This is to certify that the

dissertation entitled

A Comparison of the Oral Interview and
Behavioral Consistency Evaluation Methods
for Selecting Job Applicants

presented by

Sally Adrienne Hildebrand Mc Attee

has been accepted towards fulfillment
of the requirements for

the Ph.D. degree in Measurement, Evaluation,
and Research Design

William A. Mehrens
Major professor

Date October 22, 1985

A COMPARISON OF THE ORAL INTERVIEW AND

BEHAVIORAL CONSISTENCY EVALUATION METHODS

FOR SELECTING JOB APPLICANTS

by

Sally Adrienne Hildebrand Mc Attee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational
Psychology, and Special Education

1985

ABSTRACT

A COMPARISON OF THE ORAL INTERVIEW AND
BEHAVIORAL CONSISTENCY EVALUATION METHODS
FOR SELECTING JOB APPLICANTS

By

Sally Adrienne Hildebrand Mc Attee

This study compared the oral and behavioral consistency
examination methods in the selection process for two mana-
gerial positions. The need for such a study arose from the
researcher's desire to find a testing method which possessed
the desirable characteristics of the oral interview but
which avoided its disadvantages. The behavioral consistency
approach was used as an alternative to the oral interview
because it is parallel in development, content, and
administration but involves no interaction between raters
and candidates.

For each position, test development for both approaches
was based on a job analysis which defined the essential job
dimensions. Test content was parallel. The behavioral
consistency examination asked candidates to describe major
achievements which demonstrated their capabilities in each
job dimension. The oral examination consisted of two
questions developed by subject matter experts for each job
dimension. There were 18 subjects in the first sample and

14 in the second.

The findings were as follows:

1. The results regarding the comparability of the two methods were inconclusive. Correlations between the methods were significant and meaningful for one sample but were non-significant for the other.

2. There were no significant differences in reliability between the two methods for either the overall ratings or the dimension ratings for either sample with one exception for the dimension ratings.

3. Convergent validity results were inconclusive. The methods demonstrated convergent validity for one sample but not for the other. The methods did not demonstrate discriminant validity for either sample.

4. There were no significant differences between the methods regarding their acceptability to the raters. However, based on descriptive comparisons, the behavioral consistency method was superior in terms of rater time.

5. Based on descriptive comparisons, time efficiency for the candidate was in favor of the oral examination. However, candidate time included only actual examination time; it did not include time for travel or preparation.

DEDICATION

to

Eric Mc Attee and Janice and Stuart Hildebrand

ACKNOWLEDGMENTS

There are many people who have provided guidance and support for this study or who have participated in carrying out the research. I would first like to thank my advisor, Dr. William Mehrens, who was my dissertation committee chairman and provided advice and support throughout the project. I would also like to thank the other members of my dissertation committee, Drs. Frederick Ignatovich, William Schmidt, and Neal Schmitt, for their suggestions at the proposal stage and their review of and comments on the dissertation.

The Personnel Department and Bureau of Sanitation of the City of Milwaukee participated in this project and gave me their unstinting cooperation and support. I would like to thank James P. Springer and Fagan D. Stackhouse of the Personnel Department and William Kappel, James Kosmatka, and Will Hudson, Jr. of the Bureau of Sanitation for their participation in the study. I would also like to thank other members of the two departments who participated in the study or who supported me in carrying it out. I would like to give special thanks to Timothy Keeley, Adrian Foster, and Stephen Smith of the Personnnel Department for their support in seeing me through the entire project.

TABLE OF CONTENTS

LIST OF TABLES

Chapter One

INTRODUCTION

Need

There has been a great deal of research literature on the employee selection interview. This research has been summarized in at least eight literature reviews, which have uniformly deplored the lack of reliability and validity evidence for oral interviews. However, recent studies have found reliability and validity under the following conditions: information about the job was available to raters, the interview was structured so that the same questions were asked of all applicants, behaviorally anchored rating scales were used, and raters received training prior to the examination process.

These recent results are encouraging because there is a need for oral examinations. For jobs with interpersonal and oral communication dimensions and for managerial jobs where there are relatively few candidates and technical knowledge and managerial skills must be assessed, the oral examination is the most practical selection technique. For the first type of job, the oral examination is the only content valid method available. For the second type, content validity can be achieved, and the oral examination is efficient compared to written tests and assessment centers. Further, it

usually is accepted by candidates and hiring agencies.

Although the appropriateness of a content validity rationale for many of the examples discussed in the literature would be debatable, content validity and fairness can be supported for an oral interview when it is developed and administered according to certain procedures.

However, in spite of their potential content validity and efficiency, oral examinations have some serious problems which have been examined in the literature. First, there is evidence that race, sex, attractiveness, and other non-job-related applicant characteristics can affect oral test scores. Second, ratings may be subject to a halo effect due to the applicant's general likeability or oral communications skill over and above its importance to the job. Third, certain rater characteristics such as degree of accountability, responsibility, or authoritarianism may affect the ratings. Fourth, certain situational character-istics such as the quality of the previous applicant or the timing of unfavorable information can affect the ratings. And fifth, it is logistically difficult to assemble the raters and candidates, who may come from different parts of the country, together on the same days. In the process the best raters and/or candidates are sometimes lost.

Because of these problems, it would be desirable to identify another testing method which possessed the desirable characteristics of the oral interview but which avoided its disadvantages. Few studies have compared the

oral to other types of examinations. However, a relatively new approach to assessing training and experience (the behavioral consistency approach) appears promising as an alternative to the oral examination for the following reasons:

1. The behavioral consistency approach has been based on job analysis.

2. It covers comparable job dimensions.

3. It is based on the logical rationale that past behavior is the best predictor of future behavior.

4. It is typically scored by means of behaviorally anchored rating scales.

5. It is not affected by the race, sex or attractiveness of the applicants.

6. Because it does not assess interpersonal and oral communications skills directly, it avoids halo effect due to those factors.

7. Because raters and candidates do not have to assemble in one place, it avoids the logistics problem.

The behavioral consistency method is parallel in development and content to the oral interview method. It is also similar in administration except that the presentation format is written rather than oral and there is no interaction between raters and candidates. Therefore, it seems a logical alternative to the oral interview.

Purpose

The primary purpose of this study was to compare the oral interview evaluation method with the behavioral consistency evaluation method. The need for such a study arose from the practical need to continue conducting oral examinations combined with the practical problems they present. Because hiring agencies are typically convinced of the validity of oral examinations, it would be infeasible to substitute a training and experience method without doing a comparative study.

A secondary purpose for the proposed study was to fulfill a need for research on oral examinations in an applied setting. Much of the oral interview research has been done in laboratory settings with perhaps unwarranted assumptions regarding their generalizability to applied settings. The types of raters used in such studies are typically students or recruiters, the types of ratees used are students or simulated applicants, the stimulus material often consists of videotapes, applications, or protocols (a type of resume with additional biographical information regarding the applicant), the rating outcome is a hiring decision or lack thereof, and in many of the studies the interview content is unspecified.

Some research has been done on the generalizability issues, but results have been inconsistent. More research studies need to be done with subject matter experts as raters, job applicants as ratees, oral interviews as the

stimulus material, a continuum of ratings as the outcome measure, and structured questions based on job-related dimensions as the interview content.

This type of research would be beneficial both for private companies using oral interviews to make individual hiring decisions and for governmental agencies using oral examinations to obtain comparison ratings on candidates.

<u>Research Questions</u>

This study was designed to answer the following research questions:

1. Do the oral and behavioral consistency examination methods provide comparable ratings of applicants for employment?

2. Are the oral and behavioral consistency methods equally reliable?

3. Do the oral and behavioral consistency methods demonstrate convergent and discriminant validity?

4. Are the oral and behavioral consistency methods equally acceptable to raters and equally efficient in terms of rater time?

5. Are the oral and behavioral consistency methods equally efficient in terms of candidate time?

Chapter Two

REVIEW OF THE LITERATURE

Oral Interview Literature

Introduction

There has been a great deal of research literature on the oral interview. This research has been summarized in at least eight reviews, the earliest a 1949 review by Wagner and the most recent a 1982 review by Arvey and Campion. These reviews (Arvey and Campion, 1982; Carlson, Thayer, Mayfield and Peterson, 1971; Dunnette, 1962; Mayfield, 1964; Schmitt, 1976; Ulrich and Trumbo, 1965; Wagner, 1949; and Wright, 1969) have uniformly deplored the lack of reliability and validity evidence for oral interviews and have urged additional research, including research which compares the oral to other selection methods. Although there have been many negative research findings on oral interviews, some recent studies have found reliability and validity under the following conditions:

1. The raters receive information about the job.

2. The interview is structured so that each candidate is asked the same questions.

3. There are behaviorally anchored rating scales so that each rater is using the same definition for each rating level.

4. The raters receive training before beginning the interviewing process.

Because many of the later studies have investigated the effect of these factors and found them to improve reliability and validity, the later reviews have been more positive regarding the use of oral examinations.

This literature review is focused on research studies of the oral examination characteristics leading to reliability and validity. The model proposed by Schmitt in his 1976 review of the oral interview literature is used to summarize research on rater, applicant, and situational characteristics. This review concludes with a discussion of studies comparing the oral examination to other examination methods.

Examination Characteristics

There has been a group of studies which has investigated the effect of how the oral examination has been conducted. Interviews have varied in how much information about the job was given to the raters, how structured the questions were, what type of questions were used, what type of rating scale was used, and what type of training was received by the raters.

Job information. Studies investigating the effect of job information have found that giving the raters a job description improves the reliability and validity of the oral examination. Langdale and Weitz, in a 1973 study, found that raters who received a complete job description

had higher interrater reliability and discriminated more among the candidates than raters who received only a job title. Similarly, Rothstein and Jackson (1980) found that raters were able to discriminate more accurately between congruent and incongruent applicants (whose characteristics matched or did not match the job) when they received job descriptions than when they received only job labels.

In a related study, Wiener and Schneiderman (1974) found that raters focused on relevant applicant information when job information was available. Although the effect of irrelevant applicant information was not removed, its effect was stronger when no job information was available. In addition, Leonard (1974) found that when raters were given job descriptions, interrater reliability was higher for ratings of job relevant factors than for irrelevant factors.

Finally, a study by Osburn, Timmreck, and Bigby (1981) investigated the effect of providing job-related dimensions to the raters. Two applicants, each of whom was well-suited to one of two jobs, were rated accurately by raters who had access to specific and relevant job dimensions and inaccurately by raters who had access to only general job dimensions.

The above studies show quite clearly the positive effect of relevant job information. The literature also shows that raters with job knowledge tend to rate the same applicant information as important.

The Langdale and Weitz study showed that there was

substantial consistency among raters in rating the
importance of applicant information regardless of the amount
of job information they possessed. However, the two rater
groups (one with job information and one without) did differ
significantly in their ratings with uninformed raters giving
lower ratings on item importance. Hakel, Dobmeyer, and
Dunnette (1970) showed that rater groups which differed on
their knowledge of the job (students and interviewers)
differed on which content dimension they considered most
important. However, within groups the raters agreed on
content importance, content importance determined the
ratings, and the effect of favorable information depended on
content importance.

In a related study, Valenzi and Andrews (1973) found
that raters, all of whom had job descriptions, differed in
applicant cue use (how they weighed different types of
applicant information) and their ratings of applicants
differed according to these weights. However, the same
raters agreed in theory on cue importance. Differences in
candidate ratings were due to the inability of the raters to
apply their own rating strategies to the actual rating
situation.

As a whole, these studies provide excellent support for
the conclusion that an oral interview developed according to
a content validity model would indeed be valid. The first
group of studies cited provides support for the idea that
oral examinations which are based on job analysis and which

define the job dimensions to be evaluated will be reliable whereas interviews not so based will not be reliable. This is only logical and indeed it is not surprising that raters without a job description would define the job for themselves differently and would therefore be assessing the candidates on different factors.

The second group of studies shows that raters with job knowledge tend to rate the same dimensions as important. The inconsistent ratings of the Valenzi and Andrews' subjects, all of whom had job descriptions, were not due to disagreement on content importance but to the inability of the raters to apply their rating strategies correctly.

Interview structure. A second group of studies has investigated the effect of interview structure (asking all candidates the same questions).

In support of this approach, studies done by Reynolds (1979) and Mayfield, Brown, and Hampstra (1980) showed moderate-high interrater reliability for structured oral interviews used to select police officers and insurance agents. Latham, Saari, Pursell, and Campion (1980) demonstrated both reliability and validity for a structured situational interview. They conducted three studies using samples of hourly workers, foremen, and entry-level workers. Results of the studies showed moderate-high interrater and internal consistency reliability and concurrent and predictive validity. Additional studies of the structured situational interview by these researchers (Latham and

Saari, 1984) also showed moderate to high internal
consistency and interrater reliability and concurrent
validity.

On a logical basis, one would conclude that structure
would be bound to increase interrater reliability since one
source of variation, the questions, has been removed. In
support of this contention, Schwab and Heneman (1969)
investigated the effect of interview structure on interrater
reliability and found that degree of structure corresponded
to degree of interrater reliability for the position of
clerk-stenographer. In a similar study, Janz (1982)
investigated a patterned behavioral interview and concluded
that this type of structured interview was more valid but
less reliable than unstructured interviews. This study also
found that content differed for the two formats with
unstructured interviews focusing on credentials and
self-perception content and structured interviews focusing
on behavior descriptions. This difference in content was
largely due to the differences in training received by the
interviewers. Interviewers using the unstructured format
were trained in establishing rapport and control while those
using the structured format were trained in specific
behavioral description techniques.

In spite of these positive findings and the rationale
supporting the use of structured interviews as a way of at
least increasing interrater reliability, several studies
have had negative results. A 1971 study by Hakel showed

that interrater reliability was low to moderate even with highly structured interviews. A follow-up study by Heneman, Schwab, Huett, and Ford (1975) showed low interrater reliability and low validity for both structured and unstructured interviews for social worker jobs.

Finally, in an unpublished study, Davey (1984) investigated the effect of highly structured interviews on interview validity. He concluded that high interrater reliability does not necessarily correspond to high validity. Trained oral panels with high within-panel interrater reliability (.95 or greater) who used the same structured interview differed significantly across panels in the validity of their ratings. This study was done on a structured oral examination for State Police Trooper, which had six oral interview panels, each interviewing over 100 candidates. A high degree of structure was achieved by standardized panel training and videotape practice; standardized questions, factors and scales; and examples of good and poor responses. Oral examination ratings correlated with police academy rank .23 for 104 graduates. However, even though sample sizes were small, there were highly significant differences between the validities of individual panels, which ranged from -.04 to .79.

Although the majority of these studies support the conclusion that structured orals are at least more reliable than non-structured orals, the contradictory results obtained in the Davey, Hakel, and Heneman et al. studies are

puzzling and disturbing. Heneman et al. offered the suggestion that their negative findings may have been due to the lack of behaviorally anchored rating scales while Davey has hypothesized that high within-panel agreement was achieved by the interaction of the members throughout the process, a factor which was absent across panels. Whatever the cause of these inconsistent results, it must be concluded that structured orals are preferable to unstructured orals since different questions are a potentially undesirable source of variation in a oral examination.

Question type. Two studies have investigated the topic of question type. Latham and Saari (1984) developed an interview with both situational questions, which required applicants to state how they would behave in hypothetical situations, and questions regarding applicants' past experiences. The situational question ratings correlated significantly with job performance while the past experience question ratings did not.

The positive results for situational interviews are supported by the previous findings of Latham et al. (1980). However, any conclusions on the relative effectiveness of various question types are only tentative because of the limited evidence. Also the low validity for the past experience questions may have been affected by the small number of such questions compared to the number of situational questions.

Tengler and Jablin (1983) focused on different aspects

of question type. They investigated interviewers' use of open versus closed-ended questions and primary (introducing new topics) versus secondary (probing previously-introduced topics) questions. The researchers concluded that applicant responses were longer for open-ended and secondary questions, open-ended and secondary questions occurred mainly during the later parts of the interviews, and there was no relationship between question type and whether applicants were offered second interviews.

Results of the Tengler and Jablin study become moot in the case of structured interviews because interviewers have no leaway in the types of questions asked. Although results of the Latham and Saari study are more relevant to this research, any conclusions regarding the superiority of situational to past experience questions must be considered tentative. Question type was not a part of this research design.

Rating Method. A third group of studies investigated the effect of rating method on oral examination quality. In spite of the conflicting results cited above on the use of structured interviews, the studies of rating method effect for oral examinations give more consistently positive results, supporting the theory underlying the use of behaviorally anchored rating scales as well as showing the increased reliability and validity of such scales.

Two studies, while not providing direct support of the use of behaviorally anchored rating scales, provide

support for the theory behind their use. In a 1963 study,
Rowe showed that there are significant between-rater
differences and within-rater consistencies in where raters
set passing points. That is, raters differ among each other
but are consistent as individuals in how high a standard
they set for passing applicants. Different job standards
affected where passing points were set. This study was
conducted under the condition that passing and failing
categories were undefined for the raters. If such
categories were defined for the raters, then it follows that
the between-individual differences should decrease and the
within-individual consistencies could become an advantage in
overall consistency.

A second study (London and Hakel, 1974) showed that
ratings are affected by ideal applicant stereotypes held by
raters (whether the ideal applicant was well or not well-
qualified). It follows from this finding that if an ideal,
average, and unsatisfactory applicant stereotype were based
on job analysis and were provided to the raters through
behaviorally anchored rating scales, interrater reliability,
and validity should be improved.

There have been at least five studies which
demonstrated the increase in test quality with specifically
anchored rating scales. Maas (1965) demonstrated that a
scaled expectation rating method had significantly higher
interrater reliability than an adjective rating scale.
Results further suggested the use of rating panels rather

than individual raters. This would eliminate both question and candidate inconsistencies, which occur when different raters ask questions at different times and when candidates appear before different raters at different times.

The 1980 study by Latham, Saari, Pursell, and Campion and the 1984 study by Latham and Saari showed adequate interrater and internal consistency reliability and concurrent validity (and predictive validity for the 1980 study) for situational interviews, which were characterized by questions based on critical incidents and behavioral statements as benchmarks. And, in an investigation of behavioral versus graphic rating scales, Vance, Kuhnert and Farr (1978) found higher interrater reliability and accuracy for behavioral than for graphic rating scales. There were also more halo and leniency errors for the graphic scales. (A halo error occurs when an applicant's rating on one dimension unduly influences his or her ratings on the others. A leniency error occurs when most of the applicants are given high or low ratings.) Finally, Fay and Latham (1982) showed that two types of behaviorally anchored rating scales were less subject to rating errors of contrast and first impression than trait based scales. However, they were not less subject to halo errors. (A contrast error occurs when an applicant's rating is influenced by those of the preceding applicants. A first impression error occurs when an applicant's rating is based primarily on the first few minutes of the interview.) One of the behavioral scales

anchored rating scales had not demonstrated greater interrater reliability, greater discrimination among ratees, or fewer halo and leniency errors than other types of scales. These researchers suggested that behaviorally anchored rating scales may not be worth the time and effort necessary for development if the above criteria are paramount.

These conclusions from the performance evaluation literature substantially weaken the argument for including behaviorally anchored rating scales in a plan to strength oral examination reliability and validity. However, oral raters may benefit more from their use than performance evaluation raters since they usually know less about both the job and the ratee than performance evaluation raters. Also, research on the use of behaviorally anchored rating scales in oral interviews is certainly more directly applicable to additional research on oral interviews than is similar research from the performance evaluation literature.

For these reasons, the use of behaviorally anchored rating scales in oral examinations still seems to have potential as a way of increasing reliability and validity, and it seems reasonable to include them in the design for a content valid oral examination.

Rater training. Rater training will be discussed as an examination characteristic since it is part of the test administration process. Research on the other rater effects will be summarized later in this review.

Most of the studies of rater training have focused on the reduction of rating errors with the majority finding that training was successful in the reduction of such errors.

Several studies have compared training methods in their ability to reduce errors. In a 1973 study, Wexley, Sanders, and Yukl found that combining warnings regarding the errors with anchoring the rating points failed to eliminate contrast errors (the effect of the previous applicant). However, a workshop eliminated this type of error. A comparative study by Latham, Wexley, and Pursell in 1975 showed that a control group committed similarity, contrast and halo errors, a discussion group committed first impression errors, and a workshop group committed none of these types of errors.

In a 1981 study, Ivancevich and Smith found that training with role playing and either videotape or lecture was superior to no training in a goal setting situation. Finally, Fay and Latham in a 1982 study concluded that training reduced rating errors significantly regardless of the type of rating scale used, behavioral or trait.

There was only one contradictory study in this group of training studies: a study by Vance, Kuhnert, and Farr (1978), which concluded that training had no effect on rating errors with the use of either behavioral or graphic scales. A possible explanation for this conflicting result concerns length and intensity of training. The training

program investigated in the Vance et al. study was minimal in length and involvement while those in the studies previously cited were longer and more intensive. It is likely that a minimum amount of time and involvement is necessary for any training program to be effective.

In contrast to the above studies, which focused on rater error, Pulakos (1984) investigated the differential effectiveness of rater training programs focusing on error, accuracy, and both error and accuracy. A no training condition was also included in the study. Findings were as follows: the most accurate ratings corresponded to accuracy training while the least accurate corresponded to no training, less leniency corresponded to accuracy training and error/accuracy training, less halo corresponded to error training and error/accuracy training, and the effectiveness of training differed across dimensions.

The results of the research studies on rater training clearly show its effectiveness in increasing rater accuracy and in reducing rating errors and support its inclusion in a test administration program. Given the results of the Pulakos study, a focus on accuracy rather than error training seems appropriate.

Summary of examination characteristics results. In spite of some contradictory findings concerning the effect of interview structure and type of rating scale, the results of the research literature on the effect of examination characteristics seem to warrant the following conclusions:

that oral examinations are more reliable and valid when the

raters have detailed information about the job, when the

same questions are asked of all applicants, when there

are definitions for each rating point along the rating

continuum, and when rater training is provided. Question

type was not included in this research design.

Rater, Applicant, and Situational Characteristics

As the model proposed by Schmitt in his 1976 review of

the literature suggested, characteristics of the oral

examination itself are only one group of factors affecting

oral examination quality. Other major factors in oral

examination ratings are the effects due to the raters, the

effects due to applicants (on non-job-related factors), and

the effects due to situational factors. Findings regarding

these factors will be presented because they are major

sources of unreliability and invalidity in oral examina-

tions. Job-related applicant effects are of course positive

since examinations are designed to assess the applicants on

those factors.

Rater characteristics. Studies of rater

characteristics and their effect on orals have included

investigations of rater experience, rater selective

attention ability and memory demand, rater accountability,

rater authoritarianism and prejudice, and rater sex.

The findings of studies investigating rater experience

have not been consistent, with two investigations showing

differences between experienced and inexperienced raters and

five studies showing no differences. Rowe (1963) in a study of armed forces personnel showed that rank, which was used as an indicator of length of experience, determined where passing standards were set. Hakel, Dobmeyer, and Dunnette (1970) found that the relative importance of content categories differed between rater groups comprised of students and interviewers.

In contrast, Langdale and Weitz (1973) found no differences between experienced and inexperienced interviewers in the use of job descriptions and in ratings of dimension importance. Wiener and Schneiderman (1974) also found that experienced and inexperienced interviewers did not differ in use of relevant or irrelevant job information although experienced interviewers tended to reject applicants oftener. In a 1974 study, Moore and Lee found no difference between interviewers and managerial groups in rating errors, and Heneman et al.(1975) found no differences in the ratings of students and social worker subject matter experts in reliability and validity. Finally, Mullins (1982) found that the ratings of students and experienced interviewers were comparable.

Cardy and Kehoe (1984) investigated the relationship between rater accuracy, selective attention ability, and memory demand. They distinguished between raters who were field-independent (inclined to perceive things analytically) and field-dependent (inclined to perceive things holisti-cally). The researchers found that field-independent raters
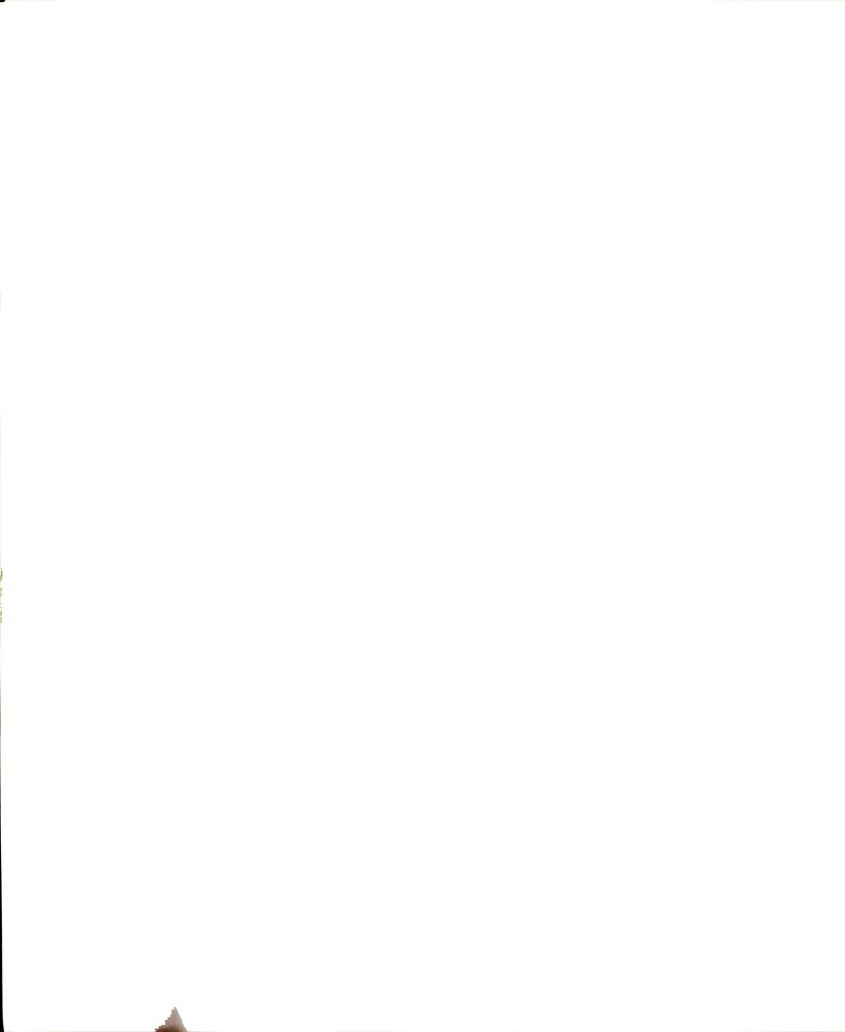
were more accurate although field-independency accounted for only a small part of the variance in rater accuracy. They also found that memory demand affected accuracy of ratings (with high memory demand corresponding to lower accuracy) but did not interact with selective attention ability. However, memory demand is not an issue in oral examinations if ratings are made immediately after the candidate is interviewed.

