

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.

U·M·I

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

Order Number 9304668

**Job knowledge, ratee familiarity, complexity of the rating
environment and the systematic distortion of performance
appraisals: A field experiment**

Gillen, Joseph John, Ph.D.

City University of New York, 1992

Copyright ©1992 by Gillen, Joseph John. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106



JOB KNOWLEDGE, RATEE FAMILIARITY, COMPLEXITY OF THE
RATING ENVIRONMENT AND THE SYSTEMATIC DISTORTION OF
PERFORMANCE APPRAISALS: A FIELD EXPERIMENT

by

JOSEPH JOHN GILLEN

A dissertation submitted to the Graduate Faculty in
Psychology in partial fulfillment of the requirements for
the degree of Doctor of Philosophy, The City University
of New York

1992

c 1992

JOSEPH JOHN GILLEN

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/16/92
Date

Roger E. Millsap
Chair of Examining Committee

9/16/92
Date

Richard D. Seay
Executive Officer

Roger E. Millsap, Ph.D.

Edwin P. Hollander, Ph.D.

Walter Reichman, Ed.D.
Supervisory Committee

Acknowledgments

The author thanks his advisory committee for their guidance throughout the duration of this research. I especially thank my committee chairman, Dr. Roger Millsap, for his continuing patience, support and encouragement; I owe him a debt of gratitude. I also thank my colleague, Karen Frost, for her continuing support and constructive criticism throughout the process.

This research would not have been possible without the permission and support of the following individuals at the Connecticut Department of Health Services: Commissioner Susan S. Addiss, Deputy Commissioner Yvette Melendez Thiesfield, Bureau Chief Stephen Harriman, Division Director Stanley K. Peck and Assistant Director Dr. Elise Gaulin-Kremer. I extend my gratitude to them all. Finally, I thank my wife and daughter, Catherine and Kelly, for their patience and support throughout the process.

Dedication

This research is dedicated to my parents: to the memory of my father, Joseph E. Gillen, and in honor of my mother, Edna Gillen.

TABLE OF CONTENTS

Title Page.....	i
Copyright Page.....	ii
Approval Page.....	iii
Acknowledgements.....	iv
Dedication.....	v
Table of Contents.....	vi
List of Tables.....	viii
List of Illustrations.....	ix
Overview.....	1
Halo Conceptualized.....	3
Measurement of Halo.....	5
Elimination and/or Control of Illusory Halo.....	7
Cognitive Processes.....	9
Contextual Factors.....	23
Summary.....	29
Hypotheses.....	32
Method.....	34
Subjects.....	34
Independent Variables.....	34
Dependent Variables.....	36
Procedure.....	37

Rater/Ratee Assignments.....	38
Rating Task and Environmental Complexity Manipulation.....	42
Results.....	44
Discussion	61
Limitations of the Present Study	67
Directions for Future Research	70
Appendix	75
References	88

LIST OF TABLES

1.	Means, Standard Deviations and 95% Confidence Intervals of Conceptual Similarity--Rating Covariation (CS-R).....	46
2.	Analysis of Variance Table Summarizing the Effects of Job Knowledge, Ratee Familiarity and Rating Environment Complexity on Conceptual Similarity--Rating Covariation (CS-R).....	48
3.	Means, Standard Deviations and 95% Confidence Intervals of Conceptual Similarity--Rating Covariation (CS-R).....	52
4.	Pearson Correlations between Halo Measure (SSD) and Conceptual Similarity--Rating Covariation (CS-R).....	54
5.	Pearson Correlations between Halo Measure (SSD) and Conceptual Similarity--Rating Covariation (CS-R).....	56
6.	Means, Standard Deviations and 95% Confidence Intervals of Halo Measure (SSD).....	57
7.	Means, Standard Deviations and 95% Confidence Intervals of Halo Measure (SSD).....	59

LIST OF ILLUSTRATIONS

1. Interaction of Job Knowledge with Ratee
Familiarity.....51

Overview

The notion of rating the performance of employees has occupied the interest of industrial and organizational psychologists for more than 60 years (Landy & Farr, 1980). Several researchers in this area of inquiry have suggested a broad-based socio-cognitive approach toward understanding the dynamic process of performance appraisal (DeCotiis & Petit, 1978; DeNisi, Cafferty & Meglino, 1984; Feldman, 1981; Landy & Farr, 1980). In addition to rater individual differences, this approach recognizes the importance of contextual/environmental influences on the rater, and their impact on the cognitive processes required in providing judgments on levels of performance. Variables such as the rater's knowledge of the job, familiarity with the ratee and the complexity of the rating environment can influence the cognitive strategies utilized by raters in arriving at evaluative judgments regarding levels of performance (Breckler, 1988; Kozlowski, Kirsch, & Chao, 1986). The objective of this study

was to examine the impact of some individual and contextual variables inherent in the performance appraisal process on raters' cognitive operations. Specifically, in light of recent findings in the performance appraisal research literature, this study focused upon the influence that levels of rater knowledge of the job, familiarity with the ratee and complexity of the rating environment have on rater reliance upon a cognitive strategy purported to be a source of halo error (Cooper, 1981a, 1981b; Kozlowski, Kirsch, & Chao, 1986; Kozlowski & Kirsch, 1987).

An inquiry into the reliability of a measure of job performance covariance beliefs [i.e., Conceptual Similarity (see Appendix B)] was also made. A discussion of the concepts of halo and halo error, as well as the variables of interest and their impact on halo error follows.

Halo Conceptualized

Halo has been referred to as the most pervasive response bias in the performance appraisal process (Cascio, 1987). Traditionally, halo has been conceptualized in two different ways. The general impression conceptualization views halo as being a result of the rater's reliance on an overall impression of the ratee (Borman, 1975; Landy, Vance, Barnes-Farrell & Steele, 1980). Under this conceptualization, the rater inappropriately provides ratings on individual performance dimensions as a function of an overall impression of the ratee. The salient feature conceptualization views the rater as being influenced by a particular feature of the ratee. Thus, the ratee's level of performance on a particular performance dimension inappropriately impacts on the rater's judgment with regard to the remaining performance dimensions. Cooper (1981b) places both conceptualizations under the rubric

"engulfing" (1981b, p.220). Either conceptualization leads to inflated interdimension correlations.

Another conceptualization of halo concerns the distinction between true and illusory halo (Bartlett, 1983; Cooper, 1981b). Essentially, this approach proposes that it is unrealistic to presume that job performance dimensions are independent. Thus, true halo is represented as valid interdimension correlation; illusory halo is said to be present in ratings when "observed halo exceeds true halo" (Cooper, 1981b, p.222). Bartlett (1983) expresses the rating response via the following equation:

$$z_R = W_T z_T + W_B z_B + W_I z_I + W_E z_E$$

The equation views the rating response (z_R) as being comprised of the true ability of the ratee ($W_T z_T$), bias of the rater ($W_B z_B$), environmental influences ($W_I z_I$) and random error ($W_E z_E$). Illusory halo is thus conceptualized as a portion of $W_B z_B$, when the observed exceeds the true interdimension

correlation. One of the aims of this study, however, was to examine the impact of environmental influences (W_{IZI}) on illusory halo.

Measurement of Halo. Several methods of operationalizing halo have been identified in the literature. Borman (1977), Bernardin & Fence (1980), Nathan & Lord (1983) and Nathan & Tippins (1990) utilized the across-dimension standard deviation (SD) averaged across ratees; variances have also been utilized in place of the SD (Cooper, 1981b; Schneier, 1977). Average interdimension correlations, sometimes transformed via Fisher's r -to- Z , have had extensive use as well (Fisicaro, 1988; Murphy & Jako, 1989). The presence of a general factor, following the examination of interdimensional factor structures, has been used as an indication of halo (Kraut, 1975), as has a Rater \times Ratee effect following a Rater \times Ratee \times Category analysis of variance (Cooper, 1981b; Pulakos, Schmitt, & Ostroff, 1986). Pulakos, Schmitt

& Ostroff, (1986) suggest the use of the average interdimension correlation as an appropriate operationalization when true scores are not available. Cooper (1981b) alluded that the standard deviation and correlational measures of halo were not equivalent; Pulakos, et al. (1986) offer compelling evidence that use of the standard deviation as an indication of halo is inappropriate. Specifically, the average standard deviation across rating dimensions is affected by the intercorrelations among rating dimensions and the differences among the means of rating dimensions utilized by raters. Thus, in certain cases, the standard deviation measure of halo is an inaccurate estimate of the level of rating dimension intercorrelations. Pulakos, et al. (1986) suggest that the average difference between observed and true dimension intercorrelations, the average observed dimension intercorrelation or the standardized standard deviation across dimensions are all appropriate measures of halo. However, the

"true" level of dimension intercorrelation is not usually available (as was the case in this study). Additionally, because, for this study, the intercorrelation matrix was based upon five dimensions across only five rates, the standardized standard deviation will be utilized as the operationalization of halo.

Elimination and/or Control of Illusory

Halo. Historically, there have been three lines of inquiry addressing the elimination or control of illusory halo: rater training, format design and statistical controls. Training raters to avoid errors [i.e., Rater Error Training, or RET (Bernardin, 1978; Borman, 1979; McIntyre, Smith & Hassett, 1984; Pulakos, 1984)], or to provide more accurate ratings [i.e., Rater Accuracy Training, or RAT (Athey & McIntyre, 1987; Bernardin & Buckley, 1981; Hauenstein & Foti, 1989; Thornton & Zorick, 1980)] has been moderately successful. The value of rater error training has, however, been questioned (Bernardin &

Pence, 1980; Zedeck & Cascio, 1982), as has the efficacy of training focused on providing accurate ratings (Hedge & Kavanagh, 1988).

Investigations of the effectiveness of different appraisal formats (Bernardin & Smith, 1981; Bernardin, Alvares & Cranny, 1976; King, Hunter & Schmidt, 1980; Sharon & Bartlett, 1969) have provided products of questionable utility and validity (Banks & Murphy, 1985; Murphy & Constans, 1987; Murphy, Martin & Garcia, 1982). Additionally, Landy & Farr (1980) suggest that format-related research has been so disappointing that a moratorium is in order.

