D. (1990). Rater confidence and social influence in performance appraisals. Journal of Applied Psychology, 75(3), 274-289.
- Zammuto, R.F., London, M., & Rowland, K.M. (1982). Organization and rater differences in performance appraisals. Personnel Psychology, 35, 643-658.
- Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. Organizational Behavior and Human Performance, 18, 269-294.