A single study has been done on the important issue of rater accountability and responsibility. Rozelle and Baxter (1981) found that under the condition of high accountability and responsibility, there was high interjudge agreement and low within-judge overlap (similar ratings by one judge of several applicants) in contrast to the low accountability and responsibility condition.

The subject of rater authoritarianism and prejudice has likewise received surprisingly little attention. Two studies have been done which linked rater authoritarianism and prejudice with applicant race and sex. Simas and McCarrey (1979) found that high authoritarian raters of both sexes rated males higher than females and made more job offers to males than did low authoritarian raters. In a 1982 study, Mullins investigated the impact of rater prejudice (measured by an attitude inventory), applicant race, and applicant quality on ratings and found that while high quality applicants were rated high regardless of race, marginally qualified blacks were rated better than

marginally qualified whites. Prejudiced raters rated blacks higher.

A final set of studies has investigated the effect of rater sex. Ferris and Gilmore (1977) found that male raters gave higher ratings than female raters. In contrast, Parsons and Lidden (1984) found that female raters gave higher ratings than males but that this effect was inconsequential compared to the effect of non-verbal applicant behavior. Interviewer sex also interacted with applicant non-verbal behavior with female interviewers giving higher ratings of non-verbal cues. Two other studies showed the interaction of rater sex and applicant non-verbal language and attractiveness. Sterrett (1978) found that men and women raters differed in their ratings of applicants with differing body language, and Baron (1983) found that males gave lower ratings to scented applicants while females gave them higher ratings.

Summary of rater characteristics results. Although the results of the studies investigating rater character-istics are not completely consistent, the following conclu-sions can be drawn: the effects of rater experience and selective attention ability are negligible, the effect of memory is substantial but not an issue, raters who are accountable and responsible produce more reliable ratings, raters who are authoritarian or prejudiced favor males and blacks, and male and female raters differ in their responses to non-verbal communication. As a whole these results are

encouraging because these factors can potentially be controlled in the test development process. Of the six factors discussed, experience, selective attention ability, and memory are not concerns, and raters can be chosen so that they are accountable and responsible. Although the effects of rater authoritarianism and prejudice and sex are disturbing, it is probable that they could be alleviated by proper training.

Applicant characteristics. In addition to characteristics of the interview and the rater, another important component of the oral interview model is applicant characteristics. Studies investigating the effect of non-job-related applicant characteristics upon oral examinations have dealt with the following topics: applicant race, sex, age, attractiveness, similarity to the rater, non-verbal factors, motivation and anxiety, and training.

Considering the importance of race effects, there has been a paucity of studies concerned with applicant race. Five studies have been done. Rand and Wexley (1975) found no significant effect for race while McDonald and Hakel (1985) found significant but inconsequential effects. Parsons and Lidden (1984) found some race effects on applicant ratings with whites receiving higher ratings, but these were inconsequential compared to non-verbal effects. Race was also related to non-verbal behavior with whites receiving higher ratings on non-verbal cues than blacks. In contrast, Mc Intyre, Moberg, and Posner (1980)

and Mullins (1982) found preferential treatment for blacks
over whites. The Mullins study, cited earlier in this
paper, showed that high quality applicants were highly rated
regardless of race but that marginally performing blacks
were rated better than comparably performing whites.

In contrast to the dearth of studies on race, there
has been a multiplicity of studies on applicant sex.
Several studies have shown that raters have a general
tendency to prefer males. In a 1975 study, Dipboye,
Fromkin, and Wiback found that both professional
interviewers and students preferred males to females. This
finding was confirmed by two other studies: McIntyre et al.
(1980) and Cann, Siegfried, and Pearce (1981) both found
that male applicants were preferred over female applicants.

The remainder of the applicant sex studies were
concerned with the interaction of applicant sex with some
other factor. Simas and McCarrey (1979) found that high
authoritarian personnel officers of both sexes rated males
more favorably than females and made more job offers to
males.

Several researchers studied the interaction of sex
with type of job. A study by Heilman (1980) showed that
women were rated lower by male and female raters when
females comprised less than twenty-five percent of the
applicant pool. Heilman concluded that the effect of sex
was mediated by the degree to which sex stereotypes
operated. In similar studies, Cash, Gillen, and Burns

(1977) found that males were rated more favorably for masculine jobs and females were rated more favorably for female jobs, and Cohen and Bunker (1975) found that males were preferred for a male-oriented position and females were preferred for a female-oriented position. Finally, Heilman and Saruwatari (1979) found that men were preferred for managerial positions while women were preferred for non-managerial positions. They further found that attractive-ness was a third interacting factor. Attractiveness was an advantage for men regardless of type of job while it was a disadvantage for women seeking managerial positions and an advantage for women seeking non-managerial positions.

A recent study by Forsythe, Drake, and Cox (1985) extended the investigation of sex effects to women's clothing. Results of their study showed that women wearing more masculine clothing received higher ratings when applying for managerial jobs than women wearing more feminine clothing.

Another factor considered by researchers to have a possible interaction with applicant sex was the strength of an organization's employment policy. The results of this study were puzzling. Rosen and Mericle (1979) found that the strength of the employment policy and sex of applicant had no effect on the hiring decision, but females received lower starting salaries from companies with strong employment policies.

This review of the literature on applicant sex effects

concludes with three studies which contradict the consistent results cited previously. Ferris and Gilmore (1977) found that a female applicant received higher overall favorability ratings from student raters than a male applicant. However, this study had a relatively high alpha level (.10). Parsons and Lidden (1984) found some sex effects on applicant ratings with females receiving higher ratings, but these were inconsequential compared to non-verbal effects. Sex was also related to non-verbal behavior with females receiving higher ratings on non-verbal behavior than males. However, the preference for females in this case may have been due to the nature of the jobs, which were in an amusement park and may have demanded stereotypical feminine traits. A last study by McDonald and Hakel (1985) found significant but inconsequential sex effects in a study involving student raters.

There was one study on applicant age effects. Rosen and Jerdee (1976) investigated the interaction of age of applicant with sex of rater. They concluded that older employees were judged less reactive, more cautious, less physically capable, less interested in technology, and less trainable than younger workers by both male and female raters.

There have been numerous studies on applicant attractiveness, all concluding that attractiveness affects ratings positively although it may interact with applicant sex. Dipboye et al. (1975) found that professional

interviewers and students both preferred attractive to unattractive applicants. The results were less strong in a study by Cash et al.(1977). Study results indicated that attractive applicants were preferred for in-role and neutral jobs but only on one of three criteria, ratings of qualifications. The other criteria were ratings of success expectancy and hiring recommendations. As mentioned earlier, Heilman and Saruwatari (1979) found that attractiveness interacted with sex; it was an advantage for men but only for women seeking non-managerial positions. Cann et al. (1981) found that male and attractive applicants were preferred.

In contrast to these findings, Carlson (1967) found no effect for appearance. However, when both appearance and job-related factors were complimentary, there was an additional component in the ratings greater than that contributed by the separate ratings alone. A final study of attractiveness was conducted by Baron (1983). Study results were that rater sex and use of scent interacted in the ratings of applicants on job-related dimensions and personal characteristics. Males assigned lower ratings to scented applicants; females assigned higher ratings to scented applicants.

Several researchers have studied the effect of similarity of the applicant to the rater. Results from these studies have been mixed. Frank and Hackman (1975) found that the effect of similarity varied according to the

rater. There was no general similarity effect. However, Baskett (1973) found that while applicant competency influenced the hiring decision and the salary offered, similarity also affected the salary. Rand and Wexley (1975) also found a significant effect for similarity.

The effect of non-verbal factors on interview ratings has also been the subject of several studies. Washburn and Hakel (1973) and Imada and Hakel (1977) both found a significant effect due to non-verbal applicant communication. Sterrett (1978), cited earlier, found that men and women differed in their ratings of applicants with different body language. Women interpreted high intensity body language to mean low ambition while men interpreted low intensity body language as meaning low ambition. Since this study confounded body language with attractiveness, these results may not be meaningful. Finally, Parsons and Lidden (1984) found that non-verbal cues were highly correlated with applicant qualification ratings and that they accounted for a large part of the variance in applicant ratings compared to objective biographical information or applicant race or sex or rater sex. There were also significant relationships between non-verbal cues and applicant sex, applicant race, and interviewer sex with females and whites receiving higher ratings and female interviewers giving higher ratings. The researchers also investigated the relative contributions of various types of non-verbal cues and found that speech characteristic cues accounted for most

of the variance in qualification ratings whereas personal appearance cues accounted for little or none of the variance after speech characteristics were taken into account.

Results of two studies contradict these previously cited findings. Hollandsworth, Kazelskis, Stevens, and Dressel (1979) concluded that non-verbal behavior was unimportant compared to content, fluency, and composure on the part of the applicant, and Rasmussen (1984) concluded that non-verbal behavior was unimportant compared to resume credentials and verbal content. Rasmussen also found an interaction between verbal content and non-verbal behavior, with an effect for non-verbal behavior when verbal content was high but not when verbal content was low.

A final group of studies on applicant effect has to do with self-esteem, anxiety, motivation, and training of the applicant. King and Manaster (1977) found that self-esteem and body satisfaction of applicants had no effect on interview ratings they received. Keenan (1978) found that anxiety had no effect on the ratings but that there was a motivation effect with intermediate motivation on the part of applicants having a positive effect on their ratings. Applicants were more confident of success when they were highly motivated and liked the interviewer.

Finally, studies of applicant training in interviewing skills were conducted by Barbee and Keil (1973) and Hollandsworth, Dressel, and Stevens (1977). According to the Barbee and Keil study, a combined treatment of videotape

feedback and behavior modification improved applicants' interview ratings over the only videotape condition or the control condition. Hollandsworth et al. (1977) investigated behavioral training versus group discussion and found that interviewees from the discussion group increased their speaking times and were superior to those from the behavioral and non-trained groups in explaining their skills and opinions.

Summary of applicant characteristics results. Studies of applicant effects lead to the following general conclusions. The effect due to applicant race is unclear, there is an applicant sex effect with males preferred, there is an applicant age effect with younger applicants preferred, there is an applicant attractivenesss effect with attractive applicants preferred except for females desiring non-traditional jobs, there is an applicant similarity to rater effect with similar applicants preferred, and there is a non-verbal communication effect with vivacious applicants preferred. There is limited evidence on the effects of applicant self-esteem, anxiety, and motivation, but training improves applicant performance.

If the desired outcome of the oral interview process is to rate all applicants fairly based on their demonstrated skills on job-related dimensions, the results of these studies on applicant effects provide reason to doubt its effectiveness. However, many of these studies suffered from problems of poor test development. If the raters were not

given a job description, if the job dimensions were undefined and if there were no behavioral anchors for the rating scales, then the Wiener and Schneiderman study (1974) suggests that ratings are more liable to be based on irrelevant information such as race and sex. Their study showed that when job information was provided, the irrelevant applicant information, while still having an effect, had considerably less effect than the relevant information.

Situational characteristics. Situational characteristics comprise the final major factor affecting interview results. Studies investigating the effect of situational characteristics upon oral examinations have dealt with the following topics: primacy/recency effects, contrast effects, typical expected applicant effects, and interview length effects.

Primacy effects refer to the predominance of information presented early in the interview while recency effects refer to the predominance of information presented late in the interview. Most of the primacy/recency studies have also investigated the interaction between information favorability and its timing. Results of these studies have been inconsistent, with some studies finding primacy effects, some finding recency effects and some finding both.

Studies by Bolster and Springbett (1961), Blakeney and Mac Naughton (1971), and Tucker and Rowe (1979) all found primacy effects. Tucker and Rowe provided the clearest

demonstration of primacy effects in their investigation of reference letters. Applicants with negative reference letters were given less credit for past successes and more blame for past failures than applicants with positive reference letters. This study also showed a greater primacy effect for negative information, with negative reference letters penalizing applicants more than positive reference letters benefitted them. Bolster and Springbett found primacy effects for the first piece of inconsistent information, especially when the information was negative. Blakeney and Mac Naughton found primacy effects for negative information but concluded that these effects were not meaningful, accounting for only a small percentage of the variance of the final results.

In contrast to these studies, investigations by London and Hakel (1974) and Okanes and Tschirgi (1978) found recency effects. London and Hakel found recency effects for unfavorable information, and, in direct contrast to the Tucker and Rowe study, Okanes and Tshirgi found that initial ratings based on other materials were changed with the addition of an interview.

Several studies found both recency and primacy effects, depending on the consistency of information. Carlson (1971) found no recency or primacy effects for consistently favorable information but primacy effects for consistently unfavorable information and recency effects for inconsistent unfavorable information. Farr (1973) found recency effects

with repeated judgments but no effect with single judgments except for a primacy effect in one condition. Farr and York (1975) found recency effects for repeated judgments and primacy effects for a single judgment.

A second set of studies on situational characteristics investigated contrast effects. These effects have to do with the effect of the previous applicants on ratings of each following applicant. Wexley, Yukl, Kovacs, and Sanders (1972) found signficant contrast effects. While this accounted for only a small part of the variance for strong or weak applicants, it accounted for a large part of the variance for average applicants. In addition, an investigation by Heneman et al. (1975) found that interviewee order affected ratings while amount of interview structure and biographical data did not.

In contrast, Landy and Bates (1973) found no contrast effects in two studies, and Wexley, Sanders, and Yukl (1973) found that a rater training workshop eliminated contrast errors. Finally, several studies (Carlson, 1970; Hakel, Ohnesorge, and Dunnette, 1970; and Kopelman, 1975) found contrast effects but also found that they accounted for an insignificant portion of the variance of ratings. Kopelman also found that contrast effects were most influential for candidates of average performance. This coincides with the Wexley finding cited earlier.

There has been a single study on the effect of the typical expected applicant. London and Hakel (1974) found

no main effect on ratings for the level of the typical expected applicant. However, there was an interaction with information favorability. A high caliber typical expected applicant led to a better rating of unfavorable information.

A final group of studies on situational effects had to do with interview length. Three studies have investigated this topic. Anderson (1960) found that applicants talked the same amount of time and interview length was constant regardless of the hiring decision, but that interviewers spoke at greater length with accepted applicants. Tullar, Mullins and Caldwell (1979) found that raters took more time to make their rating decisions with high quality applicants and that they took more time to decide when interviews were expected to last longer. Tengler and Jablin (1983) found that applicants who were offered second interviews differed on a composite measure of interview response time from those who were not offered second interviews. The successful applicants spent less time answering questions but more time talking than the unsuccessful applicants. However, these differences were not significant.

Summary of the situational characteristics results.
Results of the situational research literature cannot be summarized easily because they are inconsistent. However, several of the contrast effect studies and one of the primacy/recency effect studies showed that, while these effects were present, they did not account for a meaningful portion of the variance of ratings. Also, there was no main

effect due to the typical expected applicant and no clear
conclusion regarding differences in speaking or interview
times of successful and unsuccessful applicants. Based on
the above analysis, it seems likely that situational effects
are not meaningful and, unlike applicant effects, do not
pose a serious threat to oral interview validity.
Furthermore, the situational effect studies have in general
not described the test development process. As pointed out
in the discussion of applicant effects, if the interview
were not based on job analysis, if job dimensions were not
defined for the raters, if behaviorally anchored rating
scales were not provided, and if raters received no
training, then the Wiener and Schneiderman study (1974)
suggests that ratings are liable to be based on irrelevan-
cies such as situational characteristics. Finally, the
Wexley et al. study (1973) demonstrated the effectiveness of
rater training in eliminating contrast errors.

Comparison to Other Methods

Research comparing the oral interview to other testing
methods has been extremely limited. Tubiana and Ben Shakhar
(1982) compared the results of an objective questionnaire to
results of an interview assessing personality factors for
Israeli army officers and found that the two methods were
comparable. In a similar study, James, Campbell, and
Lovegrove (1984) compared the results of a personality test
to the results of an oral interview assessing suitability.
One of the personality test scales (social conformity)

correlated moderately with the oral suitability score for men; however, there was no relationship between the personality test and oral interview for women.

Several studies, while not providing direct evidence regarding comparison of the oral interview with other approaches, have contributed related information by trying to get at the issue of generalizability. Unfortunately, the results of these studies are inconsistent. Moore and Lee (1974) found no differences between live and videotaped interviews, and Ferris and Gilmore (1977) showed that there were no differences between ratings of resumes, videotapes or audiotapes. On the other hand, Imada and Hakel (1977) found significant differences along a rater proximity continuum (whether the rater observed a videotape, observed an interview or was him or herself the interviewer). Also, Washburn and Hakel (1973) found that there were differences due to whether the presentation mode was audiovisual, visual or transcript. Finally, a recent study by Ricchiute (1985) found that mode of presentation (visual, auditory, and visual/auditory) and task importance affected the decision-making of auditors.

The studies presented above are not only inconsistent in their findings but are not completely relevant in that the content of the oral interview was undefined, the interview was based on personality characteristics rather than job dimensions, or the oral presentation mode did not consist of an interview. Therefore, the need for a study to

assess the comparability of the oral to other testing
methods which are intended to assess job-related character-
istics of applicants has been confirmed.

An issue related to the relationship between the oral
interview and other testing methods is whether oral
interviewers should have access to application materials and
if so what effect these materials have on the ratings. A
study by Tucker and Rowe (1977) supports the use of the
application in its finding that applications increased the
amount of relevant information the raters had. However, a
second study by the same authors (Tucker and Rowe, 1979)
found that raters were liable to develop unfavorable
expectations of a candidate when presented with an
unfavorable reference letter and were then liable to give
the applicant less credit for past successes and hold the
applicant more personally responsible for past failures.
They were also more likely to reject these applicants for
jobs since they attributed their failures to internal
rather than external reasons. A negative application effect
was also shown by Dipboye, Fontenelle, and Garner (1984).
These researchers found that raters without applications
made more reliable ratings of applicants' fit to the job and
interview performance. Raters with applications gathered
more correct information but were more variable in
information gathering.

Several studies have produced results counter to the
above, showing that the effect of the interview is liable to

prevail over the effect of prior information. Carlson (1971) showed that the interview changed the effect of valid test information. Heneman et al. (1975) showed that biographical information had no effect on interview results. Similarly, Okanes and Tschirgi (1978) found that judgments made before interviewing based on application information shifted significantly due to the interview. Most of the ratings shifted either up or down from the neutral category, but half of the positive recommendations were changed to either the neutral or low category. The low category had the fewest number of changes. Based upon the results of four studies, Sackett (1982) concluded that interviewer decisions are not based upon previous hypotheses about applicants; that is, interviewers do not try to confirm their initial hypotheses about applicants based on previous information. In a final study, McDonald and Hakel (1985) found that raters did not select questions based on previous resume judgments and that resume effect was small compared to interview effect.

Although the majority of these studies showed lack of effect for application materials, it seems reasonable to conclude that the application should not be a part of most oral interviews intended to be job-related ranking devices. The interview should be planned to assess the most important and relevant job dimensions, thus eliminating the need for further information that was suggested by the 1977 Tucker and Rowe study. If application information is indeed

relevant, it should probably be assessed separately by different raters so that the dimensions to be assessed by the oral are not contaminated. Since the results of the other studies show either that application materials have a negative effect on the oral or that they have no effect, they support this conclusion.

Training and Experience Evaluation Literature

There is little research on the methods of rating training and experience questionnaires or job applications although employers have been using them for years. Five major methods have been used to rate the education and experience of job candidates. These have been discussed in three articles by Ash and Levine (Ash, 1984; Ash, 1983; and Ash and Levine, 1982) and will be summarized below.

Method

The point method. When the point method is used, each applicant's education and experience is compared to certain previously specified requirements that the applicant must meet to be considered. If the applicant meets the minimum requirements then he or she receives additional points for any experience he or she has in addition to the amount required. The extra experience or education must be of equal quality to that described by the requirements.

The grouping method. When the grouping method is used, raters make a holistic judgment about the quality of the applicants' experience and education and place them in groups according to quality. A similar method, not

discussed separately here, is the holistic method. Raters make a holistic judgment about each candidate but then place them in rank order according to quality or give them a numerical rating according to quality.

The task-based method. When the task-based method is used, applicants are asked to complete a form listing the tasks that are done on the job. The applicants indicate their degree of experience for each task by checking one of the following: they have not performed the task, they have performed it under supervision, they have performed it independently, or they delegated the task to subordinates and reviewed their performance. Applicants receive points for each task based on their degree of experience, and their total scores consist of the total of the task scores. Another method, the job element or KSA method, is similar to the task-based method except that candidates rate their degree of knowledge or skill for each of the knowledges, skills or abilities deemed necessary for efficient job functioning.

The behavioral consistency method. Use of the behavioral consistency method involves having candidates provide examples of their past achievements that relate to important job dimensions. Subject matter experts then typically rate a sample of the applicant achievements in order to provide examples of behavior along a performance continuum for each dimension. That is, the candidate achievements are identified as examples of low, high, and

moderate performance. Behavioral consistency raters then rate all of the candidates' past achievements by comparing them to the examples along the performance continuum provided by the subject matter experts.

The activity/achievement indicator. The activity/achievement indicator method is based on the behavioral consistency approach and was developed by Ash (1984) to avoid some of the problems engendered by that method. The development of the activity/achievement indicator is similar to that of the behavioral consistency questionnaire. However, the benchmarks identified by the subject matter experts are included as part of the questionnaire. The applicant is asked to choose, for each dimension, the type of accomplishment most similar to those he or she had done him or herself.

Incidence of Use

It would be hard to overestimate the degree to which applications and resumes are used throughout government, industry, and education although the formal scoring methods discussed above probably predominate in the public sector. A 1984 survey of state and municipal jurisdictions conducted by the State of Alabama showed that the use of training and experience methods to rate applicants for public sector jobs is quite high. The survey found that all except one jurisdiction used training and experience evaluations to select candidates for some job classes and that use varied from complete to seldom. The survey also showed that

seventy-five percent of the agencies used training and experience evaluations 25% of the time and that most used the point evaluation system.

Studies

Two groups of researchers (Schmidt, Caplan, Bemis, Decuir, Dunn, and Antone, 1979 and Hough, 1984) have conducted studies describing the development of the behavioral consistency approach. The purpose of the Schmidt et al. (1979) study was to explain the rationale and content validity basis for the new approach and to do a study of its empirical validity and utility. The study included a comparison of the point and KSA rating systems and was conducted on a sample of budget analysts. Since the return rate for all types of questionnaires was low (20%), the validity and utility analyses could not be done. However, reliability data showed superiority for the behavioral consistency approach which had a reliability of .78 for one rater compared to .48 and .52 for the point and KSA approaches respectively. However, with three raters the point and KSA reliabilities improved to .74 and .77 respectively. Correlations among the three rating methods showed high comparability between the point and KSA methods (.94) but low comparability between the behavioral consistency approach and both of the other two methods (.11 and .05). Efficiency of scoring (the amount of time needed for rating) was low for the behavioral consistency method in comparison to the other approaches.

A second study of this method was done by Hough (1984).
Hough applied the behavioral consistency approach to
attorneys in a federal regulatory agency, asking them to
describe their accomplishments for each of nine job
dimensions. Hough found that the reliability of the
Accomplishment Record Inventory varied from .75 to .80 for
the nine dimensions and was .85 for the total dimension
scores. Concurrent validities for the dimensions (with
performance ratings comprised of both dimension and
task ratings) were significant, varying from .17 to .25 for
the single dimensions to .25 for the overall behavioral
consistency evaluation. Behavioral consistency ratings were
related to amount and level of experience as an attorney but
not to educational variables such as scores on law aptitude
and knowledge tests, school grades, honors, or quality, or
to self-perception or other prior-experience variables. The
study also showed no effect for race or sex.

Johnson, Guffey, and Perry (1980) conducted a study
comparing the behavioral consistency method with the point
and task-based methods for selecting senior eligibility
counselors. Their study showed a significant concurrent
validity coefficient of .25 for the behavioral consistency
method and low and non-significant validity coefficients for
the other two methods. Relationships between the
traditional point method and the other two methods were low
and non-significant. Amount of experience with the hiring
agency was highly related to point method ratings but was

unrelated to ratings of the other two approaches or to ratings of job performance.

Ash conducted several studies comparing the behavioral consistency approach with other methods. A 1983 study compared the behavioral consistency method with the holistic method for the evaluation of students applying for a job of planner. Results showed that intrarater reliability was higher for the behavioral consistency approach (.95) than for the holistic approach (.77). However, interrater reliability was comparable (.84 and .83 respectively). Time for applicant completion of the questionnaires was 99.3 minutes and 43.2 minutes for the behavioral consistency and holistic approaches respectively. Scoring time was higher for the behavioral consistency method. The correlation between the two methods was .36, and their correlations with an IQ test were low.

A second study by Ash (1982) compared four training and experience methods: the point method, the grouping method, the task-based method and the behavioral consistency method. The study was carried out using three jobs: Auto Equipment Repair Foreman, Computer Operations Supervisor, and Medical Disability Examinations Supervisor. A comparison of the reliabilities of the four methods showed the task-based method to be superior with reliabilities of .98 to .99 for the three jobs. Reliabilities for the point method ranged from .77 to .92, for the grouping method from .44 to .78, and for the behavioral consistency method from .76 to .93.