Landy, Vance, Barnes-Farrell & Steele (1980) offered a method for the statistical control of illusory halo. Specifically, when individual and overall performance ratings are provided, the overall rating is partialled out of each individual performance item. This procedure results in enhanced interpretability of factor analyses, improved discriminant validity of individual

performance items and reduced halo. Several researchers have, however, questioned the appropriateness of this methodology with a primary emphasis on the implication by Landy, et al. (1980) that individual performance items are orthogonal to an overall rating (Harvey, 1982; Hulin, 1982; Murphy, 1982). Evaluative statements regarding overall performance are generally assumed to reflect, in part, dimension-specific performance (i.e., dimension-specific evaluations are non-orthogonal to an overall evaluation). By partialling out the overall from dimension-specific evaluations (having non-zero correlations with the overall evaluation), true variance must be removed (Murphy, 1982).

Cognitive Processes

Landy & Farr's (1980) review of the performance appraisal literature concluded that an increased understanding of the rating process would most likely be achieved by attending to cognitive characteristics

of the rater. Recent performance appraisal research has, indeed, taken a cognitive processing approach (Krzystofiak, Cardy, & Newman, 1988). Terms such as "cognitive style", "cognitive schema" and "cognitive categorization" are now apparent in the performance appraisal literature (Cardy & Kehoe, 1984; Lord, 1985; Mount & Thompson, 1987; Nathan & Lord, 1983).

A promising line of research addressing raters' cognitive processes focuses on the notion that, under certain conditions, halo error may, in part, be a function of raters' preexisting conceptualizations of the intercorrelations between performance dimensions without regard to the performance of the ratee (Cooper, 1981a, 1981b; Kozlowski, Kirsch, & Chao, 1986; Kozlowski & Kirsch, 1987). Shweder & D'Andrade (1980) reported on a series of studies in which memory-based clinical and/or personality rating structures were reproduced from independent similarity of meaning judgments. Specifically, judgments by an independent sample of individuals,

regarding the interdimension correlations of various rating instruments, were predictive of the intercorrelations of ratings provided by actual raters utilizing those instruments.

Shweder & D'Andrade (1980) proposed that memory-based ratings provided by the actual raters were systematically distorted as a function of their reliance upon the "representative heuristic" (Tversky & Kahneman, 1974). Tversky & Kahneman (1974) state that when judgments are made under difficult memory conditions, individuals often rely on beliefs regarding "the degree to which A is representative of B", or "the degree to which A resembles B" (1974, p. 1124). Thus, if A and B are conceptualized as being resemblant, they may be, erroneously, conceptualized as being covariant.

Drawing on Shweder & D'Andrade's (1980) systematic distortion hypothesis, Cooper (1981a) investigated the notion of raters' implicit covariance theories as being a source of illusory

halo in job performance ratings. Cooper (1981a) proposed that when raters were required to provide performance ratings under difficult memory conditions (e.g., annually), they would tend to rely on the "representative heuristic" (Tversky & Kahneman, 1974). Reliance on the heuristic would, it was hypothesized, result in the systematic distortion of ratings as a function of rater beliefs with regard to how performance appraisal dimensions "go together" or resemble each other. Cooper (1981a) asked graduate students to provide conceptual similarity judgments of rating scale dimensions from four published studies. Specifically, for each performance appraisal instrument, students provided conceptual similarity judgments for each unique pair of rating scale dimensions. Conceptual similarity was operationalized as "how likely someone who is good at dimension X is also good at dimension Y" (1981a, p. 304). Conceptual similarity measures were correlated with interdimension correlations

of the performance appraisal instruments utilized in each of the studies addressed. Results from three of the four intermatrix correlations were significant: conceptual similarity indices were predictive of interdimensional matrices derived from rater-provided ratings. Results of the fourth study, which provided normative "true score" measures (Borman, 1979), did not support the systematic distortion hypothesis: "true score" interdimension correlations were more predictive of rater-provided indices than were conceptual similarity measures. However, subjects in the fourth study provided performance ratings immediately after observing videotaped performances. Thus, the difficult memory condition requirement was not fulfilled.

Murphy & Balzar (1986) investigated the relationship between the systematic distortion of memory-based ratings and rating accuracy. Subjects viewed videotapes of "on-the-job" performance and provided either immediate or delayed performance

ratings. Consistent with the systematic distortion hypothesis (Shweder & D'Andrade, 1980), ratings provided in the delayed condition exhibited significantly greater interdimension correlation than did ratings provided in the immediate condition.

The relationship between job knowledge, rater familiarity, conceptual similarity and halo error was investigated by Kozlowski, Kirsch & Chao (1986). It was hypothesized that conceptual similarity covariation would be more strongly associated with rating covariation, and rating covariation less strongly associated with true score covariation, when raters lacked relevant job knowledge and rater familiarity. Kozlowski, Kirsch & Chao (1986) noted operational problems related to Borman's (1977) methods of presenting videotaped performances as rating stimuli and generating "true scores". Specifically, it was noted that, because of their brevity (i.e., 5 - 7 minutes) and lack of distractions, performance videotapes are unrealistic

representations of the rating process. Additionally, because all raters are provided with the same performance information, the method is not appropriate when job knowledge and ratee familiarity are variables of interest. Borman's (1977) method of pooling "expert" judgments to arrive at "true" performance scores is, Kozlowski, Kirsch & Chao (1986) noted, potentially faulty because expert judgments could be contaminated via judges' implicit covariance beliefs or some other cognitive schema. Accordingly, Kozlowski, et al. (1986) utilized objective performance indices of major league baseball players as true scores. As hypothesized, job knowledge and ratee familiarity were significantly related to a rater's tendency to rely on implicit covariance beliefs. That is, when raters lacked knowledge of the job domain or of ratee performance, conceptual similarity--rating covariance was greater than rating--true score covariance. Additionally, there was a significant negative

relationship between illusory halo and levels of job knowledge and ratee familiarity.

Noting that the systematic distortion phenomenon references the individual rater, Kozlowski & Kirsch (1987) extended the research of Kozlowski, et al. (1986) utilizing an individual, as opposed to a group-level, analysis. Raters' conceptual similarity beliefs with regard to rating instrument dimensions were compared to true scores and provided ratings. The findings were consistent with previous conceptual similarity research. That is, when rater knowledge of the job or familiarity with the ratee declined, conceptual similarity--rating covariance was greater than true score--rating covariance. A significant negative relationship between illusory halo and job knowledge or ratee familiarity was also evident.

Kozlowski & Kirsch (1987) noted a "boundary condition", inconsistent with the systematic distortion hypothesis, with regard to job knowledge and ratee familiarity. Specifically, when low job

knowledge raters rated unfamiliar ratees, their ratings were unrelated to conceptual similarity and true score indices. Kozlowski & Kirsch (1987) proposed that conceptual similarity is a relational schema dealing with distortions of dimensional covariance, as opposed to levels of rating. When a rater possesses limited information with regard to a ratee's performance, but enough to make a category assignment relative to level of performance, the conceptual similarity schema would provide a guide, with regard to the remaining performance dimensions, on which the rater lacks relevant information. Conversely, in the absence of information relative to the job and ratee, the conceptual similarity schema would not be a sufficient guide with regard to ratings; ratings provided under this condition would be primarily comprised of either random or central tendency error. This position is distinct from the original "strong" systematic distortion hypothesis proposed by Shweder (1975), and disconfirmed by

subsequent research (Block, Weiss, & Thorne, 1979; Weiss & Mendelsohn, 1986). It is, however, consistent with Cooper's (1981b) "weak" hypothesis which views the phenomenon as dealing with illusory halo, as opposed to the existence or validity of enduring personality structures. Additionally, although the Kozlowski & Kirsch (1987) "boundary condition" position was advanced post-hoc, it is consistent with the data, and addresses Cooper's concerns regarding clarification of the operation of the process. As such, this modified systematic distortion hypothesis deserves empirical investigation.

Jacobs & Kozlowski (1985) found a similar operation with regard to halo. Students rated their professors at one, five and ten weeks into the semester. Measures of halo increased as a function of familiarity with the ratee; the most significant increase occurred between weeks one and five, and slight increases were noted between weeks five and

ten. Although, on the surface, this result is in conflict with the prevailing position that increased opportunity to observe ratee performance leads to decreases in halo (Brown, 1968; Cummings & Schwab, 1973; Koltuv, 1962), a close examination of the study reveals results consistent with the systematic distortion hypothesis. Specifically, ratings of professors were provided by students; the rating dimensions were such that the raters would presumably (and by design) have no knowledge of performance at week one (e.g., Depth of Knowledge, Grading, Relevance), and limited knowledge at weeks five and ten. Thus, as raters would have some indication of levels of performance on some of the dimensions at weeks five and ten (i.e., Delivery, Interpersonal Relations), their conceptual similarity beliefs would serve as a guide with regard to the performance dimensions of which they had little or no performance-related knowledge.

Smither & Reilly (1989) investigated the relationship between raters' knowledge of the job and the reliability of measures of their conceptual similarity beliefs relative to the systematic distortion of performance ratings. Four studies were conducted. Three studies found the reliability of conceptual similarity beliefs to be positively related to level of job knowledge. The fourth study found changes in conceptual similarity beliefs over a nine-week interval.

Several unanswered questions relative to conceptual similarity and the systematic distortion of performance ratings were delineated by Smither & Reilly (1989). First, the boundary condition effect noted by Kozlowski & Kirsch (1987) may have been underestimated as a function of the unreliability of conceptual similarity judgments. That is, if the reliability of measures of conceptual similarity beliefs is low for low job knowledge raters, the difference between conceptual similarity--rating

covariance for low and high job knowledge raters would most likely increase following correction for unreliability. Second, it is possible that the unreliability of conceptual similarity is a function of temporal instability, and that conceptual similarity judgments taken at different times would covary with performance ratings taken at the same time intervals. Finally, Smither & Reilly (1989) noted, providing conceptual similarity judgments may stabilize an initially unstable conceptual similarity structure. Thus, for raters categorized as being low in job knowledge, changes in conceptual similarity ratings (from Time 1 to Time 2) may occur as a function of a cognitive shift with regard to the meaning of the dimension-pair relationships (Golembiewski, Billingsley, & Yeager, 1976).

Murphy & Jako (1989) tested the hypothesis that observed halo, as well as the tendency to over- or underpredict "true" halo, was a function of the level of true halo and conceptual similarity beliefs.

