

Training Managers to Minimize Rating Errors in the Observation of Behavior

Gary P. Latham
Weyerhaeuser Company
Tacoma, Washington

Kenneth N. Wexley
University of Akron

Elliott D. Pursell
Weyerhaeuser Company
Plymouth, North Carolina

Sixty managers in a large corporation were randomly assigned to a workshop, a group discussion, or a control group. The workshop and group discussion involved training directed toward the elimination of rating errors that occur in performance appraisal and selection interviews, namely, contrast effects, halo effect, similarity, and first impressions. Six months after the training, the managers rated hypothetical candidates who were observed on videotape. The results showed that (a) trainees in the control group committed similarity, contrast, and halo errors; (b) trainees in the group discussion committed impression errors; and (c) trainees in the workshop committed none of the errors. The importance of observer training for minimizing the "criterion problem" in industrial psychology is discussed.

A major problem in the selection and/or development of criteria as evaluative indices of human performance is the lack of reliability in the observation of behavior (Ronan & Prien, 1966, 1971). This unreliability can be largely attributed to well-known rating errors such as first impressions and the halo effect. Rating errors are errors in judgment that can occur when one individual observes another. Research designed to overcome this problem has been restricted primarily to manipulating the format of the ratings (Berkshire & Highland, 1953; Blanz & Ghiselli, 1972; Smith & Kendall, 1963); examining the ratings given by observers in different organizational roles (e.g., Kavanagh, MacKinney, & Wolins, 1971; Klimoski & London, 1974; Lawler, 1967; Ronan & Latham, 1974); and/or studying the kind of content or stimuli that are rated, namely, personality versus performance variables (Kavanagh, 1971). In spite of this work, managers and supervisors have continued to make rating errors when

observing and evaluating present or potential employees.

It would seem logical that to solve the problem of rating errors, the observer must be trained. For example, Levine and Butler (1952) worked with supervisors who were overrating employees in the higher job grades and underrating those in the lower grades. The supervisors were randomly assigned to a control, a lecture, or a discussion group. Supervisors in the lecture group were given a detailed lecture on the technique and theory of performance rating. They were also given background information on wage administration and job evaluation. The lecturer explained to the supervisors the problem caused by their previous ratings and what each supervisor needed to do to correct the problem. After the lecture, questions were encouraged and answered by the lecturer. In the discussion group, the supervisors met together to discuss the nature of the problem and how it could be solved. The discussion leader merely acted as a moderator and avoided injecting his own opinions. After expressing a number of ideas, the group then arrived at one decision acceptable to all members. The results showed that no observable behavior change occurred on the part of the supervisors in the control group. Similarly, the

The authors are grateful to Terence R. Mitchell and Gary A. Yukl for their helpful comments in writing this article. A shorter version of this paper was presented at the annual meeting of the Canadian Psychological Association, Quebec City, 1975.

Requests for reprints should be sent to Gary P. Latham, Human Resources Research, Weyerhaeuser Company, Tacoma, Washington 98402.

NUCLEAR REGULATORY COMMISSION

Docket No. 50-372-06-05 Official Exh. No. 53

In the matter of Shoreham Nuclear Power Station

Staff _____ IDENTIFIED ☒

Applicant _____ RECEIVED _____

Intervenor ☒ _____ REJECTED _____

Cont'g Off'r _____

Contractor _____ DATE 4-30-87

Other _____ Witness _____

Reporter Mary Simons

ing Errors in the for

Kenneth N. Wexley
University of Akron

assigned to a workshop,
and group discussion
rating errors that occur
namely, contrast effects,
after the training, the
ved on videotape. The
committed similarity,
discussion committed im-
mitted none of the
imizing the "criterion

d evaluating present or potential

seem logical that to solve the
rating errors, the observer must
For example, Levine and Butler
led with supervisors who were
employees in the higher job grades
ing those in the lower grades.
ors were randomly assigned to a
ecture, or a discussion group.
n the lecture group were given a
re on the technique and theory
ce rating. They were also given
information on wage administra-
evaluation. The lecturer ex-
e supervisors the problem caused
vious ratings and what each
eded to do to correct the prob-
e lecture, questions were encour-
wered by the lecturer. In the
up, the supervisors met together
e nature of the problem and
be solved. The discussion leader
as a moderator and avoided
own opinions. After expressing
ideas, the group then arrived at
acceptable to all members. The
d that no observable behavior
red on the part of the super-
control group. Similarly, the

lecture had no influence on changing the supervisors' method of rating. This finding indicated that knowledge of these errors alone will not lead raters to take effective steps to counteract them. Only the group discussion method was effective in modifying the supervisors' rating behavior.