The grouping method was superior to the others with respect to validity, which was based on peer ratings. It had a significant validity coefficient of .35 while the task-based method had a marginally significant validity of .21 and the other two methods had non-significant and low validities. Completion rate was highest for the point and grouping methods with return rates of .97. The task-based method had a completion rate of .90 and the behavioral consistency method a completion rate of .56. Raters spent less time on the grouping and task-based methods and more time on the point and behavioral consistency methods. Finally, the correlation between the behavioral consistency and the task-based method was moderate (.36 to .54), but the correlation between the behavioral consistency method and the other methods was low.

The purpose of a final study by Ash (1984) was to compare the behavioral consistency approach with a new approach (the activity/achievement indicator method) designed to overcome a major problem of the behavioral consistency approach--low completion rate on the part of applicants. The study also compared these two measures with a KSA-based questionnaire. Reliability results showed the superiority of the behavioral consistency method. Inter-rater reliability was .50 to .80 for the dimensions and .74 for the total scores while reliability for the activity/achievement indicator using coefficient alpha was .51 to .73 for the dimensions and .56 for the total scores.

The test-retest reliability for the KSA-based approach was
.60 to .89 for the dimensions and .71 for the total scores.
A multitrait-multimethod matrix showed convergent validity
for some of the dimensions. However, discriminant validity
was not shown for any dimension. Correlation of the
behavioral consistency and activity/achievement indicator
was moderate (.58). However, both methods had low
correlations with the KSA-based method. Applicant
completion time was unknown for the behavioral consistency
method, but it was under 30 minutes for both of the other
methods. Rater time was substantial for the behavioral
consistency approach while rating was done by computer for
the other two methods.

A final study of training and experience approaches was
done by Pannone (1984), who described the development of a
task-based questionnaire for the selection of applicants to
take a written examination for electrician. The tasks
performed by electricians were listed on the questionnaire,
and applicants indicated whether they had performed the
tasks and at what level. Pannone found that the reliability
(using coefficient alpha) of the task-based questionnaire
was .96 and that the validity (correlation with the written
test) was .42. Results of the task-based method were
compared with years of experience and education. Experience
and education correlated with the written test .13 and .11
respectively and with the task-based method .30 and .09
respectively. When faking was taken into account, the

validity was higher; the questionnaire and written test correlated .55 for non-fakers. The correlation for fakers was .26.

Comparison of Methods

While the results from the studies of training and experience methods have been neither consistent nor completely conclusive, they allow comparison of the methods along several important dimensions. Table 1 shows the training and experience methods, the dimensions along which they have been compared, and a summary of study results for each method-dimension intersection. Several factors not discussed above, but which are important in a comparison of the methods, have also been included.

This summary may be somewhat misleading because some of the conclusions have been based on only one study while others have been based on several. However, it is nevertheless useful in a comparison of the methods.

A primary criterion in selecting the training and experience method to be investigated in this study was that it be parallel in content and development to the oral examination proposed earlier. This necessitates a content validity base or, in other words, development based on the job tasks or dimensions determined to be relevant in a job analysis. Three methods fit that criterion--the task or KSA-based method, the behavioral consistency method, and the activity/achievement indicator method. The point and grouping methods are not based on a content validity model

Table 2.1

Comparison of Training and Experience Methods

Dimension	Method				
	Point	Group-ing	Task	BC[a]	AAI[b]
Content[c]	no	no	yes	yes	yes
Rationale	poor	poor	good	good	good
Reliability	mod/high	low/mod	high	mod/high	low/mod
Validity	low	mod	low/mod	low/mod	---
Correlation[d]	low/high	low/mod	low/mod	low/mod	mod
Completion	high	high	high	low	high
Rater time	mod	mod	low	high	low
Fakeability	no	no	yes	no	yes

Note. Mod = moderate.

[a]BC = behavioral consistency. [b]AAI = activity/
achievement indicator. [c]Content = content validity base.
[d]Correlation = correlation with other methods.

since they are not based on job analysis.

A second major criterion focused on rationale. The rationale for the behavioral consistency model, explicated by Schmidt et al. (1979), is that past behavior is the best predictor of future behavior. What the person accomplished on the job is important rather than the fact that he or she merely held the job and was exposed to the job tasks. The Schmidt et al. approach provides a way of determining the level of past performance rather than assuming that performance on similar jobs is similar. Since the activity/achievement approach is derived from the behavioral consistency approach, it has an identical rationale, and the task and KSA-based methods could be considered to have a similar rationale. However, the point and grouping methods assume that performance on similar jobs is similar.

Reliability and empirical validity are obviously critical factors in evaluating testing methods. Reliability has generally been high for the point, task, and behavioral consistency methods. However, the Schmidt et al. study showed higher reliability for the behavioral consistency than for the point system, and the Johnson et al. study showed higher validity for the behavioral consistency method than for the point or task-based methods. Empirical validity has generally been low for the point method and low to moderate for the behavioral consistency and task-based methods. Both reliability and validity are promising for the behavioral consistency and task-based

methods. However, because there have been few reliability and fewer validity studies of these methods, the evidence is inconclusive. It seems prudent to rely on a content valid method and to seek to improve reliability by having several raters. Several raters are also necessary for fairness even though for some methods there is adequate reliability with one rater.

It does seem clear according to the studies that choice of method is a meaningful decision since the correlations among the methods are generally low to moderate.

The last set of factors to be compared among the methods includes completion rate, rater time, and fake-ability. The comparisons with respect to these factors are clear. Completion rate is generally high for all methods except the behavioral consistency approach. Rater time is low for the task and activity/achievement approaches, moderate for the point and grouping approaches, and high for the behavioral consistency approach. The task and activity/achievement approaches are particularly prone to fakeability while the point and grouping methods are much less so and the behavioral consistency method least of all.

Based on the purpose of the proposed study and the factors discussed above, the behavioral consistency method was selected for comparison to the oral examination for the following reasons: it is based on a content validity model with development parallel to that of the oral approach; it is based on a well-developed rationale; it has high

interrater reliability, especially with several raters; it has shown promising though not strong empirical validity; and it is not as easily fakeable as the other methods. Problems with use of this method have to do with completion rate and rater time. However, completion rate was high for at least one study (Hough, 1984) and low completion rates may have been due in part to lack of incentive for the applicants, who were taking part in concurrent validity studies and had nothing personal to gain from participating. Rater time is higher for this method, but that is not a critical criterion since time per candidate for the oral interview is probably longer.

Chapter Three

METHODOLOGY

This research study compared the use of an oral examination to the use of a behavioral consistency examination in the selection process for two positions, Affirmative Action Officer and Sanitation District Manager, both of which were managerial positions in a large midwestern city.

Methodology for the

Affirmative Action Officer Study

Subjects

The subjects were 18 applicants for the position of Affirmative Action Officer, a managerial position in the Personnel Department in a large midwestern city. Each met the following minimum education and experience requirements: a Bachelor's Degree with a major in personnel management, public administration, business, the social sciences, or related field and five years of affirmative action experience performing duties closely related to those of the job or five years of experience in personnel management with significant responsibility for affirmative action. Subjects applied for the position based upon a job announcement listing job duties, job requirements, and type of examination.

Raters

The raters were eight subject matter experts in the
areas of personnel management or affirmative action who were
not employed by the Personnel Department and who were
willing to donate their time to the hiring process. Four of
the raters were used as oral examination raters while the
other four raters were used as behavioral consistency
examination raters. There were two minorities and two
non-minorities, two women and two men on each panel.
Because of the large time demand of each type of test method
upon the raters, it was infeasible to use the same raters
for both methods.

Job Analysis

Test development for both the oral examination and the
behavioral consistency examination was based on a job
analysis method which defined essential job tasks and
critical knowledges, skills, and abilities for the
Affirmative Action Officer job.

Development of job task statements and a job task
inventory. An initial group of task statements was taken
from a job description for the position developed when
recruitment of applicants began in November, 1984. Then a
group of four subject matter experts from the Personnel
Department (supervisors and colleagues of the position)
developed additional statements in a two to three hour
brainstorming session held during December, 1984. A final
group of applicable task statements was added to this list

from a task inventory for the position of Personnel Assess-
ment Specialist, which had been recently developed for a
professional organization of personnnel assessment
specialists.

These three sources provided 55 task statements, which
were grouped into five major areas by the researcher. These
areas were: Affirmative Action Plan Activities, Compliance
Expert Activities, Affirmative Action Status and Account-
ability Activities, Affirmative Action Project Activities,
and Supervisory/Management Activities. These groupings were
verified by a test analyst assigned to this project by the
Personnel Department.

A task inventory (see Appendix A) was then developed in
order to determine which of the job tasks were essential to
the Affirmative Action Officer job. The inventory consisted
of the job tasks grouped according to major area and three
five-point rating scales developed to ascertain the
importance, relative amount of time spent, and consequence
of error for each task.

Development of knowledges, skills, and abilities
(KSA's) and a KSA inventory. An initial list of relevant
KSA's was developed by the four subject matter experts, who
selected appropriate statements from a set of KSA's for the
position of Assistant City Personnel Director during the
same meeting in which task statements were generated. This
list of KSA's was augmented by the researcher, who developed
additional KSA's based on the previously developed tasks.

These KSA's were grouped into seven major dimensions by the researcher, with these dimensions verified by the test analyst. The dimensions were: Knowledge of Affirmative Action; Knowledge of Personnel Management; Planning and Analysis Skills; Decision-making, Judgment, and Independence Skills; Communication Skills/Interpersonal Skills; Supervision Skills; and Other Management Skills. Then three rating scales were developed to assess importance, necessity at time of hire, and usefulness in distinguishing effective from ineffective workers for each KSA. The rating scale for necessity at time of hire was a three point scale while the other two scales were five point scales.

The KSA inventory (see Appendix B) consisted of 91 KSA's grouped into seven major dimensions and three rating scales used to determine which KSA's were critical for effective job performance.

Task inventory results. Fifty-five tasks were independently evaluated by five subject matter experts from the Personnel Department (three of whom had been involved in task statement development) during January, 1985.

The initial criteria set for defining critical tasks were as follows: each task defined as critical had to receive three or more ratings at Point 4 or above (the high scale values) on each of the three scales. However, when these criteria were applied to the tasks, only 16 survived. Since so many tasks were discarded using this approach, the reasons for removal were analyzed, particularly with respect

to the second criterion, which concerned relative time spent. Of the 39 tasks omitted, 22 had been omitted because of the second criterion and 17 because of at least one other criterion.

Since a task could conceivably be critical although rarely performed, this scale was dropped for those tasks meeting more stringent criteria for the first and third scales--at least 4 ratings at Point 4 or 5. Application of the new criteria resulted in 16 additional surviving tasks for a total of 32. Appendix C contains the list of task groups and essential tasks.

KSA inventory results. Ninety-one KSA's were independently evaluated by the five subject matter experts. Critical KSA's were defined as those receiving at least three ratings of 4 or above on the importance scale, at least 3 ratings of 4 or above on the effectiveness scale, and at least 3 ratings at 3 on the necessity at time of hire scale. When these criteria were applied to the KSA's, 62 survived the criteria, and 2 which almost met the criteria were discussed by the researcher and test analyst and retained as critical KSA's.

The surviving KSA's were combined into a slightly different set of 10 dimensions. One dimension was eliminated because only one of its KSA's survived the criteria. (The surviving KSA was added to another appropriate dimension.) In addition, several dimensions were broken up to provide clearer definitions. The

resulting ten dimensions were as follows: Knowledge of
Affirmative Action; Planning and Organizing Skills;
Analytical and Quantitative Reasoning Abilities;
Decision-making, Judgment, and Independence Skills; Oral
Communication and Interpersonal Skills; Written
Communication Skills; Supervisory Skills; Initiative/
Creativity/Intelligence; Toleration of Stress; and
Professionalism. Appendix D contains the list of revised
dimensions and critical KSA's.

Task-KSA link-up. After the essential tasks and
critical KSA's had been defined based on inventory results,
the researcher completed a rational link-up of the task
groups and KSA dimensions. Each KSA dimension had a link-up
with at least one task group (see Appendix E).

Dimensions tested. Based on reported problems in
the research literature regarding low return rates for
behavioral consistency questionnaires, it was decided to
reduce the number of dimensions to be tested in the research
design to five. The researcher and test analyst assessed
each dimension for practical testability and dropped four
dimensions from consideration by discussion and consensus,
leaving 6 dimensions to be assessed. Those considered
difficult to assess, particularly by the oral examination,
were Decision-making, Judgment, and Independence Skills;
Initiative/Creativity/Intelligence; Toleration of Stress;
and Professionalism.

Since it was untestable by the oral examination method,

the Written Communication dimension was also eliminated from
the design, although, since it was so important to job
performance, it was included as a separate part of the
examination. The candidates completed a written problem
prior to the oral examination, which was scored separately
by the oral raters after they had completed scoring the oral
portion. Candidates also completed a multiple-choice test
which was not part of the research design. This test was
developed to further test the Analytical and Quantitative
Reasoning Abilities dimension. It was included as an
unweighted component in the testing process because it was
critical that anyone hired be minimally competent in this
area.

The decision to eliminate dimensions could have been
based on the number of their related task groups. However,
it was decided that this would be inappropriate since a
small number of related task groups did not imply lack of
criticality. For example, the supervisory dimension was
related to only one task group but was very important to job
performance according to consensus between the researcher
and test analyst.

The final test plan was as follows. Five dimensions
(Knowledge of Affirmative Action, Planning and Organizing
Skills, Analytical and Quantitative Reasoning Abilities,
Oral Communication and Interpersonal Skills, and Supervisory
Skills) were tested by both the behavioral consistency
and oral examination approaches. The Analytical and

Quantitative Reasoning Abilities dimension was further
tested by a multiple-choice examination, and the Written
Communication dimension was tested by a written problem.
The written problem and multiple choice test were not part
of this research study. Appendices F and G contain the list
of dimensions and KSA's tested and definitions for the
dimensions tested by the behavioral consistency and oral
examination methods.

Behavioral Consistency Examination Development

Questionnaire development. Development of the
behavioral consistency questionnaire (entitled the
Achievement History Questionnaire) was based upon examples
available in the research literature and the dimensions
generated in the job analysis. Dimension definitions were
developed from the KSA's comprising the dimensions. Because
the behavioral consistency approach consists of asking
candidates to describe their major achievements for each job
dimension, development of specific questions for each
dimension was not necessary. However, it was necessary to
provide a relevant example and to modify the instructions
found in the research literature. Criteria for selecting an
example dimension and constructing an example accomplishment
were that the dimension not be in the group of critical
dimensions to be evaluated in the examination but be one
with which the candidates would likely be familiar. The
dimension selected was that of training skill, and a
relevant accomplishment was developed based on the experi-

ence of the researcher.

Rating scale development. Several options for the development of benchmark accomplishments for the behavioral consistency questionnaire rating scales were considered.

The research literature suggested that subject matter experts use a sample of the accomplishments returned by the candidates to develop benchmarks (Ash, 1984; Hough, 1984; and Schmidt et al., 1979). Other approaches considered by the researcher included asking the behavioral consistency questionnaire rating panel to use a sample of the questionnaires to develop benchmarks, asking the subject matter experts to generate benchmarks, and asking the subject matter experts to generate general criteria.

The use of actual accomplishments provided by the candidates was desirable since such accomplishments would be realistic and meaningful. As discussed in the research literature, subject matter experts are typically asked to rate the examples on a seven point scale, and examples whose ratings have high interrater agreement are included in the dimension rating scales to be used by the behavioral consistency raters, who are usually personnel generalists or psychology students. However, development of benchmarks by the subject matter experts seemed inappropriate in this case because, with such a small group of candidates, the majority of questionnaires would have to be used for benchmark development with few left to be rated according to the benchmarks, making the task of the rating panel superfluous.

There was also a timing problem with this approach, since the accomplishments were due only one day before the rating panel was scheduled to meet.

Having the actual rating panel develop benchmarks from the accomplishments furnished by the candidates also seemed inappropriate due to the small applicant sample size. The rating panel would have had to use most of the accomplishments for benchmark development, leaving few to be rated according to the benchmarks.

These two approaches were rejected, and the researcher decided to have subject matter experts brainstorm accomplishments that they considered to be examples of high, moderate, and low performance with respect to the dimensions. Examples rather than general criteria were chosen for development since they seemed more useful to the raters.

Four subject matter experts from the Personnel Department, a different but overlapping group from those involved previously, were asked to brainstorm accomplishments of previous effective or ineffective Affirmative Action Officers. They were also asked to generate examples of high, low, and moderate accomplishments that they thought would be typical of those the applicant group would submit. The subject matter experts brainstormed 49 such accomplishments during February, 1985.

These accomplishments were placed in a randomized list, and the same subject matter experts independently sorted them into dimensions (a choice of more than one dimension

was allowed) and rated them according to level of perform-
ance on a seven point scale.

Accomplishments were assigned to a dimension if two or
more raters chose that dimension. Some accomplishments were
assigned to more than one dimension. Means and standard
deviations for the level of performance were determined
for each accomplishment. It was intended that those with
standard deviations of 1.5 or greater be eliminated.
However, there were none.

The next stage of the process consisted of developing a
matrix of accomplishments organized by dimension and level.
This was done to provide a means of ascertaining scale
coverage for each dimension and the degree of overlap of
accomplishments across dimensions. The goal at this stage
was to reduce the number of accomplishments at each level of
each dimension to two, to use as many different accomplish-
ments as possible, and to minimize the overlap of accom-
plishments across dimensions. Selection of the accomplish-
ments was done by the researcher and test analyst with these
goals in mind. This process resulted in the development of
two new accomplishments which covered empty cells in the
matrix. The final step was to rewrite the accomplishments
to provide additional clarity and to correct grammar.

The final behavioral consistency questionnaire rating
scales consisted of benchmarks along seven point rating
scales for each dimension. The number of accomplishments
per dimension varied from six to nine with 34 accomplish-

ments for all five dimensions. The scale points defined varied on each scale.

Description of the behavioral consistency examination. The behavioral consistency examination had content parallel to that of the oral examination except that candidates were asked to describe their major past achievements which were illustrative of their capabilities in each of the job dimensions identified through the job analysis. Each dimension was rated according to a behaviorally anchored rating scale with exemplary statements along various points of the scale. There was a separate rating form for each dimension on which raters recorded the ratings for all candidates.

The behavioral consistency examination (see Appendix H) consisted of a list of the job dimensions and their definitions, a form on which to provide the requested information, instructions, and an example. Applicants were requested to describe an accomplishment for each dimension which illustrated their degree of competency on that dimension. For each achievement they were asked to provide: what they did and the objective of the achievement, the outcome of the achievement, the amount of credit they claimed if there were shared responsibilities, when the achievement occurred and for what employer, and the name, address, and telephone number of a reference who could verify their statements. Acceptable accomplishments could be drawn from educational, job-related or volunteer experiences. They could be based

upon one-time incidents or long-term projects or policies.

Oral Examination Development

Question development. During February, 1985, each of the four subject matter experts was presented with the task inventory and dimension definitions and was asked to independently develop two questions for each of the five dimensions to be tested. This resulted in 64 questions.

All questions for each dimension were assembled, and the same four subject matter experts met to select two questions per dimension from the set. First each subject matter expert independently selected his or her four preferred questions for each dimension. Then differences in choice were discussed and consensus reached on the top two questions for each dimension. Several questions were switched from one dimension to another by group consensus.

The final oral examination consisted of ten questions (two per dimension) with a general question at the end allowing the candidate to add anything he or she wished.

Development of question rating scales. The goal in developing rating scales for the oral questions was to provide benchmark answers for seven point rating scales at high, mid, and low scale values for each question.

The same four subject matter experts who had developed and selected the questions were asked to brainstorm likely answers along a seven point rating scale for each question. They provided criteria for a seven point answer for all dimensions but were unable to consistently provide criteria

for the other scale values. Benchmarks for poor answers were developed for seven of the ten questions, but benchmarks for moderate-quality answers were developed for only four of the questions. However, for many of the questions it was reasonable to define moderate and/or poor answers as those which covered only some or few of the points expected in a high quality answer.

Description of the oral examination. The oral examination consisted of two questions for each of the five dimensions identified through the job analysis procedure. An additional question at the end allowed the candidates to add anything they wished. Each question was rated on a seven point numerically anchored rating scale with benchmark answers at all high scale values and at some mid and low scale values. Candidates' answers were compared to the benchmarks included on the question rating scale, and ratings were recorded on a question rating form. Then each rater combined the ratings on the relevant questions to arrive at a score for each of the dimensions. The questions pertaining to each dimension were listed on the dimension rating form. Although two questions were developed specifically for each dimension, most of the questions contributed to more than one dimension. For example, each question contributed to the rating on the Communication/ Interpersonal Skill dimension. Dimensions were rated on a scale of 60 to 100, with 70 defined as the passing score. Each candidate's final score consisted of the average of

his/her average dimension scores. After completion of the oral dimension ratings, the oral board raters also rated the Written Communication dimension separately based on a written problem which each candidate had completed prior to his or her oral examination.

Procedure

Rater training for the behavioral consistency examination. There were separate training sessions for the two groups of raters, one for those who were oral examination raters and one for those who were behavioral consistency examination raters. Training lasted approximately two hours for the behavioral consistency questionnaire raters and occurred immediately prior to the rating session.

The behavioral consistency questionnaire training session included discussion of the following points: purpose of the ratings, rationale for the selection process being used, explanation of the dimensions and KSA's and how they were developed, definition of the dimensions to be rated, explanation of the rating scale including its development, and candidate experience requirements. Discussion of the rating scales included a modification of the Supervisory Skills dimension rating scale. (The benchmark for the scale value of 6 was changed to a scale value of 4, and the wording for one of the Scale Value 5 examples was changed slightly.)

Raters were also cautioned and advised on the following

points: they were asked to rate the candidates independently without reference to the ratings of the other judges; they were asked to rate one dimension at a time, putting the questionnaires in three quality groups for that dimension and then making finer distinctions within the groups; they were asked to be lenient regarding a candidate's selection of an accomplishment for one dimension versus another since an accomplishment could be expected to relate to more than one dimension; and finally they were told that there would be a wrap-up session after all candidates had been scored with an opportunity for discussion and consensus if there were significant disagreements regarding the candidates.

The following materials were also provided to the raters: a job announcement sheet, the job task inventory, a list of dimensions with definitions, the criteria for rating dimensions, the dimension rating form, the completed candidate questionnaires, and an introductory booklet for oral board members (provided approximately a week prior to the examination) which discussed the Personnel Department's general guidelines for board members.

After discussion of the preceding points, the raters were given two candidate questionnaires and asked to score them on all dimensions. Scoring of the two questionnaires was followed by a discussion of the differences among the ratings and some consensus on points of disagreement.

Behavioral consistency evaluation. The behavioral
consistency questionnaires were sent by mail to the
candidates who met the requirements two weeks prior to the
due date for their return in mid-February, 1985.

Twenty-nine of the 35 accepted candidates (83%)
returned completed questionnaires. Six candidates who were
initially judged not to have met the requirements for the
position also returned completed questionnaires along with
supplemental application forms in which they further
explained their experience. Prior to rating the behavioral
consistency questionnaires of the accepted applicants, the
panel was asked to make a determination on whether the six
rejected applicants met the requirements. The panel
accepted one of the applicants and confirmed the rejection
decisions for the other five. The final number of
questionnaires rated by the panel was therefore 30.

Ratings of the questionnaires, unidentified by name to
the raters, were completed during mid-February, 1985. The
questionnaire raters worked individually but in a group
setting. The group setting was optimal for ensuring
that the raters were not distracted while completing the
ratings and for ensuring that the raters actually completed
the ratings, a tedious and arduous task. Raters rated
each candidate on each dimension; each candidate's final
score was the total of his or her total dimension ratings.
After independent ratings of the questionnaires had been
completed, there was discussion and some changes of

dimension scores which were substantially different.
However, only the independent ratings were used for this
study.

Rater training for the oral examination. Training
for the oral examination raters took approximately one and
one-half hours and included a brief presentation by the
Personnel Director and a discussion of the following points
by the researcher: definition and development of the job
dimensions, critical KSA's, and job tasks; candidate
experience requirements; the selection process, including
the behavioral consistency questionnaire, oral examination,
and written problem; and the questions, question rating
criteria, question rating form, and dimension rating form.

The following materials were also provided to the
raters: a job announcement sheet, definitions of the job
dimensions and the job task inventory, the oral examination
questions and written problem, the rating criteria for the
questions and written problem, question and dimension rating
forms, and a schedule of candidates. Approximately a week
prior to the oral examination, the raters were also provided
with an introductory booklet for board members which
discussed the Personnel Department's general policies
regarding ratings. Raters were not provided with any
application or behavioral consistency questionnaire
materials.

The raters were also instructed to make independent
ratings, to ask follow-up questions only for purposes of

clarification (that is, to ask no new questions), and to indicate any candidates with whom they were too familiar to rate. They were informed that there would be a wrap-up session at the end of the process for discussion of major disagreements regarding the candidates.

A mock candidate (one of the subject matter experts used in the test development process) was then presented to the raters as an example for trying out the oral examination process. He had been instructed to give answers of varying quality to the questions and the question ratings given by the board members reflected the quality of his answers. Agreement among the raters on question ratings was high; because of lack of time, dimension ratings were not completed.

Oral examination. Rater availability precluded testing all 30 of the behavioral consistency questionnaire candidates by the oral method. Therefore, candidates were selected for inclusion in the oral examination process based on their scores on the behavioral consistency questionnaire. Twenty-two candidates were invited to take the oral examination with three people declining the invitation and one not appearing at the time of examination. One person who had not been rated by the behavioral consistency examination raters was added. This resulted in 19 candidates taking the oral examination, but only 18 were included in the research sample.