Conceptual similarity beliefs were found to be unrelated to rating intercorrelations; true halo level accounted for significant variance in rating intercorrelation (i.e., 30%). It should be noted, however, that the rating dimensions utilized in the study delineated specific behaviors, as opposed to the more global trait dimensions addressed in typical systematic distortion studies. Additionally, the college student subjects in this study were, presumably, knowledgeable with regard to the target position of college instructor, and ratings were provided immediately after viewing videotaped performances. Thus, as manipulations of level of job knowledge did not occur, and memory decay was not addressed, the systematic distortion hypothesis was not adequately tested. Further, Murphy & Jako (1989) admit that the effect of "true" halo must be interpreted with caution, as the two levels chosen (i.e., .07 & .76) represent extreme, and most probably unrealistic, values.

Contextual Factors

Landy & Farr's (1980) process model of performance rating presents a systemic view of the task of providing performance ratings. Explicit in this model are rater and ratee characteristics, as well as influences upon the rater from factors such as the development and format of the rating instrument, purpose of the rating, and contextual/environmental variables. Kane & Lawler (1979) similarly note that appraisal effectiveness is affected by administrative characteristics of the appraisal system, rater and ratee individual differences and organizational conditions surrounding the appraisal system. Well recognized in the literature addressing such factors are the rater's knowledge of the requirements of the job being performed and familiarity with the ratee's performance. Clearly, raters require a knowledge of job requirements and ratee performance in order to provide adequate evaluations of performance (DeCotiis & Petit, 1978; DeNisi, Cafferty & Meglino, 1984; Ilgen

& Feldman, 1983; Kozlowski, Kirsch & Chao, 1986). Although contextual factors have been cited as being critical to performance and performance-related variables (Graves, 1985; Joyce, Slocum & von Glinow, 1982; McGrath & Rotchford, 1983; Rothstein, 1986; Steel & Mento, 1986), as well as the process of providing performance evaluations (Banks & Murphy, 1985; Ilgen & Feldman, 1983; Kane & Lawler, 1979; Zammuto, London, & Rowland, 1982), they have been largely ignored in the performance appraisal research literature (Landy & Farr, 1980; Mitchell, 1983). Indeed, in discussing the recent cognitive orientation of performance appraisal research, Banks & Murphy (1985) note that investigations of rater cognitive processes have tended to ignore the context within which performance appraisals are conducted; they suggest that performance appraisal research take into account constraining features of the organization in addition to rater cognitive processes. Such an approach, they suggest, mirrors the realities of a

functioning organization.

Ilgen & Feldman (1983) state that, in order to understand the performance appraisal process, "...one must begin with [a] realistic conception of the evaluator as an information gatherer and processor, operating in a complex environment (p. 143)." This statement recognizes that the cognitive operations required of raters in providing performance evaluations are affected by a number of contextual/environmental influences. Specifically, it recognizes that, in addition to the task of providing evaluations of subordinate performance, managers operate in an environment characterized by multiple task demands and pressure of time. Ilgen and Feldman (1983) further state that

"Rarely does the appraiser judge the performance of a single individual on a single occasion. Usually he or she is faced with working and judging several individuals over extended periods of time (p. 147)."

Thus, in addition to task and time demands unrelated to the performance appraisal process, the number of

appraisees rated at one time is a significant factor in the performance appraisal process.

The requirement of providing multiple performance appraisals does not alter the complexity of the rating task per se. Rather, the rating task environment becomes more complex. Wood (1986) defines task complexity as a linear combination of the number of discrete acts performed (component complexity), the sequencing of inputs and type of relationship between informational cues, performer acts and products (coordinative complexity), and environmental variations affecting the relationship between inputs and products (dynamic complexity). The task of providing performance ratings is thus delineated as including informational cues provided by the performer/ratee, the acts of observing, encoding, storing and retrieving performance-related information, as well as providing evaluative judgments, and the performance appraisal as an end product. As such, the performance appraisal process

is comprised of information, act and product components. In the task of providing multiple evaluations, the fundamental information, act and product components remain the same as if one were providing a single evaluation. The multi-observational nature of providing several evaluations does, however, affect the rating environment. Specifically, because the rater must attend to and observe multiple ratees, the dynamic complexity component of Wood's (1986) linear equation is affected; the rating environment is more complex. This is consistent with the notion of cognitive simplification strategies discussed by Feldman (1981), and Lord (1985). In terms of signal detection theory (SDT, see Lord, 1985 for a discussion of STD), the task of providing multiple appraisals increases absolute levels of signal and noise. Under such conditions, it is hypothesized that raters would employ a cognitive economy strategy, such as reliance upon their conceptual similarity

beliefs, in an effort to simplify the rating process. The choice of environmental complexity as a variable of interest in the context of providing performance appraisals is somewhat unique. Historically, studies investigating cognitive factors in performance appraisal have focused upon complexity as an internal individual difference variable. For example, Schneier (1977) examined rater cognitive complexity, and type of rating format, as factors affecting rating accuracy and halo. Schneier's operationalization of cognitive complexity was based upon the work of Bieri (1955), and Kelly (1955). Cronshaw and Lord (1987) investigated leadership perceptions as a function of cognitive categorization (i.e., stimulus prototypicality). Mount and Thompson (1987) investigated rater cognitive categorization, operationalized as the perceived role congruency of the target ratees. As previously indicated, this study is unique in that it will investigate the impact of environmental complexity on the cognitive

process utilized by raters. To date, the impact of environmental complexity in performance appraisal has been addressed by one other researcher (Breckler, 1988); that study utilized Wood's (1986) definition of environmental complexity.

Summary

Based upon the work of Shweder (1975) and Shweder & D'Andrade (1979, 1980), Cooper (1981a,b) and others (Kozlowski et al., 1986; Kozlowski & Kirsch, 1987) have pursued investigations of the systematic distortion of performance appraisal ratings as a function of raters' beliefs regarding the covariance of performance appraisal dimensions without reference to actual ratee performance. The original "strong" hypothesis of Shweder (1975), proposing that the systematic distortion of (personality) ratings is merely, and totally, a function of raters' preconceived notions regarding the semantic similarity of scale dimensions has

evolved to a "weaker" hypothesis. This hypothesis states that raters' beliefs regarding the conceptual similarity of (performance appraisal) rating scale dimensions contribute to the systematic distortion of ratings when raters possess a minimal amount of job-relevant knowledge or information. Thus, according to the hypothesis, when a rater's job-relevant information is fragmentary, his/her conceptual similarity beliefs (cognitive schema), with regard to how the performance appraisal dimensions "go together", function as an inappropriate guide with regard to ratings provided for the categories of which he/she possesses incomplete job-relevant knowledge. In opposition to Shweder's (1975) "strong" hypothesis, the "weaker" hypothesis states that a rater's conceptual similarity schema will not impact upon provided ratings when job-relevant knowledge is high, or totally absent. Specifically, when the rater possess adequate job and performance-related

knowledge, ratings will be relatively accurate; when the rater has no knowledge of the job requirements or ratee performance, ratings will be subject to random, or possibly central tendency, error.

In addition to raters' cognitive operations, the performance appraisal process is impacted upon by a number of individual and contextual/environmental variables (Brecker, 1988; Landy & Farr, 1980). One such variable, conspicuous by its relative absence in the research literature (Banks & Murphy, 1985; Brecker, 1988; Ilgen & Feldman, 1983), is complexity of the rating environment. The task of providing a performance appraisal consists of information, act and product components (Wood, 1986). When a rater is required to provide multiple appraisals, the basic information, act and product components remain unchanged. However, because the rater must attend to and observe multiple ratees, the dynamic complexity of the task increases; the task environment becomes more complex. Thus, the multi-observational nature of

providing several evaluations increases the cognitive demands placed upon the rater. Presumably, the rater then relies upon his/her conceptual similarity beliefs as an efficiency strategy. The current study incorporated this variable within the line of research addressing the systematic distortion hypothesis. Thus, the research extends existing laboratory research within a previously-unaddressed field experiment methodology, as well as incorporates a heretofore unaddressed environmental/contextual variable, the operationalization of which closely mirrors the realities of a functioning organization.

Based on the previous review of theory and empirical research, the following hypotheses were advanced:

H1: The CS-R relationship for raters providing ratings in high complexity rating environments will be greater than the CS-R relationship for raters providing ratings in low complexity rating environments.

- H2: The CS-R relationship for raters low in ratee familiarity will be greater than the CS-R relationship for raters high in ratee familiarity.
- H3: The CS-R relationship for raters who are low in ratee familiarity and high in job knowledge will be greater than the CS-R relationship for raters who are low in ratee familiarity and low in job knowledge.
- H4: The CS-R relationship for raters who are high in ratee familiarity and low in job knowledge will be greater than the CS-R relationship for raters who are high in ratee familiarity and high in job knowledge.
- H5: There will be a significant positive relationship between halo error and the CS-R measure. That is, as rater CS-R measures increase, the corresponding across-ratee average standardized standard deviations will decrease.

An inquiry into the reliability of the Conceptual Similarity measure (CS) was also made.

Method

Subjects

All employees within three units of a New England state government agency were the potential subject pool for this study (N=327). Pursuant to management directives, the researcher was blind to the identities of the majority of study participants (NOTE: confidentiality was partially broken in order to recruit specific personnel needed as study participants; confidentiality of responses was maintained following initial recruitment). This was accomplished with the assistance of a graduate student at the University of Connecticut (UConn). The graduate student assigned a unique code number to each study participant, and coordinated all data collection activities (see Appendix A).

Independent Variables

The three independent variables were Job Knowledge, Ratee Familiarity and Rating Environment Complexity. Job Knowledge and Ratee Familiarity were

operationalized as participant responses to the Job Knowledge/Ratee Familiarity Questionnaire (see Appendix C). Specifically, the Job Knowledge portion of the questionnaire asked participants to indicate, on a seven-point scale, their level of knowledge of the requirements of specific job classifications within one unit of the agency (additional participants from other agency units were involved as raters only, and were not included on the Job Knowledge/Ratee Familiarity questionnaire). The scale anchors were 1 = Not Very Knowledgable, 4 = Somewhat Knowledgable and 7 = Extremely Knowledgable. The Ratee Familiarity questionnaire asked participants to indicate, on a seven-point scale, how familiar they were with regard to the work performance of specific individuals performing within the aforementioned job classifications. The scale anchors were 1 = Not Very Familiar, 4 = Somewhat Familiar and 7 = Extremely Familiar. Rating Environment Complexity was operationalized by the

number of performance appraisals a participant was required to provide. Raters providing five performance appraisals were categorized as being in a low-complexity rating environment; raters providing ten performance appraisals were categorized as being in a high-complexity rating environment.

Dependent Variables.