Wexley, Sanders, and Yukl (1973) showed that despite attempts to reduce contrast effects by means of a warning (lecture) and/or an anchoring procedure (comparing applicants to a preset standard or anchor), subjects continued to make this error. Only an intensive training workshop led to a significant behavior change. Specifically, the workshop gave subjects a chance to practice observing and rating actual videotaped candidates. In addition, the subjects were given immediate feedback regarding the accuracy of their ratings.

A possible limitation of this study may have been the fact that it was conducted in a laboratory setting with undergraduate psychology students. Thus, the generalizability of the results to experienced individuals in industry was questionable. Moreover, this study, as well as Levine and Butler's (1952), only eliminated one rating error, and the effects of the training were not assessed over time.

The purpose of the present research was to evaluate the workshop approach (Wexley et al., 1973) and the group-discussion method (Levine & Butler, 1952) in training employees in a large corporation to minimize the following four rating errors: contrast effects, halo effect, similarity, and first impressions. Data were collected 6 months after the training was completed to assess the training effects over time.

METHOD

Subjects

Sixty managers in a large corporation were randomly assigned to one of three conditions: a workshop, a group discussion, or a control group that received no training. Each group contained 20 individuals. An underlying assumption of this study was that all managers need this type of training to some extent. The only criterion for participation was that the trainee's duties require frequent performance appraisals of subordinates and/or the selection of potential employees.

Procedure

Training the trainers. Six personnel supervisors who were to act as trainers were given 3 days of training. Three were randomly assigned to receive training in conducting the workshop and the remaining three were given training in conducting a group discussion. Examples of each rating error, made in a performance appraisal, a selection interview, and in an off-the-job setting (e.g., making judgments about neighbors), were demonstrated and/or discussed with each group of trainers.

The trainees were brought to one of three locations for training. At each location, one trainer conducted the workshop while another conducted the group discussion.

Workshop. The workshop included videotapes of hypothetical job candidates being appraised by a manager. The development of the videotapes was identical to the procedure discussed by Wexley, Yukl, Kovacs, and Sanders (1972). The trainees gave a rating on a 9-point scale as to how they thought the manager in the videotape evaluated the candidate, and how they themselves would rate the candidate. Group discussion followed as to the reasons for each trainee's rating of both the manager's evaluation and the candidate. In this way, the trainees had an opportunity to observe videotaped managers making observational errors, to actively participate in discovering the degree to which they themselves were prone to making such errors, to receive knowledge of results as to their own behavioral observations, and to practice job-related tasks to reduce any errors they were committing.

The first exercise focused on the similarity, or *similar-to-me* effect. This error is a tendency on the part of a rater to judge more favorably those he perceives as similar to himself (Rand & Wexley, 1975; Wexley & Nemeroff, 1974). That is, the more closely an assessee resembles the rater in attitudes or background, the stronger the tendency of the rater to judge that individual higher.

The trainees were given a job description and a list of the job requirements for a loan officer's position. They then observed a videotape in which there was a strong attitudinal and biographical similarity between the manager and a below-average candidate. Relatively little job-related information was elicited by the manager from this candidate. At the end of the videotape, a colleague of the manager requested an evaluation of the candidate. At this point the videotape was stopped. Each trainee was asked to give two ratings: (a) How would you rate the candidate? and (b) How do you think the manager rated the candidate? They then discussed their ratings among themselves in relation to the similarity error. Finally, the trainees were shown the remainder of the videotape. The tape showed the manager extolling the virtues of the candidate to his colleague and thus committing the *similar-to-me* error.

The second exercise used the same job description and requirements but focused on the *halo*

effect. This error refers to the rater's exaggeration of the homogeneity of an individual's characteristics or traits. The trainees watched a videotape in which a manager was so impressed by a candidate's nonjob-related background that he was also impressed with the candidate's job-related qualifications even though the latter were highly dubious. The trainees were asked to rate the candidate as they thought the manager would rate him and to indicate how accurate they thought the manager was in his rating by supplying their own rating of the applicant. The trainees then discussed their ratings as well as solutions for reducing this error.