The oral examination was held during late February,

1985 and consisted of a 30 minute structured oral interview for each candidate. Candidates were assigned their order of appearing before the board according to their convenience since many were from out of town. Raters alternated in asking questions of the candidates. The raters rated each candidate independently as he or she was interviewed. Raters rated each candidate on each dimension; each candidate's final score was the average of his or her average dimension ratings. The written problem was rated after the oral dimension ratings had been completed. After all of the independent ratings were completed, substantial differences were discussed. However, only the independent ratings were used in this research study.

Ratings of examination efficiency and
acceptability. Both sets of raters were asked to complete forms assessing examination efficiency and acceptability after each examination (see Appendices I and J). Since there was time available, the achievement history questionnaire raters completed these ratings immediately after the rating process. The oral examination raters received their forms by mail after completion of the process since there was no time available immediately after the examination. Rater time for both examinations and candidate time for the oral examination was recorded by the test administrator.

Methodology for the
Sanitation District Manager Study

Subjects

The subjects were 14 applicants for the position of
Sanitation District Manager, a managerial position in the
Bureau of Sanitation in a large midwestern city. Each
met an experience requirement of current status as a city
employee and two years of experience as a Sanitation
Supervisor with the Bureau of Sanitation. Subjects applied
for the position based upon a job announcement listing job
duties, job requirements, and type of examination.

Raters

The raters were nine subject matter experts in the
areas of public works or personnel management who were not
employed by the Sanitation Bureau and who were willing to
donate their time to the hiring process. Five raters
(including one white female, one black male, and three white
males) were used as oral examination raters while four
raters (one black male and three white males) were used as
the behavioral consistency examination raters. Because of
the large time demand of each type of test method upon the
raters, it was infeasible to use the same raters for both
methods.

Job Analysis

Development of the oral examination and the behavioral
consistency questionnaire were based upon Flanagan's (1954)
critical incident job analysis technique and Smith and

Kendall's (1963) retranslation of expectations technique.
Development proceeded according to the following steps:

Critical incident generation. Two groups of
subject matter experts developed examples of effective and
ineffective behavior for the job being studied. Subject
matter experts, defined as supervisors or incumbents of the
job to be assessed, were assembled in separate groups and
asked to generate examples of effective behavior and
ineffective behavior based on their observations of job
incumbents. The supervisors, consisting of three area
managers (immediate supervisors of the subject position),
one district manager on special assignment, and the three
top level managers in the Bureau, developed incidents in two
group brainstorming sessions lasting approximately three
hours each during September, 1984. Then each manager
committed to individually providing five examples of
effective behavior and five of ineffective behavior. These
were returned within the next few weeks. The incumbents,
consisting of eight district managers, met in two group
sessions during September and October, 1984. Brainstorming
was done initially to ensure understanding of the objective.
Then the managers worked individually, with each committed
to providing five examples of effective and ineffective
behaviors. For the individually produced items, the
managers were asked to provide the following information:
what the person did that was effective or ineffective; the
circumstances surrounding the incident; why the incident was

an example of effective or ineffective behavior; and when the incident occurred.

The critical incident generation process produced 246 incidents, with the supervisors providing approximately two-thirds and the incumbents providing approximately one-third. Approximately one-half were brainstormed, and one-half were produced individually.

Generation of job dimensions. Job dimensions are the general knowledges, skills, and abilities necessary to do a job. The object of this part of the process was to generate job dimensions related to the critical incidents and to group the related critical incidents within those job dimensions. Generation of preliminary job dimensions was done during October, 1984 by one of the test analysts assigned to this project by the Personnel Department with advice from the researcher. (Personnel Department staff included the researcher and two test analysts.) The same test analyst then grouped together similar instances of effective and ineffective behavior into the job dimensions. Nine dimensions were initially generated for the Sanitation District Manager job: Planning/Organizing, Analyzing/ Decision-making, Researching/Investigating, Supervision, Controlling, Communication (Relaying Information), Oral/ Written Communication Skill, Training, and Professionalism.

Retranslation of dimensions/incidents. The purpose of the retranslation phase (completed during October, 1984) was to determine whether an independent judge would make

similar judgments regarding the grouping of incidents and dimensions. This phase was carried out by having the second test analyst group the same incidents and dimensions from randomized groups of each. Agreement between the two judges was determined by dividing the number of examples both judges placed in a category by the total number of examples placed in the category by either. Dimensions with less than 70% agreement between judges were reevaluated for clarity.

Degree of agreement between the two judges on placement of the critical incidents in the nine dimensions was as follows: Planning and Organizing--80%; Analyzing and Decision-making--75%; Researching and Investigating--71%; Supervision--83%; Controlling--29%; Communication (Relaying Information)--86%; Oral/Written Communication--67%; Training--100%; and Professionalism--90%.

Dimension determination. Completion of the job analysis and test development process had been scheduled for October through December of 1984, shortly after completion of the retranslation phase described above. However, two problems intervened to delay progress. First, the managers in the Sanitation Bureau became unavailable for meetings due to snow emergencies, and, second, the researcher became involved with the development and administration of the Affirmative Action Officer examination, the other project comprising this study. Consequently, job analysis and test development were interrupted during the fall and did not recommence until the following March.

Resolution of disagreements regarding dimension definitions and placement of incidents in categories was done by discussion and consensus between the two judges and the researcher. (One judge participated in only part of the process.) The results of the discussions were to combine the Oral/Written Communication dimension with the Communication (Relaying Information) dimension, to combine the Researching/Investigating dimension with the Analyzing/Decision-making dimension, and to combine the Controlling dimension with the Supervision dimension.

These decisions were based not only upon the degree of agreement between the two judges but also upon the intended definitions of the dimensions and the overlap between them. The eliminated dimensions were considered subparts of other dimensions and not as important in their own right as the other dimensions. The decisions were also based on the researcher's decision to reduce the number of dimensions to no more than five or six based on experience with the Affirmative Action Officer examination. This seemed prudent to avoid examinee and rater fatigue, particularly in completing and rating the behavioral consistency questionnaire.

As these decisions were being made, the definitions of the other dimensions were clarified, and additional incidents (incidents initially agreed upon) were moved from one dimension to another.

The six dimensions and number of related incidents

resulting from the discussion and consensus were as follows: Planning and Organizing--35; Analyzing and Decision-making--69; Supervision--68; Communication--40; Training--21; and Professionalism--13.

Review of dimensions and critical incidents. Prior to development of the critical incident inventory, the next step in the job analysis process, the researcher reviewed the incidents in their original format to ensure that all had been included in the analysis and that there were no duplicate incidents. Incidents were also rewritten by the researcher to provide additional clarity and completeness and to correct grammatical errors. As a result of this review, several previously omitted incidents were added, several duplicate incidents were omitted, and incidents which were a combination of two or more examples were separated. This process yielded 10 additional incidents which were added to the original set of 246 incidents and placed in appropriate dimensions by the researcher.

Because of the changes made in the dimension resolution stage (the elimination of several dimensions and the changes in dimension by critical incident match) and because of the addition of new critical incidents to the dimensions, the researcher did a second review of the placement of critical incidents in dimensions. Degree of agreement with the previous decisions was as follows for the six dimensions: Planning and Organizing--85%; Analyzing and Decision-making--60%; Communication and Interpersonal Skill--95%;

Supervision--79%; Training--100%; and Professionalism/
Dedication--55%.

An analysis of these differences revealed that there
was confusion between the Analyzing/Decision-making
dimension and three other dimensions: Planning/Organizing,
Supervision, and Professionalism/Dedication. Consideration
was given to combining the Analyzing/Decision-making and
Professionalism/Dedication dimensions with the most
appropriate of the other dimensions. However, that
idea was rejected because these dimensions were logically
different from the other two and seemed meaningful as
critical dimensions for the Sanitation District Manager job.
Instead, the definitions for the four dimensions were
further clarified and distinguished from one another. It
was evident, however, that many of the critical incidents
related to more than one dimension and that they could
reasonably be placed under more than one.

Final placement of the incidents in the dimensions was
done by consensus between the researcher and senior test
analyst. The six dimensions and number of related incidents
resulting from the discussion and consensus were as follows:
Planning and Organizing--35; Analyzing and Decision-making--
60; Communication/Interpersonal Skill--39; Supervision--76;
Training--23; and Professionalism/Dedication--23.

Critical incident inventory. The purpose of the
critical incident inventory was to determine average
effectiveness ratings for the critical incidents, which

could then be used as benchmarks for dimension effectiveness scales.

The critical incident inventory (see Appendix L) consisted of six dimensions and 256 incidents randomly listed within dimension. The incidents were independently scored on a seven point effectiveness scale by five subject matter experts (two top managers and three area managers) during April, 1985.

Most of the critical incidents received average ratings near the end points of the effectiveness scale, and all dimensions contained incidents with means at both high and low scale values. Most of the standard deviations were low. (Only 9% were over 1.5.) The standard deviations were highest for critical incidents with average ratings near the midpoint. Considering how critical incidents were defined (as examples of either effective or ineffective behavior), these results were quite reasonable. Examples purposely developed to be extreme would be expected to have average ratings near the end points of the range while examples averaging a moderate rating could be expected to do so because of disagreement rather than agreement on a moderate rating.

Dimension effectivness scale development. The final stage of the job analysis process was to develop dimension effectiveness scales based on the results of the critical incident inventory. The objective in developing the dimension effectiveness scales was to provide benchmarks

(critical incidents) at each point of a seven point rating scale for each dimension. Criteria for selection of the critical incidents as benchmarks were as follows: a high degree of agreement on scale value by the subject matter experts evidenced by a small standard deviation; scale coverage as complete as possible; and reduction of the number of critical incidents at each scale point to a meaningful number.

These criteria were applied to the critical incidents as follows:

1. Twenty-four critical incidents with standard deviations above 1.5 were omitted.

2. Critical incidents with standard deviations above 1.0 were omitted if their scale values were covered by other critical incidents. Three critical incidents were omitted according to this criterion.

3. Since there was a dearth of incidents with values at the middle of the scale, critical incidents at both whole and midpoint scale values were used (e.g., at Values 5.0, 5.5, and 6.0) to provide as much information as possible to the raters.

4. The number of critical incidents at each whole or midpoint scale value was limited to five since application of the above criteria still resulted in too many incidents at various scale values to be meaningful. Where choices had to be made, the critical incidents with average ratings closest to the scale value were retained, those with the

lowest standard deviations were retained, and retention decisions due to further ties were based upon the judgment of the researcher.

Application of the above criteria to the critical incidents resulted in effectiveness scales on six dimensions with five or fewer critical incidents listed as benchmarks at as many points as possible on 7 point rating scales. There were few examples on any of the scales for the middle values between 3.0 and 5.0. The number of incidents on each scale varied from 17 to 42.

Dimensions tested. As a result of a discussion with the Sanitation Bureau managers during March, 1985, a final modification was made to the dimensions by the addition of Written Communication Skill as a separate dimension. Written Communication Skill was defined as a separate dimension because it was important to successful job performance apart from Oral Communication/Interpersonal Skill and could not be tested by an oral examination. Because of this, the Written Communication Skill dimension was not included in the research design. However, it was rated by both the behavioral consistency and oral examination raters. The oral raters evaluated a written problem produced by each candidate prior to the oral examination, and the behavioral consistency examination raters evaluated the writing ability displayed throughout each candidate's behavioral consistency questionnaire.

The final set of dimensions tested included the

following: Planning and Organizing, Analyzing and Decision-making, Oral Communication/Interpersonal Skill, Supervision, Training, Professionalism/Dedication, and Written Communication Skill (see Appendix M for definitions). The first six dimensions were tested by the behavioral consistency and oral methods. The Written Communication Skill dimension was tested by the same two sets of raters but on the basis of a written problem and the writing sample produced by the behavioral consistency questionnaire.

Behavioral Consistency Examination Development

Questionnaire development. Development of the behavioral consistency questionnaire (entitled the Achievement History Questionnaire) was based upon examples available in the research literature and the dimensions generated in the job analysis. Since the behavioral consistency approach consists of asking candidates to describe their major achievements for each job dimension, development of specific questions for each dimension was not necessary. However, it was necessary to provide a relevant example, which was based on a critical incident from the job analysis, and to modify the instructions found in the research literature.

Rating scale development. As discussed earlier in this paper, previous researchers have developed benchmarks for behavioral consistency rating scales by using samples of the accomplishments provided by the applicants themselves.

Subject matter experts are typically asked to rate the examples on a seven point scale, and examples whose ratings have high interrater agreement are included in the dimension rating scales to be used by the behavioral consistency raters, who are usually personnel generalists or psychology students. This approach is ideal if there are a large number of candidates since it provides benchmarks which are realistic and meaningful. However, as pointed out earlier, it is not appropriate with a small number of candidates because drawing a large enough sample of accomplishments to develop the benchmarks would reduce the number left to be rated by the behavioral consistency raters to the extent that their task would be superfluous. For the same reason, a second approach of having the behavioral consistency raters themselves develop benchmarks is also inappropriate.

A third approach considered by the researcher for rating scale development was having the subject matter experts brainstorm benchmarks based upon the types of accomplishments they expected to be submitted. This approach was used in the Affirmative Action Officer study for development of both the behavioral consistency and oral examination rating scales and in the present study for development of the oral examination rating scales. However, use of this approach was rejected due to the difficulty experienced by the subject matter experts in coming up with examples based on their expectations in all three instances.

The fourth approach considered by the researcher was to

use the dimension effectiveness scales developed as part of the job analysis as the basis for rating accomplishments. The job incumbents could be considered comparable to the applicants because their duties were similar although at a higher level and many duties overlapped the two types of positions. However, this approach was also rejected because of the researcher's concern that there might not be enough overlap between the experiences of applicants and incumbents and because of other scale properties that made the scales inappropriate for this use. These were the dearth of midpoint scale values against which accomplishments could be judged and the large number of negatively worded incidents on the lower ends of the scales. (Applicants could be expected to submit low value accomplishments but not negatively worded accomplishments.)

Because none of the above approaches seemed appropriate, the researcher decided to use graphic rating scales rather than behaviorally anchored rating scales. This decision seemed reasonable in light of the performance evaluation research literature which indicates that behaviorally anchored rating scales are not superior to other types of rating scales with respect to the reliability and validity of ratings.

Description of the behavioral consistency examination. The behavioral consistency examination had content parallel to that of the oral examination except that candidates were asked to describe their major past

achievements which were illustrative of their capabilities in each of the job dimensions identified through the job analysis. Each dimension was rated on a graphic rating scale with numeric and qualitative anchors. After completion of the behavioral consistency dimension ratings, the behavioral consistency raters also rated the Written Communication Skill dimension separately using the questionnaire itself as a writing sample.

The behavioral consistency examination materials consisted of a list of the job dimensions and their definitions, a form on which to provide the requested information, instructions, and an example (see Appendix N). Applicants were requested to describe an accomplishment for each dimension which illustrated their degree of competency on that dimension. For each accomplishment they were asked to provide: what they did and the objective of the achievement, the outcome of the achievement, the amount of credit they claimed if there were shared responsibilities, when the achievement occurred and for what employer, and the name and title of a reference who could verify their statements. Acceptable accomplishments could be drawn from educational, job-related, or volunteer experiences. They could be based upon one-time incidents or long-term projects or policies.

Oral Examination Development

Question development. This stage consisted of presenting the group of subject matter experts (in this case, the top three managers) with the original list of dimensions and their associated examples of effective and ineffective job behavior. The subject matter experts were asked to brainstorm at least two questions per dimension, each based on a critical behavior exemplifying that dimension. Brainstorming took place in a series of five two to three hour meetings beginning in November of 1984 and ending in March of 1985. Sixteen questions were developed through the group brainstorming process. Two questions per dimension were then selected for inclusion in the oral examination by the researcher and test developer.

Development of question rating scales. The goal in developing rating scales for the oral questions was to provide benchmark answers for seven point rating scales at high, mid, and low scale values for each question. Development of the rating scales occurred simultaneously with question development. After the generation of each question, the subject matter experts were asked to develop examples of effective, ineffective, and adequate answers for each question based on how they would expect high, low, and moderate performers to answer the questions. They provided criteria for a seven point answer for all dimensions but were unable to consistently provide criteria for the other scale values. Benchmarks for poor answers were developed

for seven of the twelve questions, but benchmarks for moderate-quality answers were developed for only three of the questions. However, for many of the questions, it was reasonable to define moderate and/or poor answers as those which covered only some or few of the points expected in a high quality answer rather than being substantially different.

Description of the oral examination. The oral examination consisted of two questions for each of the six job dimensions identified through the job analysis procedure. Each question was rated on a seven point numerically anchored rating scale with benchmark answers at all high scale values and at some mid and low scale values. Candidates' answers were compared to the benchmarks included on the question rating scale, and ratings were recorded on a question rating form. Then each rater combined the ratings on the relevant questions to arrive at a score for each of the dimensions. Although two questions were developed specifically for each dimension, most of the questions contributed to more than one dimension. For example, each question contributed to the rating on the Oral Communication/Interpersonal Skill dimension. Dimensions were rated on a scale of 60 to 100, with 70 defined as the passing score. Each candidate's final score consisted of the average of his/her average dimension scores. After completion of the oral dimension ratings, the oral board raters also rated the Written Communication dimension

separately based on a written problem which each candidate had completed prior to his or her oral examination.

Procedure

Rater training for the behavioral consistency examination. There were separate training sessions for the two groups of raters, one for those who were oral examination raters and one for those who were behavioral consistency examination raters. Training lasted approximately two hours for the behavioral consistency raters and occurred immediately prior to the rating session.

The behavioral consistency questionnaire training session included discussion of the following points: purpose of the ratings, rationale for the selection process being used, explanation of the dimensions and critical incidents and how they were developed, definition of the dimensions to be rated, explanation of the dimension effectiveness scales including their development, explanation of the graphic rating scale, and candidate experience requirements.

Raters were also cautioned and advised on the following points: they were asked to rate the candidates independently without reference to the ratings of the other judges; they were asked to rate one dimension at a time, putting the questionnaires in three quality groups for that dimension and then making finer distinctions within the groups; they were asked to be lenient regarding a candidate's selection of a behavioral example for one dimension versus another

since an example could be expected to relate to more than one dimension; and finally they were told that there would be a wrap-up session after all candidates had been scored with an opportunity for discussion and consensus if there were significant disagreements regarding the candidates.

The following materials were also provided to the raters: a job announcement sheet, a list of dimensions with definitions, the dimension effectiveness scales from the job analysis, the graphic dimension rating form, the 14 questionnaires, and an introductory booklet for board members (provided approximately a week prior to the examination) which discussed the Personnel Department's general guidelines for oral board members.

After discussion of the preceding points, the raters were given three examples to score prior to scoring the candidates. The examples were provided by three higher level supervisors in the department. The training session concluded with a discussion of the differences among the independent ratings and some consensus on the initial points of disagreement.

Behavioral consistency evaluation. The behavioral consistency questionnaire was administered to 14 candidates in May, 1985 in a group session lasting up to four hours. There were 15 original applicants with one person dropping out prior to the administration of either examination. A group session was necessary in order to prevent candidates from collaborating. For the same reason, only one session,

prior to the oral examination, was given. Although test method was confounded with order, this was preferable to candidate collaboration.

Ratings of the questionnaires, unidentified by name to the raters, were completed during May, 1985. The questionnaire raters worked individually but in a group setting. The group setting was optimal for ensuring that the raters were not distracted while completing the ratings and for ensuring that the raters actually completed the ratings, a tedious and arduous task. Raters rated each candidate on each dimension; each candidate's final score was the total of his or her total dimension ratings. After independent ratings of the questionnaires had been completed, there was discussion and some changes of dimension scores which were substantially different. However, only the independent ratings were used for this study.

Rater training for the oral examination. Training for the oral examination raters took approximately three hours and included a brief presentation by the two top Bureau managers and a discussion of the following points by the researcher: definition and development of the job dimensions, critical incidents, and dimension effectiveness scales from the job analysis; candidate experience requirements; the selection process, including the behavioral consistency questionnaire, oral examination and written problem; and a review of the questions, question

rating criteria, question rating scale, and dimension rating scale.

The following materials were also provided to the raters: a job announcement sheet, definitions of the job dimensions and the dimension effectiveness scales from the job analysis, the oral examination questions and written problem, the rating criteria for the questions and written problem, question and dimension rating forms, and a schedule of candidates. Approximately a week prior to the oral examination, the raters were also provided with an introductory booklet for board members which discussed the Personnnel Department's general policies regarding ratings. Raters were not provided with any application or behavioral consistency questionnaire materials.

The raters were also instructed to make independent ratings, to ask follow-up questions only for purposes of clarification (that is, to ask no new questions), and to indicate any candidates with whom they were too familiar to rate. They were informed that there would be a wrap-up session at the end of the process for discussion of major disagreements regarding the candidates.

A mock candidate (one of the top three managers in the Bureau) was then presented to the raters for the oral examination. He had been instructed to give answers of varying quality to the questions; however, he was unable to bring himself to do this and gave consistently excellent responses. In consequence there was high agreement among

board members on his question ratings with no real opportunity for disagreement. Because of this and because of time pressure, dimension ratings were not made.

Oral examination. The oral examination was held during May, 1985 and consisted of a 30 minute structured oral interview for each candidate. Candidates were assigned their order of appearing before the board according to the convenience of the Sanitation Bureau. Raters alternated in asking questions of the candidates. The raters rated each candidate independently as he or she was interviewed. Raters rated each candidate on each dimension; each candidate's final score was the average of his or her average dimension ratings. The written problem was rated after the oral dimension ratings had been completed. After all of the independent ratings were completed, those with substantial differences were discussed. However, only the independent ratings were used for this research study.

Ratings of examination efficiency and acceptability. Both sets of raters were asked to complete forms assessing examination efficiency and acceptability after each examination (see Appendices O and P). Since there was time available, the behavioral consistency questionnaire raters completed these ratings immediately after the rating process. The oral examination raters received their forms by mail after completion of the process since there was no time available immediately after the examination. The time devoted to each examination by both

the candidates and raters was recorded by the test
administrator.

Method of Analysis

The first research question was whether the oral and
behavioral consistency methods provided comparable ratings
of applicants for employment. This question was analyzed by
computing Pearson product-moment correlation coefficients
between final overall ratings for the two approaches and
between final dimension ratings for the two approaches.
Statistical significance tests (to determine whether the
population correlations were greater than zero) were
conducted by using a table of critical correlation values
based on the t distribution.

The second research question was whether the oral and
behavioral consistency methods were equally reliable. This
question was assessed by comparing interrater reliabilities
for the overall rating and separate dimension ratings for
the two approaches. Interrater reliability was determined
according to the intraclass correlation method suggested by
Ebel (1951) with between-raters variance omitted from the
error term. (Between-raters variance was omitted from the
error term because each rater rated each candidate and final
ratings were based on averages or totals including ratings
from all raters.)

Significant differences between dimension and overall
reliability coefficients for the two approaches were
determined according to a method developed by Feldt (1980)

to test whether Cronbach's coefficient alpha is the same for two tests administered to the same sample. Using one of the three procedures recommended by Feldt, the test statistic was:

$$t_{N-2} = (W-1)(N-2)^{1/2} / [4W(1-r_{12}^2)]^{1/2}$$

where $W = (1-r_2)/(1-r_1)$,

$\quad r_1$ = reliability for Test 1,

$\quad r_2$ = reliability for Test 2,

$\quad r_{12}$= correlation between Tests 1 and 2, and

$\quad N$ = sample size.

The null hypothesis was that the population reliabilities were equal, and the alternative hypothsis was that the population reliabilities were different for each set of reliabilities tested. The t test statistic was referred to the t distribution with N-2 degrees of freedom. Oral reliabilities were adjusted to correspond to an average of four raters. (There were five oral and four behavioral consistency raters for the Sanitation District Manager examination while there were four raters for each method for the Affirmative Action Officer examination.)

The third research question was whether the two approaches demonstrated convergent and discriminant validity. This question was explored through the multitrait-multimethod matrix approach proposed by Campbell

and Fiske (1959), which argues that, in order for tests to demonstrate validity, they should have high correlations with other measures of the same trait and should have low correlations with similar measures of different traits.

The multitrait-multimethod matrix is comprised of correlations between at least two traits and at least two methods. The correlations are organized into monomethod and heteromethod blocks. Monomethod blocks are comprised of correlations between the same and different traits which have been measured by the same method and include reliability diagonals (correlations between the same traits using the same methods). Heteromethod blocks are comprised of correlations between the same and different traits which have been measured by different methods and include validity diagonals (correlations between the same traits using different methods).

According to Campbell and Fiske, convergent validity is demonstrated when there are significant and meaningful correlations between different methods measuring the same traits. (Values in the validity diagonals were tested for statistical significance as described above for the first research question.) Discriminant validity is demonstrated when three criteria are met: correlations between different methods measuring the same traits are higher than correlations between different methods measuring different traits (values in the validity diagonals are higher than those in corresponding rows and columns of the heteromethod

blocks); correlations between different methods measuring the same traits are higher than correlations between the same methods measuring different traits (values in the validity diagonals are higher than relevant values in the monomethod blocks); and correlations between the traits should reflect the same pattern across methods.

The fourth research question was whether the oral and behavioral consistency methods were equally acceptable to the raters and equally efficient in terms of rater time. This question was analyzed by comparing mean responses to the examination evaluation forms completed by the raters for the two approaches and by comparing total rater time for the two approaches. For the acceptability ratings, a statistical significance test (to determine whether the population means of the oral and behavioral consistency rater evaluations were equal) was conducted by using a t test statistic for testing differences between means of independent samples referred to the t distribution. Differences between rater times per candidate could not be tested for significance due to the absence of standard deviations for one or both methods in both studies.

The last research question was whether the oral and behavioral consistency approaches were equally efficient in terms of candidate time. This question was assessed descriptively by comparing mean (or estimated mean) candidate times for the two approaches. Differences in candidate time were not tested for significance because

candidate time for the behavioral consistency approach was not available for the Affirmative Action Officer study (and so was estimated) and a standard deviation for the behavioral consistency approach was not available for the Sanitation District Manager study.