The dependent variables were conceptual similarity-rating covariation (CS-R) and halo. CS-R was operationalized as individual-level CS-R covariation. Specifically, one half of each rater's off diagonal intercorrelation matrix of performance appraisal ratings (R) was arrayed as a vector relating to his/her corresponding CS judgments. All (R) correlations were transformed to z scores via Fisher's r-to-z (Guilford & Fruchter, 1978; Hays, 1973). As the CS-R measure was based upon less than thirty items, individual-level CS-R

correlations were obtained via a Spearman rank correlation (Guilford & Fruchter, 1978).

Halo was operationalized as a rater's average standardized standard deviation across performance appraisal dimensions (Pulakos, Schmitt, & Ostroff, 1986). Specifically, each rater's ratings were standardized within-dimension and across-ratees, such that the mean was 0.00; the standard deviation 1.00. An across-dimension standard deviation was then calculated for each rater; an across-ratees average indicated each rater's level of halo.

Procedure

The Conceptual Similarity Questionnaire (see Appendix B) was mailed, by the graduate student, to all employees who initially agreed to participate in the research. The Conceptual Similarity Questionnaire was designed to elicit one's implicit beliefs with regard to the covariance of performance appraisal format dimensions, without reference to

actual performance. The instructions and scale were adapted from Kozlowski, et al. (1986). The Conceptual Similarity Questionnaire was also administered following the rating task. This afforded an opportunity to investigate the reliability of the measure.

Following completion of the Conceptual Similarity Questionnaire, and prior to the rating task, the Job Knowledge/Ratee Familiarity Questionnaire was mailed to all individuals who initially agreed to participate in the research (See Appendix "C ").

Rater/Ratee Assignments. The Job Knowledge/Ratee Familiarity Questionnaire was analyzed; the Job Knowledge/Ratee Familiarity scales were trichotomized as follows: scale points 1-2 = low; scale points 3-5 = moderate; scale points 6-7 = high. The moderate category was dropped out; raters were categorized as high/low on the job knowledge and

ratee familiarity factors relative to each of the ratees. It was planned that rater/ratee assignments would be made such that all raters would rate the same five "core ratees"; those ratees would be representative of the category the rater was assigned to. For example, raters who were assigned to the low knowledge/low familiarity category would rate the same five ratees; the raters, as a function of their responses to the job knowledge/ratee familiarity questionnaire, would possess little or no knowledge of the requirements of the ratees' job duties, and little or no knowledge of their job performance. Prior to the environmental complexity manipulation, the categories within the study would have been: low knowledge/low familiarity; low knowledge/high familiarity; high knowledge/low familiarity; high knowledge/high familiarity. However, analysis of the Job Knowledge/Ratee Familiarity questionnaire did not produce the desired results, in that five ratees common to all categories could not be identified.

Specifically, the categories high knowledge/low familiarity, and low knowledge/high familiarity were not well represented. In order to address this problem, several strategies were utilized. First, classification of knowledge and familiarity based upon responses to the Job Knowledge/Ratee Familiarity questionnaire was changed such that scale points 1-3 = low, scale point 4 = moderate, and scale points 5-7 = high. Next, individuals in other units of the agency who had previously held positions in the target unit, or held the same positions in their units, were identified and recruited as raters who would most likely fulfill the high knowledge/low familiarity condition. All individuals in a second agency unit were asked to participate in the study. As this second unit had limited functional interaction with the target unit, it was hoped that respondents, who would be classified in one of the two problematic categories, would be recruited. Next, those individuals from the target unit who initially

declined participation in the research were contacted a second time, in an effort to increase the probability of identifying raters in the needed categories. Finally, when performing a second analysis of the Job Knowledge/Ratee Familiarity questionnaire, the category low knowledge/low familiarity was initially ignored, as raters for this category, if needed, would be easily recruited from within the agency.

The Conceptual Similarity and Job Knowledge/Ratee Familiarity questionnaires were hand-delivered to the additional participants. Moderate levels of the knowledge and familiarity factors were dropped out of analysis of the expanded data-set. The strategies were successful in that, following analysis of the expanded Job Knowledge/Ratee Familiarity data-set, five ratees common to all categories were identified. One hundred three acceptable research participants were recruited. Gender, ethnicity and age demographics

were as follows: 27% male; 73% female; 81% white; 12% black; 7% hispanic. The mean age of the group was 36.

Rating Task and Environmental Complexity

Manipulation. Raters received their rating assignments via the mail. Raters were instructed to provide their best judgment with regard to ratee performance on each appraisal dimension for the preceding twelve month period. All raters were advised that the rating task was for research purposes only (see Appendix "D").

Half of the raters in each of the aforementioned categories were randomly assigned an additional five ratees to rate; the additional ratees were representative of the rater's assigned category; their positions on the rating list were randomly assigned to avoid order effects. Ratings of the additional ratees were not, however, utilized in data analysis relative to halo or CS-R covariation.

Assignment of five additional ratees represented the rating environment complexity manipulation. Raters were categorized as low or high on the environmental complexity variable as a function of their assignment to rate five or ten ratees respectively.

Results

The design of the study included two levels each of job knowledge, ratee familiarity and rating environment complexity. In order to ensure that raters' implicit covariance beliefs with regard to performance appraisal dimensions were equivalent across the job knowledge and ratee familiarity conditions, an ANOVA was performed on the CS ratings. Each rater's CS ratings were summed; the sums of each condition were averaged to arrive at the CS variable for each cell. None of the F ratios reached significance: for the job knowledge condition, $F(1,92) = .001, p > .05$; for the ratee familiarity condition, $F(1,92) = .34, p > .05$; for the interaction term, $F(1,92) = .04, p > .05$.

In order to arrive at the CS-R dependent variable, individual-level intercorrelation matrices were computed for the performance appraisal ratings provided by the raters. When more than one dimension lacked variance across ratees, dimensional intercorrelations could not be computed; data

relative to those raters was excluded from the study ($n = 5$). In order to arrive at a balanced design, two participants (one from the low/low category; one from the high/high category) were randomly selected and dropped from the study. As such, the design included ninety-six participants over eight cells.

One half of each rater's intercorrelation of ratings, excluding diagonals, was then arrayed as a vector corresponding to his/her respective CS judgments. All correlation values were transformed to z scores via Fisher's r -to- z (Hays, 1973). CS ratings and z values were then correlated via Spearman's rank correlation (Guilford & Fruchter, 1978). Table 1, on page 46, displays mean CS-R values, standard deviations and 95% confidence intervals for the eight cells of the design.

The CS-R covariation relationships were submitted to a $2 \times 2 \times 2$ analysis of variance (ANOVA). The results were significant for job knowledge, $F(1,88) = 5.20, p < .05$; ratee familiarity,

Table 1

Means, Standard Deviations and 95% Confidence Intervals
of
Conceptual Similarity--Rating Covariation (CS-R)

Environmental Complexity						
Low			High			
Ratee Familiarity			Ratee Familiarity			
Job Knowledge	Low	High	Low	High		
Low	(a)	.366	.682	.424	.724	.549
	(b)	.292	.142	.215	.139	
	(c)	.172-.560	.588-.776	.281-.567	.632-.816	
High	(a)	.458	.432	.483	.408	.445
	(b)	.229	.225	.181	.234	
	(c)	.306-.610	.283-.581	.363-.603	.253-.563	
		.412	.557	.454	.566	
		.485		.510		

Note: (a) = mean
(b) = standard deviation
(c) = 95% confidence interval

$F(1,88) = 8.00, p < .01$; and the job knowledge-by-ratee familiarity interaction term, $F(1,88) = 15.40, p < .01$. The main effect for rating environment complexity did not reach significance, $F(1,88) = .40, p > .05$. None of the interaction terms which included rating environment complexity reached significance: for the interaction of job knowledge and rating environment complexity, $F(1,88) = .20, p > .05$; for the interaction of ratee familiarity and rating environment complexity, $F(1,88) = .00, p > .05$; for the three-way interaction of job knowledge, ratee familiarity and rating environment complexity, $F(1,88) = .20, p > .05$. Table 2, on page 48, displays a summary of the analysis of variance results.

Hypothesis 1 stated that the CS-R relationship for raters providing ratings in high-complexity rating environments would be greater than the CS-R relationship for raters providing ratings in low complexity rating environments. This hypothesis was not supported; the environmental complexity

Table 2

Analysis of Variance Table Summarizing the Effects
of
Job Knowledge, Ratee Familiarity and Rating Environment Complexity
on
Conceptual Similarity--Rating Covariation (CS-R)

Source	SS	df	MS	F	p	ω^2
Job knowledge (JK)	.26	1	.26	5.20	<.05	.04
Ratee familiarity (RF)	.40	1	.40	8.00	<.01	.06
Rating Environment Complexity (EC)	.02	1	.02	.40	ns	
JK X RF	.77	1	.77	15.40	<.01	.13
JK X EC	.01	1	.01	.20	ns	
RF X EC	.00	1	.00	.00	ns	
JK X RF X EC	.01	1	.01	.20	ns	
Error	3.99	88	.05			
Total	5.46	95				

manipulation had no significant effect, $F(1,88) = .40$, $p > .05$.

Hypothesis 2 stated that the CS-R relationship for raters low in ratee familiarity would be greater than the CS-R relationship for raters high in ratee familiarity. Although the main effect for ratee familiarity was significant [$F(1,88) = 8.00$, $p < .01$], the result was contrary to the hypothesis. That is, the CS-R relationship for raters in the high familiarity condition was greater than the CS-R relationship for raters in the low familiarity condition. This relationship is the opposite of that predicted.

Hypotheses 3 and 4 were evaluated via a priori orthogonal t ratios (Kirk, 1968). Hypothesis 3 stated that the CS-R relationship for raters who were low in ratee familiarity and high in job knowledge would be greater than the CS-R relationship for raters who were low in ratee familiarity and low in

job knowledge. The hypothesis was not supported:

$t(2,88) = 1.169, p > .05.$

Hypothesis 4 stated that the CS-R relationship for raters who were high in ratee familiarity and low in job knowledge would be greater than the CS-R relationship for raters who were high in ratee familiarity and high in job knowledge. The hypothesis was supported: $t(2,88) = 4.354, p < .0005.$ Figure 1, on page 51, illustrates the interaction of job knowledge and ratee familiarity.

As the rating environment complexity manipulation had no significant effect on the dependent variable, the means, standard deviations and confidence intervals for the CS-R relationship displayed in Table 2 were recomputed after collapsing the rating environment complexity variable. The recomputed values are displayed in Table 3 on page 52.