The third exercise dealt with *contrast* effects (Wexley et al., 1972). The trainees were given a job description and a list of the job requirements for an accountant's job. They were then given the résumé of a highly qualified applicant and were asked to rate him. The procedure was then repeated with a second highly qualified applicant and then with an average applicant. The fact that evaluations of job applicants can be affected by the suitability of immediately preceding applicants and that subordinates are often evaluated in comparison to other subordinates rather than against an established standard was discussed. The necessity of basing ratings against a predetermined anchor or standard was emphasized. The trainees were then trained in accordance with the principles and procedures described in Wexley et al. (1973).

The final exercise concerned *first impressions* error. This error is committed when an observer evaluates someone on the basis of judgments made primarily after an initial meeting. The trainees were given a job description and a listing of the specific job requirements for an insurance rater. They were then shown a videotape of an interview in which the applicant presented a poor impression by her answers, actions, and appearance. The remainder of the videotape showed that the applicant was acceptable for the job; however, the manager continued to act on the basis of his initial impression. Again, the trainees were requested to give two types of ratings: one based on their own opinion of the candidate and the other based on their opinion of what the manager thought of the applicant. The trainees discussed their individual ratings as well as ways to reduce this observation error.

Group discussion. The format for the group discussion required the trainer to first define the four rating errors. An example of each error was given for three situations: the performance appraisal, the selection interview, and an off-the-job situation. The trainees then divided into subgroups of three or four people and generated personal examples that they could remember experiencing in each of the three situations. These examples were subsequently shared between subgroups during a general group discussion. The trainees again returned to their respective subgroups and generated solutions to each of the rating problems. The solutions were also shared between subgroups during an overall group meeting. These solutions were identical to those decided upon

in the workshop group. Thus, the primary difference between the two training programs was the method used to eliminate the rating errors.

Testing. The individuals in the workshop, group-discussion, and control groups were tested together 6 months after the training program. Two forms, A and B, were used for testing each rating error. The control, group-discussion, and workshop trainees were each randomly divided so that half of each group (i.e., 10 individuals) received Form A and half received Form B.

Before seeing the first videotape, all subjects were given the following instructions:

You are going to rate some individuals. Please rate each person in terms of the degree to which you feel that he or she is acceptable for the job. There are no tricks. For example, if you see someone who obviously does not fulfill the requirements of the job, don't say to yourself that this is obviously a trick so I will rate him as acceptable. If he is unacceptable, rate him as unacceptable. Similarly, if you see someone who is obviously terrific, don't fool yourself by saying there must be something wrong somewhere. Rate each individual the way you actually see him.

During testing, the subjects were given detailed job descriptions and lists of the requirements for each job in question. None of the videotapes had been seen previously by any of the subjects during training. In testing for similarity error, the subjects who received Form A saw a candidate who fulfilled the requirements of the job, but whose biographical background and attitudes were those of a culturally disadvantaged individual; those subjects receiving Form B saw the same individual, but in this instance his background and attitudes were those of a typical middle-class, job candidate. If a given training program were effective, the trainees should have given this individual approximately the same rating on the 9-point hiring-recommendation scale regardless of the tape they saw.

First impressions were tested by changing the order of favorable and unfavorable responses given by an individual. Form A showed a woman who presented an unfavorable first impression but later in the interview presented a favorable impression. Form B presented the same individual with a favorable-unfavorable sequence of responses. The actual responses given by the woman were identical in both forms, but the order of the questions and answers in the interview were reversed. Ratings were again made on a 9-point hiring-recommendation scale.

Halo was established by presenting trainees with an individual's folder containing an application blank, references, interview information, and work experience. Form A contained extremely unfavorable information and Form B contained highly favorable information. However, no information was given on the individual's educational background. The trainees were subsequently shown a videotape of this individual being questioned about

R
Sin
Fin
Cor
Hal

Note:
indicat
the area

ber edi
the of
were e
this in
educati
the fa
been co

Testi
videota
Form A
lowed
a low
plicant.
recomm
applicat
contras

For
rating
would
5—wou
be mad
be mad
fied as
tional
weak e
1—appl
cations

Finall
consist
the exte
benefici
months.

The
were e
reactio
mitted.
measur
(SD =
sion tr

op group. Thus, the primary difference between the two training programs was the elimination of the rating errors.

Individuals in the workshop, group-control groups were tested together in the training program. Two forms, used for testing each rating error, group-discussion, and workshop training, randomly divided so that half of (10 individuals) received Form A and half received Form B.

In the first videotape, all subjects were given instructions:

"You are going to rate some individuals. Please be consistent in terms of the degree to which you think he or she is acceptable for the job. There are no tricks. For example, if you see someone who obviously does not fulfill the requirements of the job, don't say to yourself that it is just a trick so I will rate him as unacceptable; rate him as unacceptable. Similarly, if you see someone who is terrific, don't fool yourself by saying that something is wrong somewhere. Rate him the way you actually see him."