The same methods of analysis were used for both of the research samples.

Chapter 4

RESULTS

<u>Results for the Affirmative Action Officer Study</u>

<u>Comparability</u>

The first research question was whether the oral and behavioral consistency methods provided comparable ratings of applicants for employment. This question was analyzed by computing Pearson product-moment correlation coefficients between final overall ratings for the two approaches and between final dimension ratings for the two approaches. The correlation between final overall ratings for the two approaches was .17. This figure increased to .25 when corrected for restriction in range of the behavioral consistency scores. The behavioral consistency scores were restricted in range because only 18 of the 30 candidates who took the behavioral consistency examination also took the oral examination. (The top 22 behavioral consistency candidates were invited to take the oral examination; however, four people dropped out.)

Correlations between dimension ratings for the two approaches were as follows with corrections for restriction in range in parentheses: Knowledge of Affirmative Action (Dimension 1)--.13(.13), Planning and Organizing Skills (Dimension 2)--.04(.06), Analytical and Quantitative

Reasoning Abilities (Dimension 3)--.18(.25), Oral Communication and Interpersonal Skills (Dimension 4)--.08(.09), and Supervisory Skills (Dimension 5)--.34(.39). Neither the correlation for the overall ratings nor any of the correlations for the dimension ratings was significant at the .05 level (df = 16, one-tailed test) even when corrected for restriction in range. Therefore, it must be concluded that the ratings from the two approaches were not comparable.

Reliability

The second research question was whether the oral and behavioral consistency methods were equally reliable. This question was assessed by comparing interrater reliabilities for the overall rating and separate dimension ratings for the two approaches. Interrater reliability, based on the average of four raters for both of the methods, was .84 for overall oral examination ratings and .81 for overall behavioral consistency examination ratings.

Interrater reliabilities for the dimension scores for the oral examination were as follows: Knowledge of Affirmative Action--.82, Planning and Organizing Skills--.83, Analytical and Quantitative Reasoning Abilities--.82, Oral Communication and Interpersonal Skills--.76, and Supervisory Skills--.71. Interrater reliabilities for the dimension scores for the behavioral consistency examination were: Knowledge of Affirmative Action--.64, Planning and Organizing Skills--.72, Analytical and Quantitative Reasoning Abilities--.79, Oral Communication and Interpersonal

Skills--.64, and Supervisory Skills--.75.

Testing for significant differences by using Feldt's method with an alpha level of .10 revealed no significant differences between dimension reliabilities or overall reliabilities (see Appendix K for the t values).

Convergent and Discriminant Validity

The third research question was whether the two approaches demonstrated convergent and discriminant validity. This question was explored through the multitrait-multimethod matrix approach proposed by Campbell and Fiske (1959). Table 4-1 contains a multitrait-multimethod matrix based on the two approaches and five dimensions of interest in this study.

According to the Campbell and Fiske approach, convergent validity is demonstrated when there are significant and meaningful correlations between different methods measuring the same traits. The validity diagonal in Table 4-1 does not demonstrate convergent validity for any dimension; all of the dimension correlations between the methods were low and non-significant (alpha = .05).

When convergent validity has not been demonstrated, discriminant validity is not possible. However, the discriminant validity criteria were applied to the matrix to further explore the relationships between the methods and dimensions. The first criterion was that the correlations between the different methods measuring the same traits be higher than correlations between different methods measuring

Table 4-1

Multitrait-multimethod Matrix for the Oral and Behavioral
Consistency Methods for Affirmative Action Officer

			Method							
			Oral				BC			
Oral	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
D1	.82									
D2	.96	.83								
D3	.89	.93	.82							
D4	.91	.91	.88	.76						
D5	.55	.62	.60	.75	.71					
BC										
D1	.13	.09	.01	.23	.29	.64				
	.13	.09	.01	.23	.29					
D2	.06	.04	.11	-.04	-.13	.11	.72			
	.08	.06	.15	-.06	-.18					
D3	.21	.22	.18	.07	-.10	.46	.43	.79		
	.28	.29	.25	.10	-.14					
D4	.12	.14	.10	.08	-.13	.05	.59	.43	.64	
	.13	.15	.11	.09	-.14					
D5	-.09	.01	-.02	.12	.34	.38	.36	.30	.37	.75
	-.10	.01	-.02	.14	.39					

Note. BC = behavioral consistency; D1 = Knowledge of
Affirmative Action; D2 = Planning/Organizing Skills; D3 =
Analytical/Quantitative Reasoning Abilities; D4 = Oral Com-
munication/Interpersonal Skills; D5 = Supervisory Skills;
figures below each value in the heteromethod block were cor-
rected for range restriction; monomethod block figures were
based on sample sizes of 19 and 30 for the oral and BC
approaches respectively; heteromethod block figures were
based on a sample size of 18. (Only the top behavioral
consistency candidates took the oral.) No validity diagonal
value was significant ($p > .05$, $df = 16$, one-tailed).

different traits. A comparison of the values in the
validity diagonal with the values in the corresponding rows
and columns of the heteromethod block revealed that this
criterion was met for only the Supervisory Skills dimension.
The correlations among dimensions across methods were
uniformly low.

The second criterion was that correlations between
different methods measuring the same traits be higher than
correlations between the same methods measuring different
traits. Inspection of the values in the monomethod blocks
revealed that this criterion was not met for any dimension.
The correlations between dimensions for the oral method were
considerably higher than the correlations between methods
for the same dimensions. Except for the Supervisory Skills
dimension, the correlations between dimensions for the
behavioral consistency method were also generally higher
than the correlations between methods for the same
dimensions.

The last criterion was that correlations between the
dimensions reflect the same pattern across methods. This
criterion was not met since there was no apparent pattern
across the monomethod blocks and the correlations in the
heteromethod block were all low.

Application of the three criteria clearly showed
method effect and failed to suggest anything except lack of
relationships between dimension-method combinations.

As a last step in the exploration process, the

monomethod blocks were analyzed to determine the relation-
ships between different dimensions when rated according to
the same method. For the oral examination, these relation-
ships were high (.88 to .96) with the exception of those
involving the Supervisory Skills dimension, which were
moderate. Except for the Supervisory Skills dimension, most
of the interdimension correlations exceeded the reliability
coefficients for the related dimensions. Interdimension
correlations for the Supervisory Skills dimension exceeded
its reliability coefficient in one case. For the behavioral
consistency examination, the relationships were generally
low to moderate. In this case the reliability coefficients
exceeded the interdimension correlations for every
dimension.

Rater Acceptability and Efficiency

The fourth research question was whether the oral and
behavioral consistency methods were equally acceptable to
the raters and equally efficient in terms of rater time.
This question was analyzed by comparing mean responses to
the examination evaluation forms completed by the raters for
the two approaches and by comparing total rater time for
the two approaches. The examination evaluation forms
consisted of questions regarding the effectiveness,
fairness, difficulty, and reasonableness of the two
approaches.

The mean rating for the behavioral consistency raters
was 15.75, with 20 being the highest rating possible. The

mean rating for the oral raters was 17.00 of 20. These mean
ratings were not significantly different, $t(5) = 1.38$,
$p > .05$, two-tailed test. Both sets of raters considered
their respective methods positive with respect to the
questions asked. These results were based on ratings from
all four of the behavioral consistency raters but from only
three of the four oral raters. One of the oral raters did
not return the evaluation form even after a follow-up
telephone call.

Average total time for the behavioral consistency
raters was 14.8 hours; average time per candidate was
one-half hour. These averages included training time
and lunch time in addition to the time actually spent rating
the candidates. They also included time that some of the
raters spent at home during the evening between rating
sessions. Average total time for the oral raters was 17.5
hours; average time per candidate was .9 hours. These
averages also included time for training and lunch but none
of the raters took materials home during the evening. Since
the behavioral consistency raters assessed 30 candidates
while the oral raters assessed 19, the average time per
candidate must be used to assess rater efficiency.
Differences between rater times per candidate could not be
tested for significance due to the absence of a standard
deviation for the oral group. However, the difference
between approximately one-half hour versus one hour per
candidate does have practical significance in the opinion of

the researcher, especially when raters are volunteers. If both the oral and behavioral consistency raters had rated 30 candidates, there would have been a difference of about one and one-half days in time with the oral raters taking longer. Despite this conclusion, inspection of the ratings for the question regarding time revealed that both sets of raters found the time demand reasonable.

Candidate Efficiency

The last research question was whether the oral and behavioral consistency approaches were equally efficient in terms of candidate time. This question was not formally assessed for this study because candidate time for the behavioral consistency approach was not available. However, candidate time for the oral examination was approximately one-half hour. Based on the researcher's previous experience, a likely estimate of candidate time for the achievement history questionnaire is ten hours. Therefore, time efficiency for the candidate is in favor of the oral examination unless significant travel time is involved, as it was for many of these examination candidates. In that case, time would be approximately the same. However, the majority of time used for behavioral consistency questionnaire completion would involve actual work on the questionnaire while the majority of time used for the oral examination would involve travel. The apparent time advantage of the oral may also prove illusory because application materials are required in most selection

processes. Such materials may be unnecessary if the behavioral consistency approach is used but required if an oral examination is used.

Results for the Sanitation District Manager Study

Comparability

For the Sanitation District Manager study, the correlation between final overall ratings for the oral and behavioral consistency approaches was .67. Correlations between dimension ratings for the two approaches were as follows: Planning and Organizing (Dimension 1)--.58, Analyzing and Decision-making (Dimension 2)--.29, Oral Communication/Interpersonal Skill (Dimension 3)--.62, Supervision (Dimension 4)--.74, Training (Dimension 5)--.31, and Professionalism/Dedication (Dimension 6)--.62. The correlation for the overall ratings was significant at the .01 level (df = 12, one-tailed test) as were three of the six correlations for the dimension ratings. One additional dimension correlation was significant at the .025 level. Therefore, it must be concluded that the ratings from the two methods were comparable.

Reliability

Interrater reliability for this study was based on the average of five raters for the oral examination and four raters for the behavioral consistency examination. For testing significant differences between reliabilities, the oral reliabilities were modified to correspond to an average of four raters. The numbers in parentheses indicate the

adjusted reliabilities. Interrater reliability was .92 (.90) for overall oral examination ratings and .86 for overall behavioral consistency examination ratings.

Interrater reliabilities for the dimension scores for the oral examination were as follows: Planning and Organizing--.90(.88), Analyzing/Decision-making--.91(.89), Oral Communication/Interpersonal Skill--.89(.87), Supervision--.80(.76), Training--.83(.80), and Profession-alism/Dedication--.87(.84). Interrater reliabilities for the dimension scores for the behavioral consistency examination were: Planning and Organizing--.80, Analyzing/ Decision-making--.61, Oral Communication/Interpersonal Skill--.73, Supervision--.68, Training--.54, and Profession-alism/Dedication--.88.

Testing for significant differences by using the Feldt method revealed a significant difference between the methods for one dimension--Analyzing and Decision-making--with the oral reliability higher (p < .05). There were no signifi-cant differences (at the .10 level) between reliabilities for the overall ratings or for the other five dimensions (see Appendix Q for the t values).

Convergent and Discriminant Validity

Table 4-2 contains a multitrait-multimethod matrix based on the two approaches and six dimensions of interest in this study.

According to the Campbell and Fiske approach, conver-gent validity is demonstrated when there are significant and

Table 4-2

Multitrait-multimethod Matrix for the Oral and Behavioral
Consistency Methods for Sanitation District Manager

	Method											
	Oral						BC					
Oral	D1	D2	D3	D4	D5	D6	D1	D2	D3	D4	D5	D6
D1	90											
D2	92	91										
D3	88	93	89									
D4	91	94	83	80								
D5	91	91	81	94	83							
D6	90	90	83	91	91	87						
BC												
D1	58	61*	69	48	56	42	80					
D2	30	29	47	29**	30	28	38	61				
D3	52	41	62	29**	25	35	45	58	73			
D4	80	73	79	74	81	69	58	49	51	68		
D5	43	29	46	24	31	22**	65	28	64	54	54	
D6	67	68	79	58	60	62	64	63	61	77	51	88

Note. BC= behavioral consistency; D1 = Planning and
Organizing; D2 = Analyzing and Decision-making; D3 = Oral
Communication/Interpersonal Skill; D4 = Supervision; D5 =
Training; D6 = Professsionalism/Dedication; decimal points
have been omitted from correlations; all figures in the
matrix are based on a sample size of 14.
* $p < .025$, df = 12, one-tailed. ** $p < .01$, df = 12, one-
tailed.

meaningful correlations between different methods measuring the same traits. The validity diagonal in Table 4-2 demonstrates convergent validity for four of the six dimensions with significant correlations ranging from .58 ($p < .025$) to .74 ($p < .01$).

Based on this finding, discriminant validity criteria were applied to the matrix to further explore the relationships between the methods and dimensions. The first criterion was that the correlations between the different methods measuring the same traits be higher than correlations between different methods measuring different traits. A comparison of the values in the validity diagonal with the values in the corresponding rows and columns of the heteromethod block revealed that this criterion was not met for any dimension. The correlations among the dimensions across methods varied considerably, and many were comparable to or higher than the correlations between methods for the same dimensions.

The second criterion was that correlations between different methods measuring the same traits be higher than correlations between the same methods measuring different traits. Inspection of the values in the monomethod blocks revealed that this criterion was not met for any dimension for either method. The correlations among the dimensions within method were generally comparable to or higher than the correlations between methods for the same dimensions.

The last criterion was that correlations between the

dimensions reflect the same pattern across methods. There was no pattern apparent across methods. However, the heteromethod block showed consistently high correlations between the Supervision and Professionalism/Dedication dimensions as measured by the behavioral consistency approach and all dimensions as measured by the oral. There were consistently low correlations between the Analyzing and Decision-making and Training dimensions as measured by the behavioral consistency method and the other dimensions as measured by the oral.

Application of the three criteria clearly showed lack of discriminant validity for the methods. Relationships between dimension-method combinations were generally moderate.

As a last step in the exploration process, the monomethod blocks were analyzed to determine the relationships between different dimensions when rated according to the same method. For the oral examination, these relationships were uniformly high (in the 80's and 90's), with most of the interdimension correlations equalling or exceeding the reliability coefficents for the related dimensions. For the behavioral consistency examination, the relationships were generally moderate. In this case the interdimension correlations exceeded or equalled the reliability coefficients for three of the six dimensions.

Rater Acceptability and Efficiency

The mean rating on the examination evaluation forms for the behavioral consistency raters was 29.75, with 40 being the highest rating possible. The mean rating for the oral raters was 33 of 40. Although the mean rating for the oral examination was higher, ratings for the two approaches were not significantly different, $t(7) = 1.57$, $p > .05$, two-tailed test. Both sets of raters considered their respective methods positive with respect to the questions asked. These results were based on ratings from all nine of the raters involved in the Sanitation District Manager examination.

Average total time for the behavioral consistency raters was 7.5 hours; average time per candidate was one-half hour. These averages included training time and lunch time in addition to the time actually spent rating the candidates. Average total time for the oral raters was 15 hours; average time per candidate was one hour. These averages also included time for training and lunch. Differences between rater time per method could not be tested for significance due to the absence of standard deviations for the two methods. However, the difference between one-half hour versus one hour per candidate does have practical significance, especially when raters are volunteers. There was essentially a day's difference in total time for 14 candidates with the oral raters taking longer. Despite this conclusion, inspection of the ratings

for the question regarding time revealed that both sets of raters found the time demand reasonable.

Candidate Efficiency

Candidate time for the oral examination was approximately one-half hour, and candidate time for behavioral consistency approach, completed in a group session, was approximately 3.5 hours. Differences in candidate time per method were not tested for significance due to the absence of a standard deviation for the behavioral consistency method. Based on descriptive statistics, time efficiency for the candidate was in favor of the oral examination since travel time was necessary for both methods. However, the point made earlier in the discussion of the Affirmative Action Officer study regarding application materials still holds. These may be unnecessary when the behavioral consistency approach is used but necessary when candidates take an oral examination.

Chapter 5

SUMMARY AND CONCLUSIONS

Summary

The purpose of this study was to compare the oral examination method to the behavioral consistency examination method. The need for such a study arose from the researcher's desire to find a testing method which possessed the desirable characteristics of the oral interview but which avoided its disadvantages. The oral examination is a practical selection technique for jobs with interpersonal and oral communication dimensions and for managerial jobs where there are relatively few candidates and technical knowledge and managerial skills must be assessed. Further, it can be considered content valid and fair if developed according to the content validity model proposed earlier. However, the research literature has shown that oral examinations have been plagued by several serious problems. Non-job-related applicant characteristics, rater characteristics, and situational characteristics can affect oral ratings. Ratings may also be subject to a halo effect due to the applicant's general likeability or oral communications skill over and above its importance to the job. Finally, it is logistically difficult to assemble the raters and candidates together on the same days.

The behavioral consistency approach, a relatively new approach to assessing training and experience, appeared promising as an alternative to the oral examination because it is parallel in development and content to the oral method, it is similar in administration except that the presentation format is written rather than oral, and there is no interaction between raters and candidates.

There has apparently been no previous research comparing the oral examination with the behavioral consistency approach. In fact, there has been little research comparing the oral examination with any other method, and the behavioral consistency method has been compared mainly to other training and experience approaches.

This research study compared the use of an oral examination to the use of a behavioral consistency examination in the selection process for two positions, Affirmative Action Officer and Sanitation District Manager, both of which were managerial positions in a large midwestern city.

For the Affirmative Action Officer study, there were 18 subjects and eight raters, four behavioral consistency raters and four oral examination raters. For the Sanitation District Manager study, there were 14 subjects and nine raters, four behavioral consistency raters and five oral examination raters.

For the Affirmative Action Officer study, test development for both approaches was based on a job analysis

method which defined essential job tasks and critical knowledges, skills, and abilities for the Affirmative Action Officer job. Subject matter experts developed job task statements and KSA's and completed inventories designed to assess essentiality and criticality. Similar job tasks and KSA's were combined into task groups and dimensions respectively, and a rational link-up was completed between the dimensions and task groups. Five dimensions were tested by the two approaches: Knowledge of Affirmative Action, Planning and Organizing Skills, Analytical and Quantitative Reasoning Abilities, Oral Communication and Interpersonal Skills, and Supervisory Skills.

For the Sanitation District Manager study, job analysis was done according to Flanagan's (1954) critical incident technique and Smith and Kendall's (1963) retranslation of expectations technique. Subject matter experts generated critical incidents (examples of effective and ineffective behavior). The researcher and two test analysts then developed job dimensions related to the critical incidents, grouped the critical incidents within those job dimensions, and retranslated the incidents and dimensions. The retranslation phase consisted of an independent regrouping of the dimensions and incidents and discussion and consensus on disagreements. The job analysis resulted in the identification of six job dimensions to be tested and a dimension effectiveness scale consisting of critical incidents along a seven point effectiveness scale for each

dimension. The six dimensions were as follows: Planning
and Organizing, Analyzing and Decision-making, Oral
Communication/Interpersonal Skill, Supervision, Training,
and Professionalism/Dedication.

For each study, the behavioral consistency examination
had content parallel to that of the oral examination except
that candidates were asked to describe their major past
achievements which were illustrative of their capabilities
in each of the job dimensions identified through the job
analysis. Development of the behavioral consistency
questionnaire was based upon examples available in the
research literature and the dimensions generated in the job
analysis. For the Affirmative Action Officer examination,
each dimension was rated according to a behaviorally
anchored rating scale with exemplary statements along
various points of the scale. For the Sanitation District
Manager examination, each dimension was rated on a graphic
rating scale.

The oral examination for both studies consisted of two
questions for each of the dimensions identified through the
job analysis procedure. Each question was rated on a seven
point numerically anchored rating scale with benchmark
answers at all high scale values and at some mid and low
scale values. Questions were developed by subject matter
experts who, in the first study, independently generated
several questions per dimension and then chose the final
questions through discussion and consensus. In the second

study, subject matter experts generated questions through group brainstorming. In this case, final questions were chosen by the researcher and test analyst. In both cases, the same subject matter experts brainstormed likely answers to the questions along a seven point rating scale. They provided criteria for a seven point answer for all dimensions but were unable to consistently provide criteria for the other scale values. Benchmarks for poor and moderate quality answers were provided for only some of the questions.

For both examinations, there were separate training sessions for the two groups of raters, one for those who were oral examination raters and one for those who were behavioral consistency examination raters. Training lasted from one and one-half to three hours for the four sessions.

The Affirmative Action Officer candidates completed the behavioral consistency questionnaires individually while the Sanitation District Manager candidates completed them in a group session. However, the behavioral consistency raters for both examinations were assembled as a group to rate the questionnaires. The oral examination for both studies consisted of a 30 minute structured interview for each candidate, and the raters were present as a group to evaluate each candidate.

Findings for the two studies were as follows:

1. For the Affirmative Action Officer study, the oral and behavioral consistency examination methods were not

comparable. Neither the correlation between final overall ratings for the two approaches nor any of the correlations between dimension ratings was significant.

For the Sanitation District Manager study, the oral and behavioral consistency examination methods were comparable. The correlation between final overall ratings for the two approaches and correlations between four of the six dimension ratings were significant and meaningful.

2. For the Affirmative Action Officer study, there were no significant differences between the oral and behavioral consistency methods for either dimension or overall reliabilities.

For the Sanitation District Manager study, there were significant differences in reliability between the two methods for one dimension with the oral reliability higher for that dimension. However, there were no significant differences in reliability for the overall ratings and the other five dimension ratings.

3. For the Affirmative Action Officer study, dimension correlations between the methods were low and non-significant, demonstrating lack of convergent validity. The discriminant validity criteria were not met either. Correlations between different methods measuring different traits were higher than correlations between different methods measuring the same traits, and correlations between the same methods measuring different traits were higher (considerably higher for the oral method) than the correlations between

different methods measuring the same traits. No consistent pattern was apparent across methods. Finally, many of the interdimension correlations for the oral method were higher than the respective dimension reliabilities.

Sanitation District Manager study results demonstrated convergent but not discriminant validity. Convergent validity was demonstrated by significant and meaningful correlations between the final overall ratings and four of the six dimension ratings. However, the criteria for demonstrating discriminant validity were not met. Correlations between different methods measuring different traits were higher than correlations between different methods measuring the same traits, and correlations between the same methods measuring different traits were higher than correlations between different methods measuring the same traits. No consistent pattern was apparent across methods. Finally, many of the interdimension correlations within method were higher than or equal to the respective dimension reliabilities except for three dimensions measured by the behavioral consistency approach.

4. For both studies, the two methods were equally acceptable to the raters but were not equally efficient in terms of rater time. Differences in rater time could not be tested for significance due to the absence of standard deviations for one or both approaches for both studies. Using descriptive statistics, the behavioral consistency method was more efficient than the oral method.

5. Based on descriptive comparisons, time efficiency for the candidate was in favor of the oral examination for both studies. (Differences in candidate time were not tested for significance because candidate time for the behavioral consistency method was not available for the Affirmative Action Officer study, and so was estimated, and a standard deviation for the behavioral consistency method was not available for the Sanitation District Manager study.) However, this time advantage could prove illusory if significant travel time were necessary for the oral examination or if additional application materials were necessary for the oral but not for the behavioral consistency examination.

Conclusions

Comparability

The first research question was whether the oral and behavioral consistency methods provided comparable ratings of applicants for employment. Study results were inconclusive for this question since correlations between the methods were significant and meaningful for one study but were non-significant for the other.

Reliability

The second research question was whether the oral and behavioral consistency methods were equally reliable. Results for this question were generally consistent across the two studies and indicated that there were no significant differences in reliability between the two methods for either the overall ratings or the dimension

ratings. There was one significant difference between dimension reliabilities for the two methods--for the Analyzing and Decision-making dimension in the Sanitation District Manager study, which demonstrated higher reliability for the oral method. However, the combined evidence from both studies seems to indicate that there was little difference in reliability for the two methods.

These data also indicate that the dimension and overall score reliabilities of both methods were generally adequate, comparing favorably to the reliabilities achieved for typical employee selection tests. The behavioral consistency reliabilities were also comparable to those achieved in other studies of this approach. For these two studies, the overall reliabilities were .81 and .86 for the behavioral consistency method and .84 and .92 for the oral method. Dimension reliabilities across the studies ranged from .54 to .88 for the behavioral consistency method and from .71 to .91 for the oral method. Although some of the individual dimension reliabilities were relatively low, they were acceptable because the dimension scores were part of a composite.

A final consideration in evaluating the dimension reliabilities has to do with their comparison to the other values in the monomethod blocks of the multitrait-multi-method matrices. Inspection of the monomethod blocks for the oral method revealed that, although the oral reliabilities were generally high, they were exceeded in

each case by at least one and usually more of the related

interdimension correlations. This leads to the conclusion

that the separate dimensions were not rated accurately and

to the supposition that the raters made global judgments of

the candidates, perhaps based on likeability, poise, or

interpersonal skill, and then based dimension scores on

their overall judgments rather than rating each dimension

separately.

This situation was not as true for the behavioral

consistency approach. In the Affirmative Action Officer

study, behavioral consistency dimension reliabilities were

generally moderate to low but exceeded the interdimension

correlations in every case. For the Sanitation District

Manager study, behavioral consistency dimension reliabili-

ties exceeded the related interdimension correlations in

only three of the six cases. This indicates that, for the

Sanitation District Manager examination, some global factor

such as writing skill may have been operating.

Based on the above analysis, the reliabilities of the

dimension ratings cannot be considered satisfactory for

either method. Dimension reliabilities for the oral

approach are questionable because the interdimension

correlations exceeded the respective reliabilities in every

case. This also happened for three of the six dimensions

measured by the behavioral consistency approach for the

Sanitation District Manager examination. The behavioral

consistency approach in the Affirmative Action Officer study

did not have this failing but its reliabilities were only moderate to low.

Convergent and Discriminant Validity

The third research question was whether the two approaches demonstrated convergent and discriminant validity. Results were inconclusive regarding convergent validity with the Affirmative Action Officer study demonstrating lack of convergent validity and the Sanitation District Manager study demonstrating possession of convergent validity. Neither of the studies demonstrated discriminant validity.