Figure 1

Interaction of Job Knowledge with Ratee Familiarity

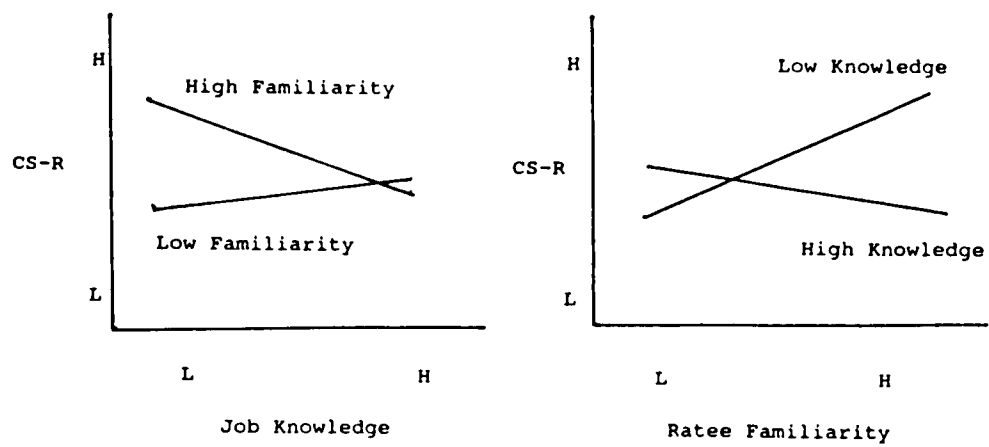


Table 3

Means, Standard Deviations and 95% Confidence Intervals
of
Conceptual Similarity--Rating Covariation (CS-R)

Job Knowledge	Ratee Familiarity		
	Low	High	
Low (a)	.395	.703	.549
(b)	.252	.139	
(c)	.279-.511	.639-.767	
High (a)	.470	.420	.445
(b)	.202	.225	
(c)	.377-.563	.357-.523	
	.433	.562	

Note: (a) = mean
(b) = standard deviation
(c) = 95% confidence interval

Hypothesis 5 stated that there would be a significant positive relationship between halo and the CS-R measure. That is, levels of halo would increase as raters tended to rely upon their implicit covariance beliefs with regard to the rating instrument dimensions. It should be noted that because the standardized standard deviation measure of halo (SSD) is inversely related to halo magnitude, support of the hypothesis would be evidenced by a negative correlation. A Pearson correlation was performed in order to determine the relationship between levels of halo (SSD) and the CS-R measure. The result, although in the hypothesized direction, was not significant, $r = -.05$, $p > .05$. A Pearson correlation between SSD and CS-R was performed by cell; none of the results was significant. The correlation for the cell low knowledge/high familiarity/high complexity did, however, approach significance. Table 4, on page 54, provides an illustration of the data.

Table 4

**Pearson Correlations between Halo Measure (SSD)
and
Conceptual Similarity--Rating Covariation (CS-R)**

Environmental Complexity					
Low			High		
Ratee Familiarity			Ratee Familiarity		
Job Knowledge	Low	High	Low	High	High
Low	-.10	.13	.32	-.45	-.03
High	.01	.25	-.33	-.12	-.05
	.08	.19	-.01	-.29	
		.07		-.15	

Note: Because the standardized standard deviation measure of halo (SSD) is inversely related to halo magnitude, a negative sign indicates a positive relationship. None of the above correlations reached significance.

A Pearson correlation between SSD and CS-R was performed by cell after collapsing the environmental complexity condition; none of the results was significant. Interestingly, however, the values in the cells representing fragmentary information (i.e., low knowledge/high familiarity & high knowledge/low familiarity) are clearly in the anticipated direction vis-à-vis the values in the cells representing full information and little information, respectively. This agrees with the notions that raters possessing fragmentary information relative to the ratee tend to rely upon their conceptual similarity beliefs, and that halo increases as a function of reliance upon one's conceptual similarity beliefs. Table 5, on page 56, provides an illustration of the data.

The average level of halo (SSD) across groups was .62 (sd = .18; 95% CI = .58-.66). Table 6, on page 57, illustrates the cell means, standard deviations and 95% confidence intervals for the SSD measure.

Table 5

Pearson Correlations between Halo Measure (SSD)
and
Conceptual Similarity--Rating Covariation (CS-R)

Job Knowledge	Ratee Familiarity		
	Low	High	
Low	.07	-.23	-.08
High	-.15	.09	-.03
	-.04	-.07	

Note: Because the standardized standard deviation measure of halo (SSD) is inversely related to halo magnitude, a negative sign indicates a positive relationship. None of the above correlations reached significance.

Table 6

**Means, Standard Deviations and 95% Confidence Intervals
of
Halo Measure (SSD)
Environmental Complexity**

		Low		High		
		Ratee Familiarity		Ratee Familiarity		
Job Knowledge		Low	High	Low	High	
Low	(a)	.65	.64	.67	.57	.63
	(b)	.21	.12	.18	.17	
	(c)	.51-.79	.56-.72	.55-.79	.46-.68	
High	(a)	.65	.62	.59	.60	.62
	(b)	.16	.24	.17	.21	
	(c)	.54-.76	.46-.78	.48-.70	.46-.74	
		.65	.63	.63	.59	
		.64		.61		

Note: (a) = mean
 (b) = standard deviation
 (c) = 95% confidence interval

Table 7, on page 59, illustrates the cell means, standard deviations and 95% confidence intervals for the SSD measure after collapsing the environmental complexity condition.

As previously indicated, an inquiry into the reliability of the Conceptual Similarity measure (CS) was made. Responses to the measure (see Appendix B) taken at Time 1 and Time 2 were subjected to a Pearson correlation. The result was significant, $r = .71$, $p < .01$.

Smither & Reilly (1989) examined the reliability of conceptual similarity schemata as a function of levels of job knowledge. In three separate studies, they found the reliability of CS judgments to be positively related to relevant job knowledge. The present research examined the difference between consistency of conceptual similarity schemata as a function of levels of job knowledge. Mean reliabilities for the high and low job knowledge conditions were .81 and .62 respectively. Consistent

Table 7

**Means, Standard Deviations and 95% Confidence Intervals
of
Halo Measure (SSD)**

Job Knowledge		Ratee Familiarity		
		Low	High	
Low	(a)	.66	.61	.64
	(b)	.19	.15	
	(c)	.58-.74	.55-.67	
High	(a)	.62	.61	.62
	(b)	.17	.22	
	(c)	.55-.69	.52-.70	
		.64	.61	

Note: (a) = mean
 (b) = standard deviation
 (c) = 95% confidence interval

with previous findings, the difference was found to be significant ($z = 1.914$, $p < .05$) with a one-tailed test. As such, the present research adds further support to the notion that level of job knowledge is positively related to consistency of conceptual similarity schemata.

Discussion

Shweder (1975) and Shweder & D'Andrade (1979, 1980) suggested that ratings of personality measures were wholly a function of raters' cognitive schemata relative to the covariance of rating scale dimensions. Specifically, they posited a "strong" hypothesis that personality ratings were merely a reflection of raters' beliefs regarding the cultural and semantic similarity of rating scale dimensions. Cooper (1981a,b) and Kozlowski (Kozlowski, Kirsch, & Chao, 1986; Kozlowski & Kirsch, 1987) integrated the notion of conceptual similarity (CS) within the area of performance ratings. Their refined, and "weaker", systematic distortion hypothesis (SDH) proposed that, under certain conditions, raters would rely upon their beliefs regarding the covariance of appraisal dimensions; their ratings would, to a certain extent, mirror such beliefs. The results of the present research provide additional support for that notion.

Bernardin and Villanova (1986) note that results found in laboratory studies may not be generalizable to real-world environments, due to critical differences between the research and criterion settings. Thus, because this research was performed within a functioning organization, the results add further support and credibility to a line of research heretofore performed exclusively via laboratory experiments.

The pattern of results achieved in this research mirror those of previous research (i.e., Kozlowski and Kirsch, 1987) in direction and relative magnitude. The effects of job-relevant knowledge, familiarity with task performance, and their interaction, were significant factors relative to raters' reliance on their conceptual similarity schemata. Absolute CS-R values in the present research were, for the most part, however, nearly double those previously reported. This may be attributable to differences between laboratory and real-world settings. However,

it is noted that the performance appraisal dimensions utilized in the previous research were specific and objective (i.e., Kozlowski & Kirsch [1987] utilized objective measures such as number of home runs, strike outs and stolen bases). The systematic distortion hypothesis (SDH) posits that the CS-R relationship will increase when performance appraisal dimensions are less specific or trait-based, as was the case in this research. Thus, SDH research supports the "boundary condition" proposition relative to job knowledge and rater familiarity, as well as the notion that specificity of appraisal dimensions aids in error reduction (Bernardin & Walter, 1977; Borman, 1979; Dunnette & Borman, 1979; Wexley & Yukl, 1984). Further, it enhances our understanding of the processes involved in error reduction. Specifically, in addition to guiding the rater to focus on well defined, and dimension-relevant, behaviors, it would appear that specificity of dimensions reduces the probability of

error due to rater reliance upon cognitive schemata relative to dimensional covariation.

A key variable in the research being reported was complexity of the rating environment. Consistent with the comments of several researchers (i.e., Banks and Murphy, 1985; Ilgen and Feldman, 1983; Kane and Lawler, 1979), it was hypothesized that as complexity of the rating environment increased, the CS-R relationship would increase. The complexity manipulation was, however, not significant. Because the research was conducted in the field, however, control of pressure-of-time was not feasible. That is, the researcher was not able to control the time parameters within which the performance appraisals were to be completed. Further, as organizational leadership demanded that all research activities be performed outside of the work setting, effect of the manipulation was most likely further attenuated, as the noise, pressures and multiple task demands inherent to the work setting were absent (Brecker,

1988). Thus, the effort to manipulate complexity of the rating environment by merely increasing the number of performance appraisals one was required to provide may have been inadequate. Alternatively, it may be that there exists a specific boundary of rating environment complexity, with regard to the number of appraisals provided, that has yet to be defined. That is, the use of five versus ten performance appraisals may have been an inadequate operationalization of rating environment complexity.