The subjects were given detailed job descriptions and lists of the requirements for each job. None of the videotapes had been used by any of the subjects during training. For similarity error, the subjects who saw a candidate who fulfilled the requirements of the job, but whose biographical data and attitudes were those of a culturally different individual; those subjects receiving the same individual, but in this videotape background and attitudes were those of a middle-class, job candidate. If a given videotape were effective, the trainees should rate the individual approximately the same on the 9-point hiring-recommendation scale as the videotape they saw.

Subjects were tested by changing the favorable and unfavorable responses given in the videotape. Form A showed a woman who gave an unfavorable first impression but later presented a favorable impression. Form B presented the same individual with a favorable sequence of responses. The responses given by the woman were identical, but the order of the questions and the interview were reversed. Ratings were made on a 9-point hiring-recommendation scale.

Subjects were established by presenting trainees with a folder containing an application letter, interview information, and work sample. Form A contained extremely unfavorable information and Form B contained highly favorable information. However, no information about the individual's educational background was given. The trainees were subsequently shown a videotape of this individual being questioned about

TABLE 1
MEAN RATINGS AND STANDARD DEVIATIONS FOR THE TRAINING AND CONTROL GROUPS

Response error	Form	Workshop		Group discussion		Control	
		\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
Similar-to-me	A	3.7	2.21	4.8	2.78	3.4	2.46
	B	4.7	2.41	4.4	1.90	5.9	1.37
First impressions	A	7.5	.85	7.0	1.89	5.8	2.04
	B	7.0	1.15	4.7	2.45	5.6	2.59
Contrast effects	A	5.8	2.20	3.9	2.33	3.0	2.45
	B	5.9	1.73	4.6	2.63	6.5	1.18
Halo	A	8.1	.99	8.2	1.23	6.8	1.48
	B	8.5	.71	8.4	.70	8.4	.97

Note. All ratings, with the exception of those involving the halo effect, were made on 9-point scales of the decision to hire, with 9 indicating an accept decision. The halo effect ratings were made on a 9-point scale of the applicant's education, with 9 indicating the greatest education.

her education. Ratings were made on education and the other job-related factors. If the trainees who were exposed to the unfavorable information gave this individual approximately the same rating on education as did the trainees who were exposed to the favorable information, halo error would have been considered eliminated.

Testing for contrast effects involved use of the videotapes from the Wexley et al. (1973) study. Form A presented a highly qualified applicant followed by an average applicant. Form B presented a low applicant followed by the same average applicant. Ratings were made on the 9-point hiring-recommendation scale. The ratings of the average applicant were analyzed to determine the extent of contrast effects.

For all test exercises except the one for halo, the rating scales were benchmarked as follows: 9—would recommend strongly that an offer be made; 5—would recommend with reservations that an offer be made; and 1—would recommend that no offer be made. The halo-exercise benchmarks were modified as follows: 9—applicant has excellent educational qualifications for the job; 5—applicant has weak educational qualifications for the job; and 1—applicant has unsatisfactory educational qualifications for the job.

Finally, subjects completed a reaction measure consisting of a 9-point scale on which they rated the extent to which they felt the program had been beneficial to them on their jobs during the past 6 months.

RESULTS

The results of the two training programs were evaluated on the basis of the trainee reaction measures and the rating errors committed. The mean rating on the reaction measure for the workshop trainees was 8.08 ($SD = .92$), while that for the group-discussion trainees was 6.25 ($SD = 1.52$); a rating

of 9 indicated the greatest benefit. The difference in the ratings between the two training programs was highly significant ($t = 5.80, p < .001$).

The means and standard deviations of the ratings given to the individuals on the videotapes are presented in Table 1. Comparisons among means were made between Forms A and B for each of the three groups using two-tailed t tests. The results indicated that the control group committed the following rating errors: similar-to-me ($t = 2.81, p < .05$); halo ($t = 2.87, p < .05$); and contrast effects ($t = 4.07, p < .01$). The ω^2 analyses (Hays, 1973) revealed that similarity, halo, and contrast errors accounted for 26%, 27%, and 44% of the variance in the ratings, respectively. The participants in the group discussion exhibited only one error. Rather than committing a first-impressions error, these individuals displayed a sizable recency or "last-impressions" error ($t = 2.35, p < .05, \omega^2 = 18\%$). The ω^2 values for similarity, halo, and contrast errors were found to be close to zero. The trainees in the workshop did not commit any of the rating errors; all t tests were nonsignificant and all ω^2 estimates were negligible (i.e., 0%-1%).