Rater Acceptability and Efficiency

The fourth research question was whether the oral and behavioral consistency methods were equally acceptable to the raters and equally efficient in terms of rater time. Results were consistent for the two studies with both methods demonstrating equal acceptability to the raters and the behavioral consistency method demonstrating superior efficiency in terms of rater time. However, the rater efficiency results were based on descriptive rather than inferential statistics. Statistical significance tests were not done due to the absence of standard deviations for rater time for one or both methods for both studies.

Candidate Efficiency

The last research question was whether the oral and behavioral consistency approaches were equally efficient in terms of candidate time. Candidate time was not available

for the behavioral consistency approach for the Affirmative

Action Officer study but based on the researcher's

experience and descriptive evidence from the Sanitation

District Manager study, the oral examination was more

efficient for candidates. (A statistical significance test

could not be done for the Sanitation District Manager study

due to the absence of a standard deviation for the

behavioral consistency approach.) However, this conclusion

was based only upon actual examination time for the

candidate. If travel or preparation time for either method

were significant, this conclusion would be affected.

Discussion

There has been limited research comparing the oral

interview to alternative selection devices. This study has

contributed to filling this gap by comparing the interview

to the behavioral consistency method, a relatively new type

of approach to the evaluation of training and experience.

The study also had the advantage of being conducted in an

applied setting as described in the introduction, with

subject matter experts as raters, job applicants as ratees,

oral interviews as the stimulus material, a continuum of

ratings as the outcome measure, and structured questions

based on job-related dimensions as the interview content.

The specific jobs selected for the study were typical low to

mid-management positions, one with an in-house candidate

pool and the other with a candidate pool based on nationwide

recruitment. Because this setting applies to many actual

selection situations, the study was more realistic than many of the typical studies conducted on selection interviews.

Although this study was based on two samples, its major limitation was small sample size and the resulting lack of power. The intention of the study was to compare the oral examination method with the behavioral consistency method. However, results regarding the primary question of comparability were inconclusive, perhaps due to the small sample size. Results showing only one significant difference in reliability between the methods may also have been affected by low power. A related limitation was the high alpha level. This was caused by doing separate significance tests for the dimensions in each study. However, the researcher was willing to accept a high alpha level in order to achieve greater power.

A third major problem centered on the issue of reliability. For the oral method in particular, dimension reliabilities were lower than related interdimension correlations. This was not as true for the behavioral consistency method but applied to three of the Sanitation District Manager dimensions. Given that there may have been legitimate relationships among the dimensions, this is still an unsatisfactory situation. Dimension ratings cannot be considered accurate if their correlations with other dimensions exceeds their correlations with themselves. This situation certainly affected the discriminant validity of

the methods and obscured what was being measured.

The study may also have been affected by several other factors, some of which were related to the above reliability problem: rater characteristics, dimension overlap, lack of behaviorally anchored rating scales, and limited administration time.

Of these problems, the most serious was rater characteristics. The previous discussion regarding the high intercorrelations among the dimensions within method suggested that some of the raters formed global judgments of the candidates and then rated the dimensions based on their overall impressions. This was borne out during the discussion and consensus session that occurred after the oral examination for Sanitation District Manager when it became apparent that two of the raters had misunderstood or disregarded the instructions and had made dimension ratings based on their overall judgment of the candidate rather than on their separate consideration of the candidate on each separate dimension. When completing his examination evaluation form, one of these raters protested that he preferred making global judgments about candidates rather than assessing them on specific dimensions.

This tendency may have been due to the limited training provided. However, since the training covered this issue, the researcher thinks it was more likely due to other factors. First, those raters who preferred the global method were experienced raters, having served on previous

oral examination panels which used the global method. These raters may have been unwilling to try a new method, preferring to use one with which they felt comfortable. Second, such raters may have been demonstrating field-dependency, a characteristic which, according to the Cardy and Kehoe study, would incline them to perceive things holistically rather than analytically. Finally, since these raters were unpaid subject matter experts rather than City employees, their accountability was relatively low and they may have felt little compunction about doing the ratings their way.

A second factor encountered in this study involved legitimate dimension overlap. The job analysis and test development for both studies were based on the assumption that the dimensions were relatively independent and could be assessed separately. However, it is more likely that at least some of the dimensions for the same job were closely related. This was evidenced in the retranslation phase of the Sanitation District Manager examination and in the high intercorrelations among dimensions for the oral examinations. These high intercorrelations among dimensions were probably caused by both legitimate relationships among the dimensions and rater unwillingness or inability to rate them separately. Except for the Analyzing/Decision-making dimension, low reliability and lack of comparability between methods for specific dimensions in the Sanitation District Manager study did not appear to be associated with dimension

definition problems encountered in the job analysis.

Another potential problem was the lack of behaviorally anchored rating scales for three of the four examinations. (There was a behaviorally anchored rating scale for the Affirmative Action Officer behavioral consistency examination.) The researcher made a case for including behaviorally anchored rating scales in a content validity model but, without large applicant samples or previous applicant answers, was unable to develop them in practice. The primary rationale for considering behaviorally anchored rating scales desirable was their supposed contribution to reliability. It stands to reason that having clear examples for various rating scale points would provide raters with common rating criteria and therefore would provide for more reliable ratings. However, the opposite was true. Interrater reliability for the Affirmative Action Officer behavioral consistency questionnaire was lower than interrater reliability for any of the other three examinations with respect to both overall scores and dimension scores. This evidence, along with evidence from the performance evaluation literature regarding the lack of superiority for behaviorally anchored rating scales, has caused the researcher to conclude that lack of behaviorally anchored rating scales did not significantly affect study results.

A final limitation of the study had to do with administration time. If the raters had had more time to

rate the candidates, at least reliability may have been improved. However, extension of administration time does not seem practical with the use of volunteer raters, which is a common practice in governmental agencies. This practice would have to be changed to one of using employees as raters or to one of using paid consultants as raters.

There were several differences between the two studies, some of which have been discussed above as limitations, which could potentially have accounted for the differences in findings regarding the primary question of comparability. These were differences in applicant group characteristics, rater characterisitics, job analysis methodology, and use of behaviorally anchored rating scales. The applicant groups differed considerably in degree of heterogeneity. The Affirmative Action Officer candidates were a nationwide group with diverse backgrounds while the Sanitation District Manager candidates were City employees occupying similar positions. The rater groups for the two studies differed in degree of expertise. Most of the Affirmative Action Officer raters had backgrounds in personnel management and were more knowledgeable although not necessarily more experienced than the Sanitation District Manager raters in the use of rating scales. Job analysis for the two studies differed with the Affirmative Action Officer study based on a task analysis and the Sanitation District Manager study based on a critical incident/retranslation of expectations approach. The last major difference between the studies was in the use

of behaviorally anchored rating scales. Both studies used modified behaviorally anchored rating scales for the oral approach, but the Affirmative Action Officer study employed behaviorally anchored rating scales for the behavioral consistency approach while the Sanitation District Manager study used a graphic rating scale for this approach.

These differences between the studies would provide a logical explanation for the differences in comparability results if the comparability differences were in the opposite direction. Except for the job analysis methodology, each of the differences between the studies (hetereogeneity of sample, rater expertise, and use of behaviorally anchored rating scales) would lead to the expectation that the oral and behavioral consistency approaches would be significantly correlated for the Affirmative Action Officer examination but not for the Sanitation District Manager examination. Because the opposite occurred, these differences did not appear to influence the study results, and the above analysis has not provided insight on the problem, leaving the researcher unable to furnish an explanation.

Implications for Future Research

The most obvious implication for future research is to replicate the study using a large sample of examination candidates. This would increase power and confidence in study results. Such a study should seek to increase the reliability of the behavioral consistency approach and the

accuracy of the separate dimension ratings. The usefulness of such a study would be increased if ratings from both approaches could be compared to job performance, enabling conclusions to be drawn regarding the empirical validity as well as the comparability of the methods.

Problems encountered in the study also point to directions for future research. Additional studies of rater characteristics, including field-dependency and experience, would be very useful to practitioners. These studies should compare subject matter experts rather than students on these factors. Studies comparing behaviorally anchored rating scales to other types of rating scales should be done for oral and behavioral consistency examinations as well as for performance evaluations.

In general, the previous research on oral interviews has not included careful development of an oral examination based on a content validity model. Interview content has been unspecified in many studies with candidates perhaps given global ratings based on different dimensions and different questions. Oral examination research in which the oral examination was developed according to a content validity model is still sorely needed.

APPENDICES

Appendices for Affirmative Action Officer

APPENDIX A

This Task Inventory consists of lists of task statements grouped into five major areas and three rating scales which will be used to determine which tasks are critical to the Affirmative Action Officer job.

1. *Please begin by detaching the* <u>Task Inventory Rating Scales</u> *page (last page of the inventory) from the rest of the inventory.*

2. *Apply Rating Scale A to each task statement, indicating how important each task is relative to other tasks to be performed. Use the task inventory answer sheet for Rating Scale A to record your response.*

3. *Apply Rating Scale B to each task statement, indicating the relative amount of time spent for each task. Use the task inventory answer sheet for Rating Scale B to record your response.*

4. *Apply Rating Scale C to each task statement, indicating the consequence of error which may result from inadequate or incorrect performance of each task. Use the task inventory answer sheet for Rating Scale C to record your response.*

5. *After rating each task on each rating scale, please consider whether any tasks are missing. Please include any missing tasks on the last page of the inventory.*

AFFIRMATIVE ACTION OFFICER TASK INVENTORY

AFFIRMATIVE ACTION PLAN ACTIVITIES

1. *Responsible for the development, implementation and annual updating of a comprehensive affirmative action plan for the City of Milwaukee.*

2. *Responsible for assisting in developing, amending or rejecting affirmative action plans of the various City departments and ensuring that each plan is consistent with the overall City affirmative action plan.*

3. *Responsible for reviewing the employment practices of the various City departments and evaluating their progress in meeting their affirmative action goals.*

4. *Reviews the City's affirmative action goals in terms of fairness and feasibility and makes recommendations to the Common Council and City Service Commission on proposed modifications.*

COMPLIANCE EXPERT ACTIVITIES

5. *Reviews all state and federal rules and regulations concerning equal employment opportunity to ensure that the City is in conformance.*

6. *Serves as subject matter expert and provides authoritative advice on fair employment practices.*

7. *Keeps current on affirmative action related legislation/court decisions and implements necessary changes in department policies/practices.*

8. *Serves as affirmative action advocate in Personnel Department and City.*

9. *Consults with legal counsel on matters related to EEO and/or affirmative action.*

10. *Testifies as an expert witness in court or by deposition regarding personnel practices.*

11. *Identifies problem areas in affirmative action.*

12. *Writes/rewrites personnel policies or procedures or suggestions for changes related to EEO/affirmative action.*

13. *Keeps current on affirmative action programs, policies and issues.*

AFFIRMATIVE ACTION STATUS AND ACCOUNTABILITY ACTIVITIES

14. *Maintains all necessary statistics such as the proportion of affected and underrepresented group members at all levels and job classifications in the City's workforce and the availability of affected and under-represented group members in the relevant labor force.*

15. *Presents information in written and oral form to CSC, MCCR, Finance Committee and other official bodies.*

16. *Responds to charges, challenges, complaints, allegations regarding the City's affirmative action plan.*

17. *Responsible for answering questions from a variety of sources on affirmative action.*

18. *Writes reports on EEO/affirmative action matters.*

19. *Responsible for computation of mathematical sums, averages, or percentages.*

20. *Designs new or modifies existing records management systems.*

21. *Responsible for verifying the accuracy of numerical data.*

22. *Selects, applies, and interprets statistical indices appropriate to the situation.*

AFFIRMATIVE ACTION PROJECT ACTIVITIES

23. *Responsible for receiving and investigating complaints of discriminatory employment practices from employees and prospective employees in con-junction with the City Attorney.*

24. *Develops and implements special programs such as departmental succession plans and career ladder plans to accelerate training and experience in underrepresented areas.*

25. *Responsible for designing and implementing training programs related to affirmative action.*

26. *Coordinates and oversees affirmative action related recruitment activities.*

27. *Provides various types of counseling to current and prospective employees and to City supervisors and managers.*

28. *Researches and remedies problems associated with retaining protected class members in the workforce.*

29. *Develops and implements new programs and policies to increase the employment and retention of minorities and women.*

30. *Supervises the Disabled Employees Placement Program.*

31. *Plans and develops recruiting networks of minorities and females.*

32. *Responsible for presentations before groups of potential minority and female applicants to explain job opportunities, requirements, procedures, etc.*

33. *Reviews resumes and/or makes reference checks.*

34. *Responsible for administering alternative selection devices/tests to handicapped applicants.*

35. *Supervises training needs assessment related to affirmative action.*

36. *Plans and budgets affirmative action projects and programs.*

37. *Evaluates the effectiveness of affirmative action programs.*

38. *Supervises the preparation of requests for proposals, evaluates proposals and specifications from vendors of consulting services, equipment, supplies, etc., and monitors affirmative action related contracts.*

39. *Identifies specific topics for research.*

40. *Writes research proposals and grant applications.*

41. *Works with Testing staff on recruitment and other issues pertaining to affirmative action.*

42. *Recommends or determines organizational or geographical area of competition based on affirmative action concerns.*

43. *Discusses the qualifications and/or suitability of minority/female candidates for positions to be filled with hiring managers or supervisors.*

SUPERVISORY/MANAGEMENT ACTIVITIES

44. *Supervises Affirmative Action Program staff--motivating, training, assigning work, evaluating, directing and disciplining.*

45. *Develops and monitors Affirmative Action Program budget.*

46. *Monitors work unit expenditures to insure overall compliance with budget.*

47. *Sets goals and objectives for employees in the work unit.*

48. *Assigns or adjusts work responsibilities of employees based on organizational needs, experience and competency of staff, developmental needs of staff, emergencies, and other factors.*

49. *Checks (monitors) the progress of work assignments periodically or at critical points to insure objectives and timetables are being met.*

50. *Reviews work products, correspondence, recommendations, and other written materials prepared by staff to insure that the quality is satisfactory, that policy is being followed or interpreted correctly, that they are technically correct, etc.*

51. *Evaluates work of employees against criteria identifying strengths and deficiencies in products, performance, or other dimensions of importance to the unit or organization.*

52. *Defines courses of action to correct deficiencies in performance and provides positive feedback and reinforcement for successful employee performance.*

53. *Counsels employees with regard to developmental objectives, career plans, promotional opportunities, etc.*

54. *Serves as a primary representative and communication link between the work unit and other work units in the department and City.*

55. *Prepares reports detailing work unit activities, program status, or reportable statistics for other work units, outside agencies, or management information.*

ADDITIONAL TASK STATEMENTS

If any Affirmative Action Officer tasks have been omitted from the inventory, please add them below.

TASK INVENTORY RATING SCALES

RATING SCALE A	RATING SCALE B	RATING SCALE C
Importance of task accomplishment to the function of the position.	Relative amount of time spent	Consequence of error
How important do you consider this task relative to other tasks in carrying out the major functions of the position?	What is the relative amount of time you would expect this task to take compared to other tasks?	What amount of harm would be done by poor or incorrect performance of the task?
Record response on the task inventory answer sheet for Rating Scale A.	Record response on the task inventory answer sheet for Rating Scale B.	Record response on the task inventory answer sheet for Rating Scale C.
1. Unimportant	1. Very small amount	1. Minimal or negligible harm done
2. Somewhat important	2. Small amount	2. Slight harm done
3. Moderately important	3. Average amount	3. Some harm done
4. Considerably important	4. Large amount	4. Somewhat serious harm done
5. Very important	5. Very large amount	5. Considerable harm done

APPENDIX B

This KSA Inventory consists of lists of KSAs grouped into seven major areas and three rating scales which will be used to determine which KSAs should be tested for the Affirmative Action Officer job.

1. *Please begin by detaching the KSA Inventory Rating Scales page (last page of the inventory) from the rest of the inventory.*

2. *Apply Rating Scale A to each KSA statement, indicating how important each KSA is relative to other KSAs. Use the KSA inventory answer sheet for Rating Scale A to record your response.*

3. *Apply Rating Scale B to each KSA statement, indicating the necessity for each KSA at time of hire. Use the KSA inventory answer sheet for Rating Scale B to record your response.*

4. *Apply Rating Scale C to each KSA statement, indicating which KSAs distinguish between effective and ineffective workers. Use the KSA inventory answer sheet for Rating Scale C to record your response.*

5. *After rating each KSA on each rating scale, please consider whether any KSAs are missing. Please include any missing KSAs on the last page of the inventory.*

AFFIRMATIVE ACTION OFFICER KSA INVENTORY

KNOWLEDGE OF AFFIRMATIVE ACTION

1. *Knowledge of affirmative action goal setting*

2. *Knowledge of affirmative action plans*

3. *Knowledge of labor force availability methods*

4. *Knowledge of state/federal affirmative action regulations*

5. *Knowledge of affirmative action information sources*

6. *Knowledge of City of Milwaukee affirmative action plans and policies*

7. *Ability to provide authoritative advice on fair employment practices on short notice*

8. *Knowledge of concepts and concerns relevant to affirmative action and equal employment opportunity in the public sector*

9. *Knowledge of/sensitivity to needs of disadvantaged*

10. *Knowledge of comparable worth*

11. *Ability to research affirmative action issues*

12. *Ability to recognize discriminatory practices (systemic)*

13. *Commitment to affirmative action*

KNOWLEDGE OF PERSONNEL MANAGEMENT

14. *Ability to design/implement training programs*

15. *Knowledge of recruitment methods*

16. *Ability to counsel*

17. *Knowledge of retention problems*

18. *Knowledge of personnel testing*

19. *Knowledge of human resources planning*

20. *Knowledge of labor relations*

21. *Knowledge of wage and salary administration, compensation systems*

22. *Knowledge of merit system concepts*

23. *Knowledge of impact of laws and other environmental influences on recruitment and selection*

24. *Knowledge of job classification/evaluation systems*

25. *Knowledge of ways of making job evaluation more relevant to employee and organizational productivity.*

26. *Knowledge of the legal environment of disciplinary actions and grievance procedures.*

27. *Knowledge of the relationship of disciplinary actions to other personnel activities.*

28. *Knowledge of internal personnel maintenance issues (promotions, transfers, demotions, terminations, lay-offs, retirements...)*

29. *Ability to construct and conduct surveys.*

PLANNING, ANALYSIS SKILLS

30. *Planning/organizing ability*

31. *Knowledge of descriptive statistics*

32. *Accuracy in performing mathematical computations*

33. *Ability to determine impact of decisions on others or other components of the organization*

34. *Skill in identifying problems, securing relevant information and identifying possible causes of problems*

35. *Ability to critically analyze issues and challenges*

36. *Ability to assess a course of action in terms of its long-range effects*

37. *Ability to draw valid conclusions from data*

38. *Ability to analyse problems and develop alternatives*

39. *Ability to meet deadlines*

40. *Ability to establish priorities*

41. *Ability to use data processing resources effectively*

42. *Ability to anticipate and solve problems*

43. *Ability to develop alternative solutions to problems*

DECISION-MAKING, JUDGMENT, INDEPENDENCE SKILLS

44. *Ability to work independently*

45. *Ability to stand up for ideas*

46. *Ability to make decisions within guidelines*

47. *Decisiveness: Readiness to make decisions, render judgments, take action, or commit oneself.*

48. *Ability to apply policies*

49. *Willingness to accept responsibility*

50. *Self-confidence*

51. *Judgment*

COMMUNICATION SKILLS/INTERPERSONAL SKILLS

52. *Oral presentation skills*

53. *Writing skills*

54. *Interpersonal/persuasive skills*

55. *Oral communication skill*

56. *Listening skill*

57. *Sensitivity*

58. *Ability to negotiate*

59. *Ability to relate to policy makers*

60. *Ability to read and understand written material*

61. *Impact*

62. *Leadership*

63. *Ability to receive information tactfully from abrasive people*

64. *Ability to testify effectively*

65. *Ability to see more than one side of an issue*

SUPERVISION SKILLS

66. *Knowledge of motivation and ability to motivate staff members*

67. *Ability to train staff*

68. *Knowledge of disciplinary methods, grievance procedures*

69. *Ability to delegate*

70. *Ability to develop quality/timeliness controls*

71. *Ability to evaluate work performance*

72. *Ability to develop and monitor budget*

73. *Knowledge of methods to increase productivity*

74. *Ability to assign work based on appropriate criteria*

75. *Knowledge of Management by Objectives*

76. *Ability to analyze reasons for poor or untimely work performance*

77. *Ability to give appropriate feedback to employees*

78. *Ability to counsel subordinates*

OTHER MANAGEMENT SKILLS

79. *Creativity*

80. *Ability to generate and support good ideas*

81. *Ability to perform under pressure/extreme stress*

82. *Ability to learn rapidly*

83. *Ability to work with frequent interruptions*

84. *Ability to accept constructive criticism*

85. *Ability to contribute effectively as a member of a team; ability to cooperate*

86. *Ability to act in accordance with high ethical and professional standards*

87. *Ability to work in a political system*

88. *High motivation for professional growth*

89. *Conscientiousness*

90. *Discretion*

91. *Initiative*

ADDITIONAL KSA's

If any Affirmative Action Officer KSA's have been omitted from the inventory, please add them below.

KSA INVENTORY RATING SCALES

RATING SCALE A	RATING SCALE B	RATING SCALE C
IMPORTANCE	NECESSITY AT TIME OF HIRE	DISTINGUISHING EFFECTIVE FROM INEFFECTIVE WORKERS
How important is the KSA in the performance of the job tasks?	How necessary for successful performance is the KSA at time of hire?	How useful is this KSA in distinguishing effective from ineffective workers?
Record response on the KSA inventory answer sheet for Rating Scale A.	Record response on the KSA inventory answer sheet for Rating Scale B.	Record response on the KSA inventory answer sheet for Rating Scale C.
1. Unimportant.	1. Not needed at time of hire.	1. Not useful.
2. Somewhat important.	2. Useful at time of hire.	2. Somewhat useful.
3. Moderately important.	3. Essential at time of hire.	3. Moderately useful.
4. Considerably important.		4. Considerably useful.
5. Very important.		5. Very useful.

APPENDIX C

Task Groups and Essential Tasks

Task Groups	Essential Tasks
A. Affirmative Action Plan Activities	1-4
B. Compliance Expert Activities	5-9, 11-13
C. Affirmative Action Status and Accountability Activities	14-19,21,22
D. Affirmative Action Project Activities	24,32,37
E. Supervisory/Management Activities	44,47-52,54,55

APPENDIX D

Revised Dimensions and Critical KSA's

Revised Dimensions	Critical KSA's
1. Knowledge of Affirmative Action	1,2,4,7,8,11,12,13,23
2. Planning and Organizing Skills	30,39,40,42
3. Analytical and Quantitative Reasoning Abilities	31,32,33,34,35,36,37, 38,43,60
4. Decision-Making, Judgment and Independence Skills	44-51
5. Oral Communication and Interpersonal Skills	52,54,55,56,57,58,59, 61,62,63,64,65
6. Written Communication Skills	53
7. Supervisory Skills	66,69,70,71,74,76,77
8. Initiative/Creativity/Intelligence	79,80,82,91
9. Toleration of Stress	81,83
10. Professionalism	84-90

APPENDIX E

Link up of Task Groups/KSA Dimensions

| | Task Groups* | | | | |
Dimensions	A	B	C	D	E
Knowledge of Affirmative Action	X	X	X	X	
Planning and Organizing Skills	X	X	X	X	X
Analytical and Quantitative Reasoning Abilities	X	X	X	X	X
Decision-Making, Judgment and Independence Skills	X	X	X	X	X
Oral Communication and Interpersonal Skills	X	X	X	X	X
Written Communication Skills	X	X	X	X	X
Supervisory Skills					X
Initiative/Intelligence/Creativity	X			X	
Toleration of Stress	X		X		
Professionalism	X	X	X	X	X

*Task Group A: Affirmative Action Plan Activities
 Task Group B: Compliance Expert Activities
 Task Group C: Affirmative Action Status and Accountability Activies
 Task Group D: Affirmative Action Project Activities
 Task Group E: Supervisory/Management Activities

Dimensions and KSA's Tested

Dimensions	KSA's
1. Knowledge of Affirmative Action	1,2,4,7,8,11,12,13,23
2. Planning and Organizing Skills	30,39,40,42
3. Analytical and Quantitative Reasoning Skills	31,32,33,34,35,36,37, 38,43
4. Oral Communication and Inter-personal Skills	52,54,55,56,57,58,59, 61,62,63,64,65
5. Written Communication Skill	53
6. Supervisory Skills	66,69,70,71,74,76,77

APPENDIX G

DIMENSIONS

I. Knowledge of Affirmative Action

The Affirmative Action Officer must be able to develop affirmative action goals and plans. He or she must have current knowledge of state and federal laws, regulations and court cases related to affirmative action and be able to provide authoritative advice on fair employment practices on short notice. In addition, he or she must be knowledgeable regarding concepts and concerns relevant to affirmative action and equal employment opportunity in the public sector and be able to recognize discriminatory practices and to research affirmative action issues. The Affirmative Action Officer must be committed to affirmative action.

II. Planning and Organizing Skills

The Affirmative Action Officer must be able to plan and organize work effectively. This involves the ability to establish a situationally appropriate plan or course of action for one's self and others to attain specific goals or accomplish defined tasks. It involves the ability to employ a systematic approach to the work, to set meaningful priorities, to manage time and resources effectively and to meet established deadlines.

III. Analytical and Quantitative Reasoning Abilities

The Affirmative Action Officer must have analysis and quantitative reasoning skills. This involves the ability to identify problems, secure relevant information and identify possible causes of problems; the ability to develop alternative solutions to problems; the ability to assess a course of action in terms of its long-range effects and to determine the impact of decisions on others or other components of the organization. It also involves accuracy in performing mathematical computations, knowledge of descriptive statistics, the ability to draw valid conclusions from data, and the ability to relate and compare data from a variety of sources.