It was hypothesized that as reliance upon one's implicit covariance beliefs increased, level of halo would increase. As indicated earlier, although the relationship was in the hypothesized direction, it did not reach significance. Two points deserve elaboration with regard to this result. First, previous research (i.e., Kozlowski, Kirsch, & Chao, 1986; Kozlowski & Kirsch, 1987) operationalized ratee familiarity as how familiar the raters were with the target ratees (i.e., how aware they were of the

ratees by name alone). The current research, however, asked respondents to indicate how familiar they were with the job performance (or product of that performance) of the target ratees. Thus, differences in operationalizing ratee familiarity may have affected the results. Second, and more importantly, differences between laboratory and field methodology may have affected levels of halo. Specifically, although raters in the current study were categorized with regard to the job knowledge and ratee familiarity factors, the existing social system (Muchinsky, 1987; Trist, 1981) within which raters and ratees operated had the potential to impact upon levels of classical halo (i.e., general impression and salient feature halo error). That is, even though a rater may have lacked job knowledge or familiarity vis-à-vis a specific ratee, opportunities for direct social interaction, or third-party reports regarding the ratee, existed. Such social system interactions certainly have the potential to influence a rater's

judgment regarding ratee performance. However, within a laboratory experiment, it is doubtful that socially-induced influence would exert as significant, if any, an effect.

Limitations of the Present Study. The strengths and weaknesses of laboratory and field research have been well documented (Campbell & Stanley, 1963; Cook & Campbell, 1979; Kerlinger, 1986). Because prior research relative to the topic utilized laboratory methodology exclusively, that the present research relied upon a field-experimental design is positive in terms of integration of results. However, the present research was subject to the weaknesses of less-controlled paradigms. Specifically, because participants were required to respond to instrumentation outside of the work setting, much of the realism sought was lost. Such was the case with regard to the environmental complexity variable. Two important factors related to the work setting,

however, (i.e., job knowledge and rater familiarity) remained.

Campbell & Stanley (1963) and Cook & Campbell (1979) offer caveats with regard to history occurring between measures of the same construct taken at time 1 and time 2. Although such a threat was not significant in the present research, historical factors relative to all participants deserve mention within the context of generalizability. The present research was performed amidst a fiscal crisis affecting the organization of which the participants were members. During the course of the research, legislative mandates resulted in a significant number of personnel layoffs; threats of additional layoffs were rife. Such a climate should certainly be taken into account when an attempt to generalize is made. Thus, although it is generally recognized that appraisals for research purposes are more valid than for administrative purposes (Bass & Turner, 1973; Boehm, 1977; Zedeck & Cascio, 1982), the climate

within which the research was conducted may have affected participants' willingness to provide accurate ratings. This may have affected the results with regard to the relationship between halo and reliance on conceptual similarity schemata. Specifically, although raters in the categories low knowledge/low familiarity and high knowledge/high familiarity demonstrated less reliance upon their conceptual similarity schemata (see Tables 1 and 3 on pages 51 & 52), their levels of halo were not significantly different from the remaining groups (Tables 6 and 7, on pages 57 & 59 illustrate that levels of halo were relatively stable across cells). Perhaps the threatening climate caused the raters to provide ratings contaminated by halo in an effort to protect their colleagues. In addition to the general climate, the researcher's position within the organization may have impacted upon the results. During data collection, the researcher, unhappily, became aware of an effort by one individual to

persuade others not to participate in the research, or to sabotage the research, only because the researcher was a manager.

As previously mentioned, social interactions between participants had the potential to affect the relationship between levels of halo and conceptual similarity. Previous laboratory studies found these measures to be positively related; the present research revealed no significant relationship (see Tables 4 and 5 on pages 54 & 56). Thus, within the context of investigating such a relationship, interaction between participants may be considered a negative, and potentially contaminating factor.

Directions for Future Research. The present research is notable in that it supports, via a field methodology, the results of previous laboratory research. However, a number of critical questions remain unanswered. It is suggested that future research investigate the relationship of complexity

of the rating environment and reliance upon one's conceptual similarity schemata. The present research operationalized rating environment complexity as the number of appraisals one was required to provide. The use of five versus ten appraisals as an operationalization did not aid in achieving significance. Perhaps there is a critical boundary, relative to number of appraisals provided, that has yet to be discovered. Thus, future research should examine the effect of a range of numbers of appraisals performed, and its effect on the CS-R relationship. Additionally, as the present research lacked realism in terms of organizational pressures, it is suggested that future field research focused upon this topic be conducted such that appraisals are provided within the work setting. Alternately, laboratory research could be designed to simulate the pressures of a real-world work setting. Brecker (1988) successfully simulated complexity of the rating environment within a laboratory setting.

Performance appraisal researchers are in general agreement that as appraisal dimensions become more specific, rating accuracy increases, and rating error decreases (Muchinsky, 1987). The systematic distortion hypothesis (SDH) posits an inverse relationship between rating dimension specificity and reliance upon one's conceptual similarity schemata. Again, additional research is indicated relative to reliance upon conceptual similarity schemata and levels of specificity of appraisal dimensions. Perhaps such a line of research could identify a critical point, with regard to dimension specificity, at which reliance upon conceptual similarity schemata is activated.

Smither & Reilly (1989) proposed several questions with regard to the reliability of measures of conceptual similarity beliefs. The present research examined general consistency, as well as consistency as a function of level of job knowledge. However, the notion of schemata reconceptualization

has yet to be addressed. Smither & Reilly (1989) suggested that differences in conceptual schemata, from Time 1 to Time 2, may reflect a reconceptualization. That is, having considered dimension relationships at Time 1, raters may reconceptualize dimension relationships as a function of having been exposed to them. Future research utilizing measures at Times 1, 2 and 3 may be able to speak to this issue. Additionally, it has been suggested that conceptual similarity reliability is temporally related, and that CS measures at a specific point in time will correlate well with contiguous appraisals. Thus, future research should examine the CS-R relationship at several points in time.

Finally, research with regard to training for appraisal - specifically rater error training (RET) - has yet to address the operation of conceptual similarity schemata. It would seem, given what has been revealed with regard to the job knowledge/ratee

familiarity--conceptual similarity relationship, that error training relative to the systematic distortion hypothesis could be tailored to "boundary condition" raters (i.e., raters in the low knowledge/high familiarity or high knowledge/low familiarity categories).

APPENDIX A

Dear Colleague:

As you may know, I am near completion of the requirements for the Ph.D. degree; the last step in that process is an original research project. The study I will be conducting concerns how, and under what conditions, people make certain decisions when providing performance appraisals. Participation in the study is voluntary; I hope you will agree to assist me in this important endeavor. Please note that this is not a Department project. However, the Department and your union representatives have been made aware of the details of the study; they have not objected to it.

If you agree to participate, you will be asked to respond to several questionnaires. Additionally, you will either be asked to rate the performance of several employees of the Division, or you will be one of the individuals rated. Please be aware that the identity of all study participants will be completely confidential. A graduate student at UConn will collect the data and assign code numbers to all study participants. You will receive your materials in the mail, and will return them to the graduate student in a pre-addressed, stamped envelope. I WILL NEVER BE AWARE OF YOUR IDENTITY. NOBODY IN THE DEPARTMENT OR DIVISION WILL BE AWARE OF YOUR IDENTITY. When the study is completed, a summary of the research will be made available for your review. The summary will not identify any individual.

I hope you will agree to participate in this study. It is very important that I have as many people participate as possible. Should you have questions or concerns regarding this project, please feel free to contact me. I will make every effort to respond to all your concerns.

If you agree to participate in the study, please sign your name in the space provided below. Additionally, please provide the address where you wish to receive the study materials. Please do not use your work address. Time spent on the study must be non-work time (the total time should be less than two hours). If you agree to participate in the study, return this letter in the pre-addressed, stamped envelope. If you do not wish to participate, discard this letter.

Thank you,

Joseph J. Gillen, Chief
Applications, Examinations & Licensure

PRINT YOUR NAME HERE: _____

PRINT THE ADDRESS WHERE YOU WILL BE RECEIVING THE
STUDY MATERIALS HERE:

APPENDIX B

CONCEPTUAL SIMILARITY QUESTIONNAIRE

The Administrative-Clerical Performance Appraisal used by this agency contains five performance dimensions: Knowledge of Work, Quantity of Work, Quality of Work, Ability to Deal With People and Attendance. This questionnaire is designed to measure how you see these dimensions as being "similar" or "resembling one another".

On the opposite side of this page, you will find a listing of all the possible pairs of performance dimensions. You will be asked to provide ratings of how you feel they are similar or dissimilar. For example, if you feel Quantity and Quality are similar or "go_together", then you would expect that someone who is rated high on Quantity would also be be rated high on Quality. If you feel that Knowledge and Quantity are not similar, or "don't go together", then you would expect that someone who is high on Knowledge would not necessarily be high on Quantity.

Use the scale on the top of the opposite side of this page as a guide in making your judgments. **YOUR RESPONSES WILL BE STRICTLY CONFIDENTIAL, AND WILL BE UTILIZED FOR RESEARCH PURPOSES ONLY.** Please note that you have been assigned a code number. The code number has been assigned by an individual not employed by the Department. I will only be aware of your code number - I will not be able to identify you by name.

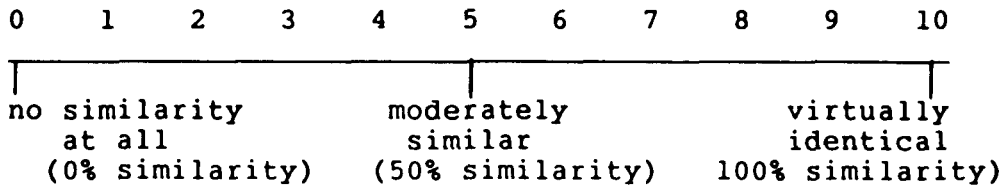
Following completion of the research, a summary will be available upon request.

Please return the completed questionnaire in the enclosed pre-addressed, stamped envelope.:

CODE # _____

SIMILARITY RATING SCALE

Use the following scale to make your judgments. Each of the numerical points can be regarded as a percentage of similarity. For example, if the first dimension in the pair is identical to the second, then a rating of 10 indicates 100% similarity. Likewise, if the dimension pairs are moderately similar, then a rating of 5 would indicate 50% similarity, and so forth.



WHEN RATING THE PAIRS, THINK OF PEOPLE IN GENERAL, NOT A SPECIFIC PERSON. Please consider all points on the scale. Place the number that represents your rating on the line to the right of the dimension pair.