DISCUSSION

The difficulty in developing adequate measures of an individual's job performance is generally referred to as the "criterion problem" by industrial psychologists. Lifson (1953) showed that up to one third of per-

formance-measurement variance was due to rater differences.¹ As Ronan and Prien (1971) have pointed out, the fact that this finding was obtained "in a controlled experimental situation makes it conceivable that even more such unwanted variance occurs in field studies having less stringent attempts at control" (p. 151).

The significance of the present study is that it has shown that this aspect of the criterion problem can be reduced by training observers to minimize rating errors. In addition, the study provides external validity (Campbell & Stanley, 1966) for the previous research by Levine and Butler (1952) and Wexley et al. (1973). Specifically, these results show that training programs that carefully apply basic principles of learning in accordance with the task being taught can eliminate more than one rating error committed by subjects from a different population. Levine and Butler (1952) trained first-line foremen, Wexley et al. (1973) trained college students, and the present authors trained managers in a corporate setting. Moreover, the present study showed that the results of this training are sustained over time.

A possible limitation of this study is that the testing was a simulation rather than an actual measure of the trainee's on-the-job behavior. It would have been practically infeasible to assess the degree to which these managers actually make all of these errors in a real performance appraisal. The present approach, however, is consistent with Hemphill's claim made in 1964 (cited in Wallace, 1965) that the only hope for evaluating managers is to establish simulated criteria in which their performance can be compared to known standards under controlled, standardized conditions. Moreover, research cited by Lifson (1953) has shown that filmed performances are rated the same as live performances.

¹ Lifson (1953) was concerned with the nature of errors that are involved in time-study pace ratings. It is interesting to note that his raters were similar to those used in our control group, in that both sets of individuals had considerable experience in observing and evaluating workers, but had not received training in ways to reduce rating errors.

The finding that the group-discussion trainees rather than the control group committed the error of first impressions is puzzling. Anecdotal evidence from the trainers indicated that the trainees in this group considered discussion of the error irrelevant because "the error is so obvious that no experienced manager commits it." Consequently, there was relatively little group discussion on it. Inspection of Table 1 indicates that during testing, these trainees actually made a recency error. It appears that in the process of proving their point they "leaned over backwards."

That the workshop appears to be slightly more effective than the group discussion in eliminating rating errors could be a result of one or more combinations of factors. First, the trainees in the group discussion were able to obtain knowledge of results from the trainer and from each other as to their personal understanding of the errors and their solutions, but unlike the trainees in the workshop, they did not receive feedback as to their own specific behavior in committing an error.

Second, the trainees in the workshop reported that the highly structured format of their program made them feel that the time required away from their job for this training was being used wisely. This was not always the case with trainees in the group discussion. Although this group was given greater freedom to participate in the structuring of the training content, they wanted more feedback from the trainer than from each other. Consequently, many of them expressed a lack of interest in the program. This finding is consistent with the trainee reaction data.

Third, the workshop training lasted 9 hours, while the group discussion lasted 6 hours. This difference was due to the time required in the workshop to set up and show the various videotapes. Although this time difference could have explained the superiority of the workshop trainees if testing had occurred immediately after training, it is difficult to believe that this difference by itself could explain test results obtained 6 months later.

Fourth, it is possible that the trainees in

the gr
had m
compa
indiv
too, se
exper
task tl
ivities
their r
and th
job sec

Fina
the sin
and th
have t
error
accordi
because
that th
what e
given v

It is
this stu
program
advant
costly
expense
renting
proach
the ma
worksh
pared
ers, be
program
light o
the gro
positive
worksh
near to
rating
cost ar
training
discuss
benefici
tween
will be
circums

Berkshire
perform
Person

ding that the group-discussion rather than the control group error of first impressions is puzzle. Evidence from the trainers that the trainees in this group discussion of the error irrelevant behavior is so obvious that no experienced manager commits it." Consequently, relatively little group discussion on of Table 1 indicates that during these trainees actually made a reversal. It appears that in the process of their point they "leaned over back."