IV. Oral Communication and Interpersonal Skills

The Affirmative Action Officer must be able to effectively and accurately communicate ideas and information to others orally in both formal and informal one-to-one and group situations, extemporaneously or with prior preparation.

In addition, he or she must possess interpersonal and persuasive skills. This involves the ability to interact effectively with peers, superiors and subordinates, staff of other departments and members of the various "publics" served. It also involves the ability to see more than one side of an issue, to work cooperatively, to adapt one's behavior to enable the effective pursuit of goals despite obstructions presented by conflicting attitudes, opinions or the action of others and the ability to generate the trust and confidence needed to obtain agreement or acceptance with respect to ideas, plans or programs.

V. Supervisory Skills

The Affirmative Action Officer must be able to supervise both professional and clerical staff members. He or she must be able to delegate work and to assign work based on appropriate criteria, to develop quality and timeliness controls, to motivate staff members and evaluate work performance, to analyze reasons for poor or untimely work performance and give appropriate feedback to employees.

CITY OF MILWAUKEE

ACHIEVEMENT HISTORY QUESTIONNAIRE

FOR

AFFIRMATIVE ACTION OFFICER

City
of
Milwaukee

NAME _____

ADDRESS _____

PHONE (DAY) _____

(EVENING) _____

NOTE: Deadline for return of this questionnaire is Tuesday, February 12, 1985 in order to be guaranteed consideration by the rating panel.

GENERAL INSTRUCTIONS

<u>PLEASE NOTE</u>: It is very important to read these instructions very carefully before you begin to complete the questionnaire.

1. Please answer the sets of questions which appear on the attached pages. Each set of questions relates to an important dimension of the Affirmative Action Officer position. The questions ask you to describe what you consider to be your major <u>achievement(s)</u>which would demonstrate that you possess the job-related knowledge, skills or abilities identified. In other words, we are looking for concrete examples of things you have actually done or accomplished pertaining to each dimension rather than a general description of your skills. These achievements may be either specific incidents or examples of sustained high performance over a period of time. An example of an achievement demonstrating skills related to another dimension is attached.

2. You should select your strongest achievements regardless of where they were attained. You need not restrict yourself to those related to affirmative action positions.

3. If you cannot have your responses typewritten, please write as neatly as possible in black ink. Attach additional pages as necessary. Your responses will be photocopied for distribution to the panel of raters.

4. Do <u>not</u> substitute your resume for any responses to the questionnaire. All responses must be in the format provided in order to ensure a fair evaluation.

5. You are asked to describe <u>one</u> achievement for each of the five dimensions in the questionnaire; however, if you would like, you may describe more than one. If so, attach a separate sheet with the same format.

6. Return your completed questionnaire to:

> Timothy J. Keeley
> City of Milwaukee Personnel Department
> Room 706, City Hall
> 200 East Wells Street
> Milwaukee, WI 53202

7. Questionnaires must be <u>received</u> in our office (not postmarked) by
   Tuesday, February 12, 1985 to be guaranteed consideration by the rating panel.

<u>PLEASE READ AND SIGN THE FOLLOWING STATEMENT AND RETURN WITH THE QUESTIONNAIRE</u>

    I certify that all information provided herein relating to my own achievements and experience is true to the best of my knowledge, and that the information can be verified through persons I have so listed in the questionnaire. I understand that falsification of information may result in disqualification or removal from a City position.

DATE _____

SIGNATURE _____

EXAMPLE

This is an example of an achievement demonstrating skills related to the training dimension for another position (that of Personnel Officer). You do <u>not</u> need to describe an achievement for this dimension.

Example Dimension: Training Skills

The Personnel Officer must be able to plan and implement training programs for groups of employees as well as orient new employees to their jobs.

1. Describe an <u>achievement</u> which would show that you have the knowledges, skills or abilities described above. Tell us <u>what you actually did,</u> and include the objective or the problem.

While working as a personnel administrator for the State of Ohio, I was responsible for supervising a federally funded project whose objective was to provide assistance to State and local governments in the area of personnel testing. This assistance was provided through training seminars and workshops as well as research projects.

The objective of the training function was to determine training needs of local jurisdictions, develop training programs to meet these needs and implement the training programs.

The project began with a needs assessment survey of local jurisdictions throughout Ohio. Based on the survey, we determined that two sets of training programs were needed - one on oral examinations and legal guidelines for testing and the other on job analysis and test development.

Development of the seminar on oral examinations and legal guidelines was carried out by a staff of four personnel management specialists under my direction. We used several techniques to present the material including lecture, tryout, group discussion and role-playing. This was a one day seminar given in five locations across the state. I was one of five presenters.

Development of the seminar on job analysis and test development was carried out by myself and one of my staff members. The primary techniques used were lecture, tryout and group discussion. This was a one day seminar given in one location. I was one of two presenters.

2. What was the outcome or result?

   *The oral examination and legal guidelines seminars were attended by groups
   varying in size from 10 to 25. The job analysis seminar drew a group
   of 75.*

   *Evaluations from seminar participants showed that they considered the
   seminars useful, informative and well-presented.*

3. Was this achievement entirely attributable to you? Yes _____ No _x___

   If no, what is the estimated percentage of this achievement which is due
   to the efforts of other people (excluding clerical staff)? ____60____
   *However, this was all done under my supervision.*

4. (a) When did this achievement take place (approximate date)?
    *The oral examination seminars were given in April and May of 1975. The job
    analysis seminar was given in September of 1975. Seminar development took
    place throughout the Spring and Summer of 1975.*

   (b) For what employer? _____ *State of Ohio* _____

   _____

5. Please give the name, address and telephone number of someone who can
   verify this information.

   *Joseph Jones
   Ohio Department of
   Administrative Services
   P. O. Box 30007
   Columbus, Ohio 43215
   (614) 466-1567*
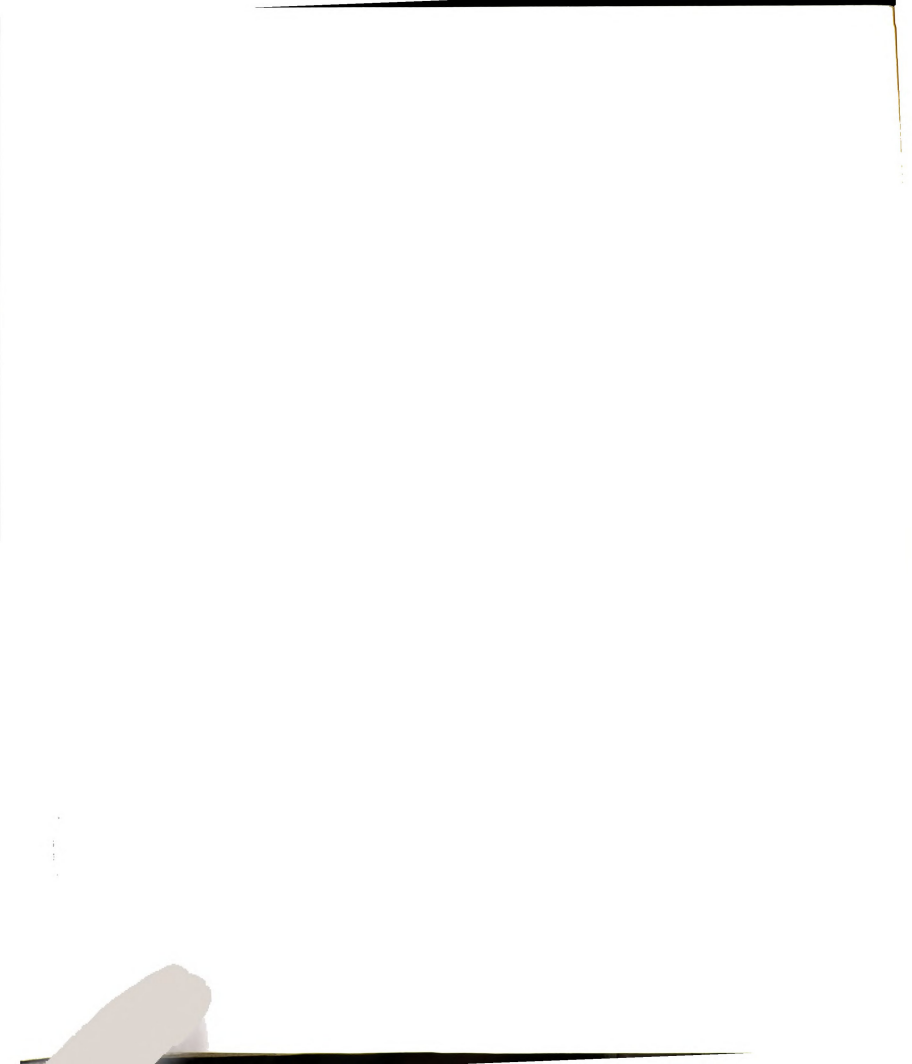
I. Knowledge of Affirmative Action

The Affirmative Action Officer must be able to develop affirmative action goals and plans. He or she must have current knowledge of state and federal laws, regulations and court cases related to affirmative action and be able to provide authoritative advice on fair employment practices on short notice. In addition, he or she must be knowledgeable regarding concepts and concerns relevant to affirmative action and equal employment opportunity in the public sector and be able to recognize discriminatory practices and to research affirmative action issues. The Affirmative Action Officer must be committed to affirmative action.

1. Describe an achievement which would show that you have the knowledges, skills or abilities described above. Tell us what you actually did, and include the objective or the problem.

2. What was the outcome or result?

3. Was this achievement entirely attributable to you?   Yes _____   No _____

   If no, what is the estimated percentage of this achievement which is due
   to the efforts of other people (excluding clerical staff)? _____

4. (a) When did this achievement take place (approximate date)?

   _____

   (b) For what employer? _____

   _____

5. Please give the name, address and telephone number of someone who can
   verify this information.

II. Planning and Organizing Skills

The Affirmative Action Officer must be able to plan and organize work effectively. This involves the ability to establish a situationally appropriate plan or course of action for one's self and others to attain specific goals or accomplish defined tasks. It involves the ability to employ a systematic approach to the work, to set meaningful priorities, to manage time and resources effectively and to meet established deadlines.

1. Describe an <u>achievement</u> which would show that you have the knowledges, skills or abilities described above. Tell us <u>what you actually did</u>, and include the objective or the problem.

2. What was the outcome or result?

3. Was this achievement entirely attributable to you?   Yes _____   No _____

   If no, what is the estimated percentage of this achievement which is due
   to the efforts of other people (excluding clerical staff)? _____

4. (a) When did this achievement take place (approximate date)?

   _____

   (b) For what employer?   _____

   _____

5. Please give the name, address and telephone number of someone who can
   verify this information.

III. Analytical and Quantitative Reasoning Abilities

The Affirmative Action Officer must have analysis and quantitative reasoning skills. This involves the ability to identify problems, secure relevant information and identify possible causes of problems; the ability to develop alternative solutions to problems; the ability to assess a course of action in terms of its long-range effects and to determine the impact of decisions on others or other components of the organization. It also involves accuracy in performing mathematical computations, knowledge of descriptive statistics, the ability to draw valid conclusions from data, and the ability to relate and compare data from a variety of sources.

1. Describe an <u>achievement</u> which would show that you have the knowledges, skills or abilities described above. Tell us <u>what you actually did</u>, and include the objective or the problem.

2. What was the outcome or result?

3. Was this achievement entirely attributable to you?   Yes _____   No _____

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people (excluding clerical staff)? _____

4. (a) When did this achievement take place (approximate date)?

   _____

   (b) For what employer? _____

   _____

5. Please give the name, address and telephone number of someone who can verify this information.

IV. Oral Communication and Interpersonal Skills

The Affirmative Action Officer must be able to effectively and accurately communicate ideas and information to others orally in both formal and informal one-to-one and group situations, extemporaneously or with prior preparation.

In addition, he or she must possess interpersonal and persuasive skills. This involves the ability to interact effectively with peers, superiors and subordinates, staff of other departments and members of the various "publics" served. It also involves the ability to see more than one side of an issue, to work cooperatively, to adapt one's behavior to enable the effective pursuit of goals despite obstructions presented by conflicting attitudes, opinions or the action of others and the ability to generate the trust and confidence needed to obtain agreement or acceptance with respect to ideas, plans or programs.

1. Describe an achievement which would show that you have the knowledges, skills or abilities described above. Tell us what you actually did, and include the objective or the problem.

2. What was the outcome or result?

3. Was this achievement entirely attributable to you?   Yes _____   No _____

   If no, what is the estimated percentage of this achievement which is due
   to the efforts of other people (excluding clerical staff)? _____

4. (a) When did this achievement take place (approximate date)?

   _____

   (b) For what employer? _____

   _____

5. Please give the name, address and telephone number of someone who can
   verify this information.

V. Supervisory Skills

The Affirmative Action Officer must be able to supervise both professional and clerical staff members. He or she must be able to delegate work and to assign work based on appropriate criteria, to develop quality and time-liness controls, to motivate staff members and evaluate work performance, to analyze reasons for poor or untimely work performance and give appropriate feedback to employees.

1. Describe an achievement which would show that you have the knowledges, skills or abilities described above. Tell us what you actually did, and include the objective or the problem.

2. What was the outcome or result?

3. Was this achievement entirely attributable to you?   Yes _____   No _____

    If no, what is the estimated percentage of this achievement which is due
    to the efforts of other people (excluding clerical staff)? _____

4. (a) When did this achievement take place (approximate date)?

    _____

    (b) For what employer? _____

    _____

5. Please give the name, address and telephone number of someone who can
    verify this information.

APPENDIX I

Rater Evaluations

We appreciate your serving as a rater for the Affirmative Action Officer examination. Your answers to the following questions will be used to improve our selection process.

1. Do you consider it fair to use accomplishment ratings to determine which applicants will participate in an oral examination?

    A. Very fair   D. Unfair
    B. Fair   E. Very unfair
    C. Somewhat fair

2. Do you consider it fair to use accomplishment ratings to determine final rankings in the examination process?

    A. Very fair   D. Unfair
    B. Fair   E. Very unfair
    C. Somewhat fair

3. How effective do you consider the accomplishment rating process in determining the best qualified applicants to be called for an oral examination?

    A. Very effective   D. Ineffective
    B. Effective   E. Very ineffective
    C. Somewhat effective

4. How effective do you consider the accomplishment rating process in determining the best qualified applicants for a job?

    A. Very effective   D. Ineffective
    B. Effective   E. Very ineffective
    C. Somewhat effective

5. How difficult did you consider the accomplishments to rate?

    A. Very easy   D. Difficult
    B. Easy   E. Very difficult
    C. Somewhat difficult

6. How reasonable do you consider the time you spent rating accomplishments for a position of this level and type?

    A. Very reasonable   D. Unreasonable
    B. Reasonable   E. Very unreasonable
    C. Somewhat reasonable

APPENDIX J

Rater Evaluation

We appreciate your serving as a rater for the Affirmative
Action Officer oral examination. Your answers to the
following questions will be used to improve our selection
process.

1. Do you consider it fair to use oral dimension ratings to
   determine final rankings in the examination process?

   A. Very fair   D. Unfair
   B. Fair   E. Very unfair
   C. Somewhat fair

2. How effective do you consider the oral dimension rating
   process in determining the best qualified applicants for
   a job?

   A. Very effective   D. Ineffective
   B. Effective   E. Very ineffective
   C. Somewhat effective

3. How difficult did you consider the dimensions to rate?

   A. Very easy   D. Difficult
   B. Easy   E. Very difficult
   C. Somewhat difficult

4. How reasonable do you consider the time you spent
   rating dimensions for a position of this level and type?

   A. Very reasonable   D. Unreasonable
   B. Reasonable   E. Very unreasonable
   C. Somewhat reasonable

5. Please add any comments in the space below.

Significance Tests for Differences Between Oral and
Behavioral Consistency Reliability Coefficients for
Dimension and Overall Reliabilities for the Affirmative
Action Officer Study

	r_1	r_2	r_{12}	W	t
D1	.82	.64	.13	2.00	1.43
D2	.83	.72	.04	1.65	1.01
D3	.82	.79	.18	1.17	.32
D4	.76	.64	.08	1.50	.82
D5	.71	.75	.34	.86	-.32
O	.84	.81	.17	1.19	.35

Note. D1 = Knowledge of Affirmative Action; D2 =
Planning and Organizing; D3 = Analytical and Quantitative
Reasoning Abilities; D4 = Oral Communication and
Interpersonal Skills; D5 = Supervisory Skills; O = Overall
Score; column definitions are as follows:

r_1 = oral reliability,

r_2 = behavioral consistency reliability,

r_{12} = correlation between oral and behavioral
consistency ratings,

W = $(1-r_2)/(1-r_1)$,

t_{N-2} = $(W-1)(N-2)^{1/2} / [4W(1-r_{12}^2)]^{1/2}$, and

N = sample size;

none of the t values were significant (p > .10, df = 16,
two-tailed).

Appendices for Sanitation District Manager

APPENDIX L

INSTRUCTIONS FOR THE

SANITATION DISTRICT MANAGER

CRITICAL INCIDENT INVENTORY

This Critical Incident Inventory consists of lists of critical incidents grouped into six major areas. Determine how effective each incident is compared to other incidents in carrying out the major functions of the position.

Please rate the effectiveness of each critical incident using the following scale.

1. extremely ineffective
2. considerably ineffective
3. moderately ineffective
4. somewhat effective
5. moderately effective
6. considerably effective
7. extremely effective

Record your rating in the space next to each incident on the inventory.

Please rate each incident even if it duplicates or closely resembles another incident. After rating each incident, please consider whether you would like to add any incidents. If so, please include them on the last page of the inventory.

SANITATION DISTRICT MANAGER
CRITICAL INCIDENTS
PLANNING AND ORGANIZING

_____ 1. Assigns all regular sanitation supervisors to work during the initial wave of a snow emergency thereby not having any supervisors to staff interim operations.

_____ 2. Updates and routes weed-cutting locations for mowers and consolidates a list for mowers to follow.

_____ 3. Anticipates need for pick-up of garbage if special occasion occurs.

_____ 4. Assigns and requests appropriate equipment for bulky item collection.

_____ 5. Holds informal meetings periodically for district staff to plan seasonal needs.

_____ 6. Gives inaccurate indication of supply needs.

_____ 7. Unavailable when staff meetings with the area manager are necessary to plan manpower needs during the winter resulting in inadequate manpower and infrequent collections.

_____ 8. Has up-to-date list of streets prone to develop drifting problems.

_____ 9. Uses a cluster technique to insure all carts in district are serviced without overtime. (Crews that finish daily route assist crews that are behind schedule.)

_____ 10. Assigns, moves and reassigns equipment and/or manpower as needed.

_____ 11. Ignores routing for bulky item collection.

_____ 12. Does not attempt to reduce manpower in using the end-loader and hopper for leaf collection.

_____ 13. Does not adequately plan manpower needs or adjust to shortage of collectors.

_____ 14. Makes up the winter supervisor duty roster ensuring that holiday assignments are shared equally and that no supervisor has two duty weeks in a row.

_____ 15. Balances supervisors' assignments to avoid hardship, overworking.

_____ 16. Establishes inappropriate priorities and policies for vehicle and operator assignment.

_____ 17. Develops weed route books and maps indicating salt routes and wall maps indicating limits of collection routes.

_____ 18. Does not set up or enforce good office procedures in district office.

_____ 19. Does not have maps of salt or plow routes in district and the individual route sheets are not up to date.

_____ 20. Does not follow pre-planned snow operation, assigning snowplow drivers to routes in the order in which routes were numbered and drivers arrived.

_____ 21. Does not establish policies for vehicle and operator assignment.

_____ 22. Balances overtime hours of supervisors.

_____ 23. Maintains and updates an efficient, accurate, up-to-date routing system, reviewed annually (or as need develops).

_____ 24. Establishes and maintains an accurate filing system and record keeping procedure in district headquarters.

_____ 25. Does not anticipate clean-up assignments of equipment, personnel and material delivery after snow storm.

_____ 26. Routing system for ice control is inefficient or not current.

_____ 27. Anticipates and sets up clean-up assignments of equipment, personnel and material delivery after snow storm.

_____ 28. Revises filing system for cart chits by filing all chits by quarter section numbers.

_____ 29. Does not maintain accurate records due to inaccurate documentation or ignoring records.

_____ 30. Ensures adequate staffing for collection operations during daytime salting operations.

_____ 31. Establishes priorities and policies for assignment of vehicles and operators according to expertise and capability.

_____ 32. Exceeds the allotment of scheduled vacations causing a shortage of manpower by granting an unscheduled vacation request.

_____ 33. Utilizes resources across district to ensure that weekly collection schedules are met.

_____ 34. Gradually reduces manpower in the fall to avoid unnecessary costs for the Bureau.

_____ 35. Ignores balancing overtime hours or inappropriately balances hours.

ANALYZING & DECISION MAKING

_____ 36. Utilizes resources available to solve collection problem.

_____ 37. Investigates complaint regarding a damaged metal bushel basket and finds crew to be innocent. Follows through by contacting complainant.

_____ 38. Appropriately assigns bulky item collection to regular residential crews.

172

_____ 39. Does not evaluate effectiveness of ice control operations or utilize tools at his disposal.

_____ 40. Inappropriately responds to an emergency snow plowing situation by not providing information or a course of action to be taken at initial phone call to his home.

_____ 41. Orders leaf equipment before there are enough leaf piles to use the equipment for a full day to ensure having the equipment at times of greater demand.

_____ 42. Acts immediately to clear a private parking lot at Children's Hospital due to emergency situation, permitting a helicopter to land.

_____ 43. Takes initiative to address emergency ice control situation immediately.

_____ 44. Calls superintendent at home to recommend plowing due to developing poor conditions after the area manager was informed and rejected recommendation.

_____ 45. Assigns a county worker to help the yardman collect bulky items in order to bring the large number of bulky item requests down to a manageable limit.

_____ 46. Analyzes why progress in collection is not effective (volume, manpower and equipment).

_____ 47. Orders crews to plow snow while salting the streets on a weekend when there is a low salt supply and snow from a previous storm is present.

_____ 48. Reviews plow routes and eliminates several streets from the "Mains" list that no longer carry traffic to justify priority treatment.

_____ 49. Formulates a selection method, has it approved by the superintendent and makes a recommendation for appointment following a newly established procedure in filling a job.

_____ 50. Ensures that all citizen hardship requests are handled in person and not over the phone.

_____ 51. Assigns a supervisor to an Alternate Plowing Situation who is unfamiliar with the problem and cannot be relied upon, resulting in trucks being used incorrectly and problems not being addressed.

_____ 52. Denies request for use of a bulky collection packer to help a crew which had its regular packer break down for one hour.

_____ 53. Does not pre-check bulky collection items (just assigns items to be picked up).

_____ 54. Drives through district on a daily basis to determine where problem areas exist and what might have occurred overnight that needs immediate attention, e.g. litter problems.

_____ 55. Waits for direction from main office regarding ice control operations.

_____ 56. Takes appropriate follow-up measures regarding an elderly person who could not move brush to proper location.

_____ 57. Puts unresolved problems on hold.

_____ 58. Grants request of additional snow equipment without delay during an ice control operation.

_____ 59. Checks field conditions first hand and orders out heavy equipment to open a subdivision cut off by wind-driven snow.

_____ 60. Interprets litter policy literally; sends trucks to empty containers already empty.

_____ 61. Ignores policy on Friday litter; lets trash accumulate.

_____ 62. Makes no recommendations or inappropriate recommendations regarding operational needs during an ice control operation.

_____ 63. Utilizes resources in order to reduce a growing number of bulky collection requests by using regular garbage collection crews to make two furniture stops daily while the bulky collection truck handles brush pick-ups.

_____ 64. Is unable to make decisions when necessary due to procrastination or letting problem take care of itself.

_____ 65. Makes reliable recommendations for operational needs, e.g. equipment, during ice control operations.

_____ 66. Ignores ineffective collection methods and does not attempt to improve them.

_____ 67. Does not physically check status of buildings and issues free carts to multiple unit buildings.

_____ 68. Transfers men and equipment where needed to make up for deficiencies in manpower in order to maintain schedule.

_____ 69. Pre-checks bulky material to be collected to ensure appropriate action.

_____ 70. Designs Hardship Factor for cart hardships determining how long it takes to service a hardship vs. a regular cart collection.

_____ 71. Based on Aldermanic complaint, dispatches truck and helper to plow street without investigating, resulting in unnecessary overtime since street had already been plowed.

_____ 72. Begins use of a hopper and end loader for the leaf program rather than hand or vac pickups to increase efficiency.

_____ 73. Ensures that regular crew picks up special collection when warranted.

_____ 74. Ignores requests, does not respond immediately, or action taken is incorrect.

_____ 75. Orders out men and equipment for weekend work to clean up after a parade on his own authority when he could not reach and get approval of area manager or a higher level supervisor.

_____ 76. Routinely checks district for potential problems using interaction with employees and subordinate supervisors.

_____ 77. Recommends trial use of log loader trucks in the curb pick-up of brush to increase efficiency.

_____ 78. Requests, schedules, reassigns equipment and manpower as conditions warrant during snow plowing.

_____ 79. Removes all Dead Ends and Cul-De-Sacs from regular plow routes and places them on special end-loaders' routes for plowing, resulting in too much work for end-loaders.

_____ 80. Evaluates/critiques individual salting (plowing) operations in terms of time, quality and complaints.

_____ 81. Takes appropriate action when discovers complaint being mishandled.

_____ 82. Makes recommendations for policy change to area manager.

_____ 83. Sends inappropriate truck to pick up bulky collection items.

_____ 84. Does not move equipment and manpower when needed.

_____ 85. Investigates aldermanic complaint regarding weeds before acting and finds complaint to be unjustified, the result of a feud between neighbors.

_____ 86. Allows personnel to act alone in emergency situation.

_____ 87. Recommends that two more salt trucks be parked at the District Headquarters, making use of two empty stalls.

_____ 88. Handles complaints uniformly.

_____ 89. Fails to investigate breakdown of cart collection truck and requests unnecessary replacement for minor breakdown.