<u>DIMENSION PAIRS</u>	<u>RATINGS</u>
Knowledge of Work - Quantity of Work	_____
Knowledge of Work - Quality of Work	_____
Knowledge of Work - Attendance	_____
Knowledge of Work - Ability to Deal With People	_____
Quantity of Work - Quality of Work	_____
Quantity of Work - Attendance	_____
Quantity of Work - Ability to Deal With People	_____
Quality of Work - Attendance	_____
Quality of Work - Ability to Deal With People	_____
Attendance - Ability to Deal With People	_____

CODE # _____

APPENDIX CJOB KNOWLEDGE/RATEE FAMILIARITY QUESTIONNAIRE

In order to provide a Performance Appraisal, one must possess an understanding of the job being performed by the employee, as well as the employee's performance on the job. Individuals providing Performance Appraisals differ in the extent to which they have knowledge of the job being performed and/or the performance of the employee performing that job. For example, it is unlikely that a new manager/supervisor would have a complete understanding of what it is that the employee actually does. Additionally, knowledge of the employee's performance is dependent upon the opportunity to observe the actual performance and/or the product of that performance. Managers/supervisors differ with regard to opportunity to observe employee performance and/or the end product of that performance. Further, if the employee is relatively new, the manager/supervisor may not have had adequate opportunity to observe his/her performance prior to providing a Performance Appraisal.

This questionnaire is designed to measure how much you know about specific job classifications within the Division, and how much of an opportunity you have to observe the actual job performance and/or the end product of job performance of specific persons. Please answer the questions honestly, and to the best of your ability. Use the scale provided with each question. **YOUR RESPONSES WILL BE STRICTLY CONFIDENTIAL AND WILL BE UTILIZED FOR RESEARCH PURPOSES ONLY.** Individual responses will not be shared with the Division or Department. Please return the completed questionnaire in the enclosed pre-addressed, stamped envelope.

**** go on to the next page ****

CODE # _____

JOB KNOWLEDGE

Using the scale on the bottom of this page, please indicate, on the line to the right of each job title, your level of knowledge of the jobs listed on the next page. That is, indicate how much knowledge you have regarding what a person in that title actually does. You will note that not all of the scale points have a written description. However, if you feel your level of knowledge falls at a particular point between two scale points that have a description, you should feel free to use that scale point. For example, if you feel that your level of knowledge for a particular job title is less than, but close to, "Extremely Knowledgeable", you would probably choose scale point "6".

Please consider all points on the scale, and provide your best response. **YOUR RESPONSES WILL BE COMPLETELY CONFIDENTIAL AND WILL BE UTILIZED FOR RESEARCH PURPOSES ONLY.**

Rating Scale

-
- 1 - Not Very Knowledgable
 - 2 -
 - 3 -
 - 4 - Somewhat Knowledgable
 - 5 -
 - 6 -
 - 7 - Extremely Knowledgable

JOB KNOWLEDGE RATINGS

<u>Job Titles</u>	<u>Ratings</u>
Administrative Assistant, MQA Administration...	_____
Administrative Hearing Attorney II, MQA.....	_____
Paralegal Specialist, MQA Administration.....	_____
Health Program Associate, MQA Administration...	_____
Clerk/Clerk Typist, MQA Administration.....	_____
Secretary I, MQA Administration.....	_____
Administrative Attorney, Hearings Office.....	_____
Head Clerk, Hearings Office.....	_____
Senior Clerk, Hearings Office.....	_____
Clerk/Clerk Typist, Hearings Office.....	_____
Secretary I, Hearings Office.....	_____
Paralegal Specialist, Hearings Office.....	_____
Investigator, Hearings Office.....	_____
Section Chief, APEX.....	_____
Licensing Examination Assistant, APEX.....	_____
MQA Investigator, APEX.....	_____
Senior Clerk, APEX.....	_____
Clerk/Clerk Typist, APEX.....	_____
Secretary I, APEX.....	_____
Secretary II, L&R.....	_____

JOB KNOWLEDGE RATINGS
(continued)

<u>Job Titles</u>	<u>Ratings</u>
Head Clerk, L&R.....	_____
Senior Clerk, L&R.....	_____
Clerk/Clerk Typist, L&R.....	_____
Fiscal Administrative Asst., L&R.....	_____
Administrative Asst., Nursing Board.....	_____

RATEE FAMILIARITY

The attached pages list personnel assigned to the Division. Using the scale provided on the bottom of this page, please indicate, on the line to the right of the individual's name, your degree of familiarity with his/her work performance on his/her current job. That is, how much of an opportunity you have to observe that person actually performing the job, or to observe the end product of the performance. You will note that not all scale points have a written description. However if you feel your level of familiarity with a person's work falls at a particular point between two scale points that have descriptions, you should certainly feel free to use that scale point. For example, if you feel that your level of familiarity with an individual's work is less than, but close to "Extremely Familiar", you would probably choose scale point "6".

Please consider all points on the scale and provide your best response. **YOUR RESPONSES WILL BE COMPLETELY CONFIDENTIAL AND WILL BE UTILIZED FOR RESEARCH PURPOSES ONLY.**

If your name appears on the list, **DO NOT RATE YOURSELF.** Additionally, **PLEASE DO NOT DISCUSS THIS QUESTIONNAIRE WITH ANYONE.** The information is being used for research purposes only. None of the individual responses from this research will be shared with the Division or Department. Please return the completed questionnaire in the enclosed pre-addressed, stamped envelope.

Rating Scale

-
- 1 - Not Very Familiar
 - 2 -
 - 3 -
 - 4 - Somewhat Familiar
 - 5 -
 - 6 -
 - 7 - Extremely Familiar

Name

Familiarity Rating

**PLEASE RETURN THE COMPLETED QUESTIONNAIRE IN THE
ENCLOSED PRE-ADDRESSED, STAMPED ENVELOPE**

Code # _____

APPENDIX D

Dear Colleague:

This packet contains a number of Performance Appraisal forms and a list of individuals that you are being asked to rate. Each individual on the list has been assigned a code number; you should record that number - not the individual's name - on the line on the top of the Performance Appraisal form entitled "CODE # of person you are rating: _____". Only the identification number, and the ratings for Knowledge of Work, Quantity of Work, Quality of Work, Attendance and Ability to Deal With People need to be recorded. You have also been assigned a code number which appears on the Performance Appraisal. The code numbers have been assigned by an individual not employed by the Department. I will NOT be aware of your identity, or the identities of the individuals you are rating. You should NOT discuss the ratings, or the names of the individuals you have rated, with anyone. **ALL INFORMATION USED IN THIS STUDY WILL BE STRICTLY CONFIDENTIAL AND USED FOR RESEARCH PURPOSES ONLY. THE RESEARCH IS BEING CONDUCTED FOR THE PURPOSE OF FULFILLING THE REQUIREMENTS FOR AN ACADEMIC DEGREE. INDIVIDUAL RESPONSES WILL NOT BE SHARED WITH THE DIVISION OR DEPARTMENT.**

For each individual on the list, please provide your best ratings, for the last twelve months, on each of the five dimensions. You may feel that you have little basis to provide ratings for one or more individuals. In that case, please do the best you can in providing ratings. **IT IS IMPORTANT THAT YOU KEEP THE NAMES OF THE INDIVIDUALS YOU HAVE RATED, AND THEIR RATINGS, STRICTLY CONFIDENTIAL.**

PLEASE RETURN THE COMPLETED MATERIALS IN THE ENCLOSED PRE-ADDRESSED, STAMPED ENVELOPE

RATER CODE # _____

RATEE ASSIGNMENT

<u>NAME</u>	<u>CODE #</u>
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

*NOTE: HALF OF THE RATERS WILL RECEIVE AN ADDITIONAL FIVE INDIVIDUALS TO RATE; THIS WILL REPRESENT THE ENVIRONMENTAL COMPLEXITY MANIPULATION; THE ADDITIONAL RATEES WILL BE ASSIGNED TO THE RATING LIST ON A RANDOM BASIS.

CODE # _____

CODE # OF PERSON YOU ARE RATING: _____

INSTRUCTIONS

Evaluate the employee on the job now being performed. Check the box above the horizontal line which most nearly coincides with your overall judgment of each quality. The care and accuracy with which this appraisal is made will determine its value to you, to the employee and to the agency.

JOB ELEMENTS	LESS THAN GOOD				
	EXCELLENT	VERY GOOD	GOOD	FAIR	UNSATISFACTORY
KNOWLEDGE OF WORK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consider knowledge of job gained through experience, general education, specialized training.	Well informed on all phases of work	Knowledge thorough enough to perform without assistance	Adequate grasp of essentials, some assistance	Requires considerable assistance	Inadequate knowledge
QUANTITY OF WORK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consider the volume of work produced under normal conditions. Disregard errors.	Rapid worker unusually large volume	Turns out large volume	Average	Volume below average	Very slow worker
QUALITY OF WORK	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consider neatness accuracy and dependability of results regardless of volume.	Exceptionally accurate	Seldom necessary to check work	Acceptable, usually neat, few errors	Often unacceptable. Frequent errors or rejections	Too many errors or rejections
ATTENDANCE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consider tardiness and occasions of absenteeism.	Rarely late, rarely absent	Very seldom late, seldom absent	Seldom late, absent only occasionally	Frequently tardy or routinely absent	Habitually tardy or excessive absenteeism
ABILITY TO DEAL WITH PEOPLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Relationships with other staff and the public. Consider harmony in and outside the work unit.	Brings credit to state service through cooperative dealings with staff and public	Harmonious relationships with staff and/or public	Adequate work relationships with staff and/or public	Reluctant to cooperate with staff and/or public	Creates problems through personal interaction with staff and/or public

REFERENCES

- American Psychological Association (1981). Ethical principles of psychologists (revised). American Psychologist, 36, 633-638.
- Anastasi, Anne (1988). Psychological Testing (sixth edition). New York:Macmillan.
- Athey, T.R., & McIntyre, R.M. (1980). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. Journal of Applied Psychology, 72, 567-572.
- Banks, C.G., & Murphy, K.R. (1985). Toward narrowing the research-practice gap in performance appraisal. Personnel Psychology, 38, 335-345.
- Bartlett, C.J. (1983). What's the difference between valid and invalid halo? Forced-choice measurement without forcing a choice. Journal of Applied Psychology, 68, 218-226.
- Bass, A.R., & Turner, J.N. (1973). Ethnic group differences in relationship among criteria of job performance. Journal of Applied Psychology, 57, 101-109.
- Bernardin, H.J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 63, 301-308.
- Bernardin, H.J., Alvares, K.M., & Cranny, C.J. (1976). A recomparison of behavioral expectation scales to summated scales. Journal of Applied Psychology, 61, 564-570.