workshop appears to be slightly more than the group discussion in rating errors could be a result of combinations of factors. First, in the group discussion were able knowledge of results from the from each other as to their understanding of the errors and their unlike the trainees in the workshop did not receive feedback as to specific behavior in committing an

the trainees in the workshop received the highly structured format of made them feel that the time from their job for this training used wisely. This was not case with trainees in the group. Although this group was given om to participate in the structured training content, they wanted from the trainer than from. Consequently, many of them lack of interest in the program. is consistent with the trainee

workshop training lasted 9 the group discussion lasted 6 difference was due to the time the workshop to set up and show videotapes. Although this time would have explained the superior workshop trainees if testing had immediately after training, it is believe that this difference by to plain test results obtained 6

is possible that the trainees in

the group discussion and the control group had more anxiety during testing due to their comparative lack of experience in observing individuals on videotapes. This explanation, too, seems doubtful, in that the subjects were experienced managers who were engaging in a task that was similar to their day-to-day activities. Moreover, the managers knew that their ratings during testing were anonymous and thus could not have any impact on their job security.

Finally, the findings may be an artifact of the similarity between the workshop training and the testing. That is, the subjects may have been able to recognize when a given error was being tested and thus responded accordingly. This explanation was ruled out because postexperimental interviews indicated that the trainees had no idea whatsoever of what error was actually being tested on a given videotape.

It is difficult to conclude on the basis of this study that only one of the two training programs should be used. The primary disadvantage of the workshop was that it was costly and time-consuming to develop. The expense of preparing the videotapes and the renting and/or purchasing of equipment approached \$1000. On the other hand, once the materials were developed, training the workshop leaders was relatively easy compared to training the group-discussion leaders, because the major part of the training program was on the use of the videotapes. In light of the one rating error committed by the group-discussion trainees and the more positive reaction to the workshop by the workshop trainees, the workshop would appear to be slightly more effective for reducing rating errors. On the other hand, when the cost and time for developing a structured training program is prohibitive, the group-discussion method appears to be a highly beneficial alternative. Thus, the choice between the two methods, in most instances, will be contingent upon a given organization's circumstances.

REFERENCES

- Berkshire, J. R., & Higbland, R. W. Forced-choice performance rating—A methodological study. *Personnel Psychology*, 1953, 6, 355-378.
- Blanz, F., & Ghiselli, E. E. The mixed standard scale: A new rating system. *Personnel Psychology*, 1972, 25, 185-199.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.
- Hays, W. L. *Statistics for the social sciences*. New York: Holt, Rinehart & Winston, 1973.
- Kavanagh, M. J. The content issue in performance appraisal: A review. *Personnel Psychology*, 1971, 24, 653-668.
- Kavanagh, M. J., MacKinney, A. C., & Wolins, L. Issues in managerial performance: Multitrait-multimethod analyses of ratings. *Psychological Bulletin*, 1971, 75, 34-49.
- Klimoski, R. J., & London, M. Role of the rater in performance appraisal. *Journal of Applied Psychology*, 1974, 59, 445-451.
- Lawler, E. E. The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, 1967, 51, 369-381.
- Levine, J., & Butler, J. Lecture versus group discussion in changing behavior. *Journal of Applied Psychology*, 1952, 36, 29-33.
- Lifson, K. A. Errors in time-study judgments of industrial work pace. *Psychological Monographs*, 1953, 67(5, Whole No. 355).
- Rand, T. M., & Wexley, K. N. A demonstration of the Byrne similarity hypothesis in simulated employment interviews. *Psychological Reports*, 1975, 36, 535-544.
- Ronan, W. W., & Latham, G. P. The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology*, 1974, 6, 53-64.
- Ronan, W. W., & Prien, E. P. *Toward a criterion theory: A review of research and opinion*. Greensboro, N.C.: Creativity Research Institute, Smith Richardson Foundation, 1966.
- Ronan, W. W., & Prien, E. P. *Perspectives on the measurement of human performance*. New York: Appleton Century Crofts, 1971.
- Smith, P., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, 47, 149-155.
- Wallace, S. R. Criteria for what? *American Psychologist*, 1965, 20, 411-417.
- Wexley, K. N., & Nemeroff, W. F. Effects of racial prejudice, race of applicant, and biographical similarity on interviewer evaluations of job applicants. *Journal of Social and Behavioral Sciences*, 1974, 20, 66-78.
- Wexley, K. N., Sanders, R. E., & Yukl, G. A. Training interviewers to eliminate contrast effects in employment interviews. *Journal of Applied Psychology*, 1973, 57, 233-236.
- Wexley, K. N., Yukl, G. A., Kovacs, S. Z., & Sanders, R. E. Importance of contrast effects in employment interviews. *Journal of Applied Psychology*, 1972, 56, 45-48.

(Received December 9, 1974)