- Bernardin, H.J., & Buckley, M.R. (1981). A consideration of strategies in rater training. Academy of Management Review, 6, 205-212.
- Bernardin, H.J., & Pence, E.C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.
- Bernardin, H.J., & Smith, P.C. (1981). A clarification of some issues regarding the development and use of behaviorally-anchored rating scales (BARS). Journal of Applied Psychology, 66, 458-463.
- Bernardin, H.J., & Walter, C.S. (1977). Effects of rater training and diary-keeping on psychometric error in rating. Journal of Applied Psychology, 62, 64-69.
- Bieri, J. (1955). Cognitive complexity-simplicity and predictive behavior. Journal of Abnormal and Social Psychology, 51, 263-268.
- Block, J., Weiss, D.S., & Thorne, A. (1979). How relevant is a semantic similarity interpretation of personality ratings? Journal of Personality and Social Psychology, 37, 6, 1055-1074.
- Boehm, V.R. (1977). Differential prediction: A methodological artifact? Journal of Applied Psychology, 62, 146-154.
- 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. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 238-252.
- Borman, W.C. (1979). Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.
- Borman, W.C. (1979). Format and training effects on rating accuracy and rating errors. Journal of Applied Psychology, 64, 4, 410-421.
- Brecker, N. (1988). The effects of rater training, environmental complexity, cognitive complexity and rater intelligence on performance appraisal accuracy. University Microfilms International (order # 8817333).
- Brown, E.M. (1968). Influence of training, method, and relationship on the halo effect. Journal of Applied Psychology, 52, 195-199.
- Cardy, R.L., & Kehoe, J.F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of a cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 69, 589-594.
- Cascio, W.F. (1987). Applied Psychology in Personnel Management (third edition). Englewood Cliffs, N.J.:Prentice-Hall.
- Cooper, W.H. (1981a). Conceptual similarity as a source of illusory halo in job performance ratings. Journal of Applied Psychology, 66, 302-307.

- Cooper, W.H. (1981b). Ubiquitous halo. Psychological Bulletin, 90, 218-244.
- Cronshaw, S.F., & Lord, R.G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. Journal of Applied Psychology, 72, 1, 97-106.
- Cummings, L.L., & Schwab, D.P. (1973). Performance in organizations: Determinants and appraisal. Glenview, IL:Scott, Foresman.
- DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 21, 635-646.
- DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive view of the performance appraisal process: A model and some research propositions. Organizational Behavior and Human Performance, 33, 360-396.
- Dunnette, M.D., & Borman, W.C. (1979). Personnel selection and classification systems. Annual Review of Psychology, 30, 474-525.
- Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Fisicaro, S.A. (1988). A reexamination of the relation between halo error and accuracy. Journal of Applied Psychology, 73, 239-244.
- Friedman, B.A., & Cornelius, E.T. (1976). Effect of rater participation in scale construction on the psychometric properties of two rating scale formats. Journal of Applied Psychology, 61, 210-216.

- Golembiewski, R.T., Billingsley, K., & Yaeger, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. Journal of Applied Behavioral Science, 12, 133-157.
- Graves, L.M. (1985). Effects of leader persistence and environmental complexity on leadership perceptions: Do implicit beliefs discourage adaptation to complex environments? Group and Organizational Studies, 10, 19-36
- Guilford, J.P., & Fruchter, B. (1978). Fundamental Statistics in Psychology and Education (sixth edition). New York:McGraw Hill.
- Harvey, R.J. (1982). The future of partial correlation as a means to reduce halo in performance ratings. Journal of Applied Psychology, 67, 171-176.
- Hauenstein, N.M.A., & Foti, R.J. (1989). From laboratory to practice: Neglected issues in implementing frame-of-reference rater training. Personnel Psychology, 42, 359-376.
- Hays, W.L. (1973). Statistics for the Social Sciences (second edition). New York:Holt, Rinehart & Winston.
- Hedge, J.W., & Kavanagh, M.J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. Journal of Applied Psychology, 73, 68-73.
- Hulin, C.L. (1982). Some reflections on general performance dimensions and halo rating error. Journal of Applied Psychology, 67, 165-170.

- Ilgen, D.R., & Feldman, J.M. (1983). Performance appraisal: A process focus. In L.L. Cummings & B.M. Staw (Eds.), Research in Organizational Behavior, Vol. 5, Greenwich, Ct.:JAI Press.
- Jacobs, R., & Kozlowski, S.W.J. (1985). A closer look at halo error in performance ratings. Academy of Management Journal, 28, 201-212.
- Joyce, W.F., Slocum, J.W., & von Glinow, M.A. (1982). Person-situation interactions: Competing models of fit. Journal of Occupational Behaviour, 3, 265-280.
- Kane, J.S., & Lawler, E.E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In L.L. Cummings & B.M. Staw (Eds.), Research in Organizational Behavior, Vol. 1., Greenwich, CT.:JAI Press.
- Kelly, G.A. (1955). The psychology of personal constructs (Vols. 1 & 2). New York:Norton.
- Kerlinger, F.N. (1986). Foundations of Behavioral Research (third edition). New York:Holt, Rinehart and Winston.
- King, L.M., Hunter, J.E., & Schmidt, F.L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. Journal of Applied Psychology, 65, 507-516.
- Kirk, R. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA:Brooks/Cole.
- Koltuv, B.B. (1962). Some characteristics of intrajudge trait intercorrelations. Psychological Monographs, 76, 552.

- Kozlowski, S.W.J., Kirsch, M.P., & Chao, G.T. (1986). Job knowledge, rater familiarity, conceptual similarity and halo error: An exploration. Journal of Applied Psychology, 71, 45-49.
- Kozlowski, S.W.J., & Kirsch, M.P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. Journal of Applied Psychology, 72, 252-261.
- Kraut, A.J. (1975). Prediction of managerial success by peer and training-staff ratings. Journal of Applied Psychology, 60, 14-19.
- Krzystofiak, F., Cardy, R., & Newman, J. (1988). Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior. Journal of Applied Psychology, 73, 515-521.
- Landy, F.J., Vance, R.J., Barnes-Farrell, J.L., & Steele, J.W. (1980). Statistical control of halo error in performance ratings. Journal of Applied Psychology, 65, 501-506.
- Landy, F.J., & Farr, J.L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Lord, R.G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. Journal of Applied Psychology, 70, 66-71.
- McGrath, J.E., & Rotchford, N.L. (1983). Time and behavior in organizations. In L.L. Cummings and B.M. Staw (Eds.), Research in Organizational Behavior, Vol. 5, Greenwich, CT:JAI Press.
- McIntyre, R.M. Smith, D.E., & Hassett, C.E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.

- Mitchell, T.R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), Performance Measurement and Theory, Hillsdale, NJ:Erlbaum.
- Mount, M.K., & Thompson, D.E. (1987). Cognitive categorization and quality of performance ratings. Journal of Applied Psychology, 72, 240-246.
- Murphy, K.R. (1982). Difficulties in the statistical control of halo. Journal of Applied Psychology, 67, 161-164.
- Murphy, K.R., & Balzer, W.K. (1986). Systematic distortion in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.
- Murphy, K.R., & Constans, J.I. (1987). Behavioral anchors as a source of bias in rating. Journal of Applied Psychology, 72, 573-577.
- Murphy, K.R., & Jako, R. (1989). Under what conditions are observed intercorrelations greater or smaller than true intercorrelations? Journal of Applied Psychology, 74, 827-830.
- Murphy, K.R., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observations? Journal of Applied Psychology, 67, 562-567.
- Nathan, B.R., & Lord, R.G. (1983). Cognitive categorization and dimensional schemata: A process approach in performance ratings. Journal of Applied Psychology, 68, 102-114.

- 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, 290-296.
- Pulakos, E.D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.
- Pulakos, E., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within rates to measure halo. Journal of Applied Psychology, 71, 29-32.
- Rothstein, H.J. (1986). The effects of time pressure on judgment in multiple cue probability learning. Organizational Behavior and Human Decision Processes, 37, 83-92.
- Saal, F.E., & Landy, F.J. (1977). The mixed standard rating scale: An evaluation. Organizational Behavior and Human Performance, 18, 19-35.
- Saal, F.E., Downey, R.G., & Lahey, M.A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 88, 413-428.
- Schneier, C.E. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 62, 541-548.
- Sharon, A.T., & Bartlett, C.J. (1969). Effects of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22, 251-263

- Shweder, R.A. (1975). How relevant is an individual difference theory of personality? Journal of Personality, 43, 455-484.
- Shweder, R.A., & D'Andrade, R.G. (1979). Accurate reflection or systematic distortion? A reply to Block, Weiss, & Thorne. Journal of Personality and Social Psychology, 37, 1075-1084.
- Shweder, R.A., & D'Andrade, R.G. (1980). The systematic distortion hypothesis. In R.A. Shweder & D.W. Fiske (Eds.), New directions for methodology of behavioral science: Fallible judgment in behavioral research (pp. 37-58). San Francisco: Jossey-Bass.
- Smither, J.W., & Reilly, R.R. (1989). Relationship between job knowledge and the reliability of conceptual similarity schemata. Journal of Applied Psychology, 74, 530-534.
- Steel, R.P., & Mento, A.J. (1986). Impact of situational constraints on subjective and objective criteria of managerial job performance. Organizational Behavior and Human Decision Processes, 37, 254-265.
- Tabachnick, B.G., & Fidell, L.S. (1989). Using Multivariate Statistics (Second Edition). New York: Harper & Row.
- Thornton, G.C., & Zorick, S. (1980). Training to improve observer accuracy. Journal of Applied Psychology, 65, 351-354.
- Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. Science, 85, 1124-1131.

- Weiss, D.S., & Mendelsohn, G.A. (1986). An empirical demonstration of the implausibility of the semantic similarity explanation of how trait ratings are made and what they mean. Journal of Personality and Social Psychology, 50, 3, 595-601.
- Wexley K.N., Yukl, G.A. (1984). Organizational Behavior and Personnel Psychology. Homewood, IL:Richard D. Irwin, Inc.
- Wood, R.E. (1986). Task complexity: Definition of the construct. Organizational Behavior and Human Decision Processes, 37, 60-82.
- Zammuto, R.F., London, M., & Rowland, K.M. (1982). Organization and rater differences in performance appraisals. Personnel Psychology, 35, 643-658.
- Zedeck, S., & Cascio, W.F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.