_____ 90. Waits until main office gives approval to address emergency ice control situation.

_____ 91. Does not make any bulky item assignments to residential crews.

_____ 92. Learns about a new area not by hearsay, but rather through actual observation to determine what is going on in the area.

_____ 93. Addresses any problem or complaint that is still unresolved the day after a snow operation.

_____ 94. Pulls plows off a route where the main streets were complete and moves them to a route that has equipment shortages.

_____ 95. Does not order clean up plows to handle snow islands, complaints and missed streets the day after a storm.

COMMUNICATION/INTERPERSONAL SKILL

_____ 96. Fails to follow-up on a citizen complaint, not informing the complainant of the legitimate reason for the lack of collection and the appropriate action that will be taken.

_____ 97. Does not participate during a brainstorming session related to improving bureau services.

_____ 98. Communicates problems to superiors.

_____ 99. Does not correct misinformation being given out.

_____ 100. Maintains good relations and communication with other bureaus during snow operations.

_____ 101. Informs Alderman of results regarding a citizen/Aldermanic complaint.

_____ 102. Fails to pass along an ice control alert to the next district manager.

_____ 103. Informs main office of changing conditions.

_____ 104. Promises to handle complaint and get back to person.

_____ 105. Follows up on a complaint handed up from the route supervisor regarding crews setting out cans at the alley line ahead of the truck. Compromises with the complainant by not setting out cans in her block and notifies supervisor of the complaint and action taken.

_____ 106. Does not meet with supervisors at all or infrequently.

_____ 107. Warns higher level supervisors of problem that was not solved regarding a citizen that refused to move garbage to curb.

_____ 108. Inappropriately responds to an Aldermanic service request by not explaining why the City does not plow alleys and not suggesting any alternatives.

_____ 109. Responds in writing to Aldermanic service request, clearly and concisely, and to the heart of the problem.

_____ 110. Becomes involved in staff meeting by voicing opinions and informing others what is going on in the field.

_____ 111. Explains that everyone in neighborhood has same problem when answering complaint.

_____ 112. Informs main office of action taken in emergency situations.

_____ 113. Establishes poor relations/communications with other bureaus or departments.

_____ 114. Makes timely and appropriate response to Aldermanic requests during storm and emergency situations.

_____ 115. Introduces himself/herself to the employees and supervisors when going into the new area and lets them know what he/she expects.

_____ 116. Does not have open communication with field supervisors and does not inform field supervisors of changes in various procedural matters.

_____ 117. Conveys incorrect information to subordinate supervisors regarding how to conduct the survey prior to the installation of carts.

_____ 118. Maintains an open door policy to all employees to allow them to come and speak freely.

_____ 119. Examines and writes down critical complaints (cart-related, property damage, rude behavior) from other City bureaus/departments, Aldermen, area manager.

_____ 120. Subordinate supervisors do not know what to expect or are unsure how to proceed.

_____ 121. Does not take notes during the monthly District-Area Managers meeting, thus not communicating the essence of the meeting to subordinate supervisors.

_____ 122. Returns a call in a timely manner to a citizen or an Alderman on a matter relating to a Sanitation practice or procedure.

_____ 123. Fails to inform upper management that crews are unable to collect garbage due to unpassable alleys.

_____ 124. Conducts monthly meetings with supervisors in the district insuring broader downward communication.

_____ 125. Stifles communication with subordinates and has a poor listening ability.

_____ 126. Tactfully handles a citizen complaint regarding the use of carts for garbage collection.

_____ 127. During an ice control, notifies headquarters that a street has drifted shut and he has taken action to plow.

_____ 128. Does not inform main office of action taken during an ice control.

_____ 129. Calls Water Department and main office in case of water main breaks.

_____ 130. Allots inadequate number of vacations/weeks due to lack of communication.

_____ 131. Does not communicate ideas for policy change to area manager.

_____ 132. Holds a "pre-season" meeting with staff to discuss procedures and M.O.'s.

_____ 133. Meets with supervisors when new routes are being changed over in the area to permit as few complications as possible.

_____ 134. Takes detailed notes at monthly staff meeting to ensure that information will be relayed accurately to subordinate supervisors.

SUPERVISION

_____ 135. Spot-checks routing for collection of bulky items.

_____ 136. Maintains inventory control of supplies.

_____ 137. Does not ensure that crew picks up special collection when warranted.

_____ 138. Meets with subordinate supervisors regularly.

_____ 139. Backs up and reinforces subordinate supervisors' judgement when questioned (if appropriate).

_____ 140. Checks salt application rates (# of lbs./mile).

_____ 141. Does not allow subordinate supervisors to make decisions on their own.

_____ 142. Does not reprimand or send home an intoxicated employee brought into District Headquarters by employee's supervisor.

_____ 143. Ensures that litter cans are empty on Friday.

_____ 144. Shows a lack of trust and respect in subordinates. "Subordinates are treated as pawns to be used-discarded!"

_____ 145. Checks daily progress reports.

_____ 146. Investigates citizen complaint regarding abusive language of crew member before giving disciplinary action.

_____ 147. Sits in the office and does not get involved in the operation during a plowing operation.

_____ 148. Assigns plowing routes to drivers he favors, resulting in drivers being unfamiliar with the streets and streets remaining unplowed.

_____ 149. Allows collection routes to fall short of weekly collection objectives.

_____ 150. Does not ask for/require recommendations from line supervisors regarding disciplinary actions.

_____ 151. Delegates responsibility for snow plowing operations or goes home.

_____ 152. Reprimands a subordinate supervisor for referring to a female employee in a derogatory manner.

_____ 153. Fails to monitor salt application rates.

_____ 154. Delegates responsibility of reworking salt routes to line supervisors, giving them guidelines and monitoring the operation.

_____ 155. Makes periodic checks during crew coffee and lunch breaks to determine if bureau policy in this area is being carried out.

_____ 156. Makes adjustments in salt application rates as necessary.

_____ 157. Inappropriately advises employee to file a grievance even though employee's action clearly violates bureau policies.

_____ 158. Instructs supervisors to inform all employees of job opening and especially encourages minorities to apply.

_____ 159. Follows up on an assignment delegated to a subordinate to make sure that the job was completed and done properly.

_____ 160. Does not check daily progress reports.

_____ 161. Does not refer employee to the Employee Assistance Program, ignoring obvious employee problems that he was aware of.

_____ 162. Justifies and receives bureau permission to issue "Favorable Occurances" to deserving collection crews.

_____ 163. Ensures maintenance of accurate records during snow plowing operations (e.g. drivers' time sheets, payroll records, emergency personnel hours).

_____ 164. Violates union contract by having non-union personnel perform duties done regularly by union personnel.

_____ 165. Allows a Sanitation crew to take a mid-morning break less than one hour after starting time.

_____ 166. Ensures that procedures are consistently and uniformly applied.

_____ 167. Asks employee if he would be able to move the time of a dental appointment to late afternoon to avoid loss of his services.

_____ 168. Corrects the directions of a subordinate supervisor which created an unsafe situation for the collection crew.

_____ 169. Ensures that the schedule for biweekly maintenance checks for garbage trucks is in the truck, in the immediate supervisor's route book and on the District Manager's calendar.

_____ 170. Leaves meeting weekly collection schedules to subordinate supervisors.

_____ 171. Does not follow a course of progressive discipline, allowing an employee who was 10 minutes late to work to begin working without giving him any disciplinary action.

_____ 172. Tolerates or endorses drinking in the field office at the end of the day allowing employees to wind down.

_____ 173. Spot-checks performance and accuracy of first line supervisors and reports progress of crew.

_____ 174. Does not take action to correct long breaks taken by crew on a daily basis.

_____ 175. Does not immediately correct the unsafe work habits of an employee which results in a fellow employee getting injured.

_____ 176. Does not check the total hours from an employee's time sheet against the actual hours via the tacograph.

_____ 177. Requires recommendations from line supervisors regarding disciplinary action.

_____ 178. Fails to review previous disciplinary action resulting in an inappropriately lenient action being taken.

_____ 179. Gives incomplete orders and instructions to subordinates and tries to hold them accountable.

_____ 180. Does not check back to ensure promised action (as result of a complaint) was carried out.

_____ 181. Allows unauthorized use of supplies.

_____ 182. Does not check daily weight sheets, resulting in a truck with low weight continuing to go to the transfer station twice per day when once would have been sufficient.

_____ 183. Based on a citizen complaint, corrects behavior of subordinate supervisor.

_____ 184. Follows guidelines/policies of progressive discipline, avoiding potential grievances.

_____ 185. Corrects improper procedures followed by clerk on radio and instructs clerk on how situation should have been handled.

_____ 186. Makes an impromptu visit with a different garbage collection crew every day to inspect and develop rapport with the crew.

_____ 187. Spends several days in the field observing collection crews to ensure that directives are being carried out.

_____ 188. Assumes control from start to finish of snow plowing operation.

_____ 189. Fails to check up to see if an assignment is complete after delegating it to someone.

_____ 190. Does not make adjustments in salt application rates.

_____ 191. Investigates employee complaint and handles on a one-on-one basis.

_____ 192. Ensures that correct information is being given out.

_____ 193. Directs and monitors activity of emergency personnel from other bureaus.

_____ 194. Takes no action when complaint being mishandled.

_____ 195. Favors specific first line supervisors in district.

_____ 196. Checks to determine if all work is completed before operation is ended.

_____ 197. Makes arrangements for injured county worker to receive medical attention and transportation to the hospital.

_____ 198. Utilizes supplemental supervisors appropriately.

_____ 199. Approves a written warning for an employee without first checking the employee's record.

_____ 200. Ends operations before verifying if all work is complete.

_____ 201. Is not consistent or uniform in applying procedures.

_____ 202. Remains knowledgeable of status of equipment and progress of operation during snowplowing.

_____ 203. Follows up and completes investigation based upon a citizen complaint about a supervisor before taking action.

_____ 204. Takes time to meet with an employee who has drinking problems, his family and priest to insist on formal treatment to prevent the employee from being fired.

_____ 205. Does not use supplemental supervisors appropriately. (Gives them larger or more difficult routes than they can handle, lets them sit around, or doesn't use them at all.)

_____ 206. Fails to be compassionate, not allowing subordinate supervisor to take 2 hours off due to his child having been in an accident.

_____ 207. Shows empathy to terminally ill employee by disregarding disciplinary action ordered by another district manager.

_____ 208. Utilizes the Employee Assistance Program to save an employee's life.

_____ 209. Shows empathy for employee in a predicament, going to help him when he had car trouble on a freeway on his way to work.

_____ 210. Allows the mid-morning rest period and/or paid lunch hour to exceed what is specified in the union contract.

TRAINING

_____ 211. Fails to train first line supervisors resulting in wasting time answering questions.

_____ 212. Leaves policy training to subordinates' peers.

_____ 213. Knows why training is necessary.

_____ 214. Ensures that subordinates can anticipate problems.

_____ 215. Shares training opportunities with employees.

_____ 216. Does not keep subordinates informed of training opportunities.

_____ 217. Encourages supplemental training for subordinates.

_____ 218. Personally trains subordinates in Bureau policies and procedures.

_____ 219. Trains subordinate supervisor to handle manpower, equipment, resolve problems and citizen complaints.

_____ 220. Ensures that transfer subordinates from the Bureau know its policies and procedures.

_____ 221. Does not show subordinates how to make adjustments in the salt application rates.

_____ 222. Delegates nuts-and-bolts training to subordinate supervisors.

_____ 223. Talks to new supervisor about how to deal with employees, providing the supervisor with a better idea of how to get results without ordering subordinates around.

_____ 224. Explains to new supervisor how to handle a water main break problem.

_____ 225. Trains subordinates on how to make adjustments in the salt application rates.

_____ 226. Subordinates transferred lack training in Bureau policies and procedures.

_____ 227. Ensures that subordinates are familiar with list of streets prone to develop drifting problems.

_____ 228. Lets subordinates learn job on their own.

_____ 229. Trains subordinates to take his/her place if absent, giving authority along with responsibility.

_____ 230. Communicates with subordinates to know their training needs.

_____ 231. Fails to take responsibility for training subordinate supervisor I,

turning job of orientation and training over to a sanitation laborer.

_____ 232. Does not train subordinates to investigate complaints, resulting in crew picking up hazardous material.

_____ 233. Discourages or avoids discussion of supplemental training.

PROFESSIONALISM/DEDICATION

_____ 234. Responds inappropriately when no collection for 3 weeks--response: we'll pick it up next time.

_____ 235. Refuses to take ownership for policy responsibility ("downtown says").

_____ 236. Takes ownership and responsibility for policies.

_____ 237. When garbage not picked up and street under construction, his/her response is "we can't get in".

_____ 238. Minimizes importance of complaints.

_____ 239. Contacts duty area manager when off duty to inform manager of freezing rain that had started to fall near his home.

_____ 240. Does not support a new collection program that is being field tested, therby not providing a true indication of what the program could do.

_____ 241. Has not used any sick leave for 12 straight years which indirectly attributed to other supervisors in that district being reluctant to use sick leave.

_____ 242. Does not comply with several bureasu policies, distorts policies and is unwilling to make efforts involved in policy changes.

_____ 243. Voices disapproval of policies outside of staff meetings.

_____ 244. Does not provide assistance when asked what to do concerning a situation involving sexual harrassment in another bureau.

_____ 245. Supports the Bureau's affirmative action plan, explaining it to subordinates.

_____ 246. Arrives late for a plowing intoxicated. End result is delayed plowing activities.

_____ 247. Is available day and night for call.

_____ 248. Treats complaints as problems to be solved.

_____ 249. Shows approval/support of policies when outside staff meetings.

_____ 250. Unavailable at certain times.

_____ 251. Investigates a citizen complain regarding vehicle used by another bureau and handles in an effective manner.

_____ 252. Handles complaints differently based on where located.

_____ 253. Ensures that catch basins stay open in case of water main breaks.

_____ 254. Volunteers use of special service crew and truck to other districts after rapid drop-off of special service requests.

_____ 255. Is intoxicated during working hours setting a bad example for subordinates and being unable to perform duties.

_____ 256. When brush not picked up, his/her response is "It's too large for us to handle."

SANITATION DISTRICT MANAGER DIMENSIONS

PLANNING AND ORGANIZING SKILLS

The Sanitation District Manager must be able to plan and organize work effectively. This involves the ability to plan ahead in order to attain specific goals or accomplish defined tasks. It involves the ability to set meaningful priorities, to manage time and resources effectively and to meet established deadlines.

ANALYZING AND DECISION-MAKING SKILLS

The Sanitation District Manager often encounters situations requiring immediate analysis and decision-making skills. This involves the ability to identify problems, obtain relevant information and identify possible causes of problems; develop alternative solutions to problems; and determine the possible effects of an action. It also involves the ability to choose from among alternatives based on the facts of the situation and to commit oneself to a course of action.

ORAL COMMUNICATION AND INTERPERSONAL SKILLS

The Sanitation District Manager must be able to speak clearly and communicate ideas and information to others in both one-to-one and group situations such as discussing how a snow emergency will be handled.

In addition, he or she must possess good interpersonal skills and must be persuasive in dealing with people such as irate citizens. This involves the ability to interact effectively with Aldermen, citizens, supervisors, subordinates, peers and staff of other departments. It also involves the ability to see more than one side of an issue, to work cooperatively and to obtain the trust and confidence needed to reach agreement.

SUPERVISORY SKILLS

The Sanitation District Manager must be able to supervise subordinate sanitation supervisors and sanitation laborers. This involves the ability to direct district operations, assign work and delegate work to lower level supervisors, support and motivate staff, check to see if work is done properly and evaluate staff and discipline when necessary.

TRAINING ABILITY

The Sanitation District Manager must have the ability to train or provide training for subordinate supervisors and sanitation laborers. This involves recognizing the need for training, a commitment to providing training opportunities for others and the ability to clearly explain policies and procedures.

PROFESSIONALISM/DEDICATION

The Sanitation District Manager must act as a professional who is dedicated to his/her job. This involves the willingness to accept responsibility, to support department policies, to work beyond one's job description if necessary, to set high goals for one's performance, to strive for accuracy and thoroughness in one's approach to work, to exhibit a positive attitude and to set a good example for others.

WRITTEN COMMUNICATION SKILLS

The ability to effectively and accurately communicate ideas and information to others in writing. Involves clarity and conciseness, use of appropriate vocabulary and grammar, appropriate punctuation and acceptable business style.

APPENDIX N

CITY OF MILWAUKEE

ACHIEVEMENT HISTORY QUESTIONAIRE

FOR

SANITATION DISTRICT MANAGER

GENERAL INSTRUCTIONS

PLEASE NOTE: It is very important to read these instructions very carefully before you begin to complete the questionnaire.

1. *Please answer the sets of questions which appear on the attached pages. Each set of questions relates to an important dimension of the Sanitation District Manager position. The questions ask you to describe what you consider to be your major achievement(s) which would demonstrate that you possess the job-related knowledge, skills or abilities identified. In other words, we are looking for concrete examples of things you have actually done or accomplished pertaining to each dimension rather than a general description of your skills. These achievements may be either specific incidents or examples of sustained high performance over a period of time. Examples of achievements demonstrating skills related to the Planning and Organizing and Analyzing and Decision Making dimensions are attached.*

2. *You should select your strongest achievements regardless of where they were attained. You need not restrict yourself to those attained as a Sanitation Supervisor.*

3. *Please write as neatly as possible in black ink. Attach additional pages as necessary. Your responses will be photocopied for distribution to a panel of raters.*

4. *You are asked to describe one achievement for each of the six dimensions in the questionnaire; however, if you would like, you may describe more than one. If so, attach a separate sheet with the same format.*

PLEASE READ AND SIGN THE FOLLOWING STATEMENT AND RETURN WITH THE QUESTIONNAIRE

I certify that all information provided herein relating to my own achievements and experience is true to the best of my knowledge, and that the information can be verified through persons I have so listed in the questionnaire. I understand that falsification of information may result in disqualification for this examination.

DATE _____

SIGNATURE _____

*This is an example of an achievement demonstrating skills related to the planning
and organizing dimension.*

Example Dimension: Planning and Organizing

*The Sanitation District Manager must be able to plan and organize work effectively.
This involves the ability to plan ahead in order to attain specific goals or
accomplish defined tasks. It involves the ability to set meaningful priorities,
to manage time and resources effectively and to meet established deadlines.*

1. Describe an <u>achievement</u> which would show that you have the ability to plan and
    organize. Tell us <u>what you actually did,</u> and include the objective or the
    problem.

> While working as a Sanitation District Manager, I was responsible
> for weed-cutting in my district. We operated from a card file
> of weed-cutting locations, but the cards were out-of-date with
> some including locations no longer to be cut and the cards didn't
> indicate the most efficient route for the tractor mowers. Also
> I either had to make out a daily list for the mower operator or
> give him the loose cards. This made keeping records for charging
> for weed cutting difficult.
>
> I investigated the locations to see which still needed to be cut
> and developed the best route for the mower operator to take. I
> also set up a weed cutting book which included all information
> about each lot including a diagram and which provided a record
> of when each lot was cut for charging purposes.

2. *What was the outcome or result?*

   Weed-cutting became more efficient because this action:

   1) Removed locations that no longer were to be cut from the list.
   2) Cut down on travel time for mowers by routing locations.
   3) Provided easy reference for any questions regarding the locations, including cutting charges.
   and 4) Cut the time spent by the supervisor making out daily lists.

3. *Was this achievement entirely due to you? Yes __X__ No _____*

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. *a) When did this achievement take place (approximate date)?*
   Summer, 1984

   (b) In what district or for what employer?
   City of Milwaukee, Bureau of Sanitation, Central Area I

5. *Please give the name and title of someone who can verify this information.*

   Joseph Jones
   Central Sanitation Area Manager

EXAMPLE

This is an example of an achievement demonstrating skills related to the analyzing and decision-making dimension.

Example Dimension: Analyzing and Decision-Making

The Sanitation District Manager often encounters situations requiring immediate analysis and decision-making skills. This involves the ability to identify problems, obtain relevant information and identify possible causes of problems; develop alternative solutions to problems; and determine the possible effects of an action. It also involves the ability to choose from among alternatives based on the facts of the situation and to commit oneself to a course of action.

1. _Describe an **achievement** which would show that you have the ability to analyze situations and make good decisions. Tell us **what you actually did**, and include the objective or the problem._

> While working as a Sanitation District Manager for the City of
> Milwaukee, I was responsible for bulky item collection as well
> as regular household garbage collection in my district. During
> a time of heavy demand for bulky item collection, which is
> done by a separate crew, I was able to reduce a growing number
> of requests efficiently by separating brush stops from furniture
> stops and having the regular garbage collection crew also
> collect furniture.
>
> There had been a lot of tree waste to be collected during the
> month of July because of the severe thunderstorm which had swept
> through the Milwaukee area. Bulky collection requests mounted.
> This is also a typically high volume period for household garbage
> collection. However, I thought that the collection crews could
> still afford to help out in this situation and asked my super-
> visors to give two furniture stops to each crew on a daily basis
> until such time as the requests could be handled reasonably well
> by the bulky collection crew. In the meantime, I had the bulky
> truck concentrate on the more time-consuming brush pick ups.
> I was able to reduce the number of bulky collection requests
> in a significantly shorter amount of time than other districts
> and was able to lend a hand to a neighboring district which had
> a greater quantity of requests to handle.

2. *What was the outcome or result?*

I was able to handle the situation without requesting additional resources, which cost money. I also prevented complaints of delayed pickup of the bulky items either by the individual citizen who made the request or the Alderman's office who sometimes forwards complaints of this nature.

3. *Was this achievement entirely due to you? Yes* ___X___ *No* _____

If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

I was solely responsible for analyzing the situation and making the decision. However the lower level supervisors and sanitation laborers carried it out.

4. *a) When did this achievement take place (approximate date)?*

This occurred during July and August, 1984.

(b) In what district or for what employer?

City of Milwaukee, Bureau of Sanitation, South Area 2.

5. *Please give the name and title of someone who can verify this information.*

Joseph Jones
South Sanitation Area Manager

I. Planning and Organizing Skills

The Sanitation District Manager must be able to plan and organize work effectively. This involves the ability to plan ahead in order to attain specific goals or accomplish defined tasks. It involves the ability to set meaningful priorities, to manage time and resouces effectively and to meet established deadlines.

1. *Describe an <u>achievement</u> which would show that you have the ability to plan and organize. Tell us <u>what you actually did,</u> and include the objective or the problem.*

2. *What was the outcome or result?*

3. *Was this achievement entirely due to you? Yes _____ No _____*

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. *a) When did this achievement take place (approximate date)?*

   (b) In what district or for what employer?

5. *Please give the name and title of someone who can verify this information.*

II. *Analyzing and Decision-Making Skills*

The Sanitation District Manager often encounters situations requiring immediate analysis and decision-making skills. This involves the ability to identify problems, obtain relevant information and identify possible causes of problems; develop alternative solutions to problems; and determine the possible effects of an action. It also involves the ability to choose from among alternatives based on the facts of the situation and to commit oneself to a course of action.

1. *Describe an <u>achievement</u> which would show that you have analyzing and decision-making skills. Tell us <u>what you actually did</u>, and include the objective or the problem.*

2. *What was the outcome or result?*

3. *Was this achievement entirely due to you? Yes _____ No _____*

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. *a) When did this achievement take place (approximate date)?*

   (b) In what district or for what employer?

5. *Please give the name and title of someone who can verify this information.*

III. *Oral Communication and Interpersonal Skills*

The Sanitation District Manager must be able to speak clearly and communicate ideas and information to others in both one-to-one and group situations such as discussing how a snow emergency will be handled.

In addition, he or she must possess good interpersonal skills and must be persuasive in dealing with people such as irate citizens. This involves the ability to interact effectively with Aldermen, citizens, supervisors, subordinates, peers and staff of other departments. It also involves the ability to see more than one side of an issue, to work cooperatively and to obtain the trust and confidence needed to reach agreement.

1. *Describe an* <u>*achievement*</u> *which would show that you have oral communication and interpersonal skills. Tell us* <u>*what you actually did*</u>*, and include the objective or the problem.*

2. What was the outcome or result?

3. Was this achievement entirely due to you? Yes _____ No _____

   If no, what is the estimated percentage of this achievement which is due to
   the efforts of other people? _____%

4. a) When did this achievement take place (approximate date)?

   (b) In what district or for what employer?

5. Please give the name and title of someone who can verify this information.

IV. *Supervisory Skills*

The Sanitation District Manager must be able to supervise subordinate sanitation supervisors and sanitation laborers. This involves the ability to direct district operations, assign work and delegate work to lower level supervisors, support and motivate staff, check to see if work is done properly and evaluate staff and discipline when necessary.

1. *Describe an <u>achievement</u> which would show that you have supervisory skills. Tell us <u>what you actually did</u>, and include the objective or the problem.*

2. What was the outcome or result?

3. Was this achievement entirely due to you? Yes _____ No _____

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. a) When did this achievement take place (approximate date)?

   (b) In what district or for what employer?

5. Please give the name and title of someone who can verify this information.

V. Training Ability

The Sanitation District Manager must have the ability to train or provide training for subordinate supervisors and sanitation laborers. This involves recognizing the need for training, a commitment to providing training opportunities for others and the ability to clearly explain policies and procedures.

1. *Describe an <u>achievement</u> which would show that you have the ability to train subordinates. Tell us <u>what you actually did</u>, and include the objective or the problem.*

2. What was the outcome or result?

3. Was this achievement entirely due to you? Yes _____ No _____

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. a) When did this achievement take place (approximate date)?

   (b) In what district or for what employer?

5. Please give the name and title of someone who can verify this information.

VI. *Professionalism/Dedication*

The Sanitation District Manager must act as a professional who is dedicated to his/her job. This involves the willingness to accept responsibility, to support department policies, to work beyond one's job description if necessary, to set high goals for one's performance, to strive for accuracy and thoroughness in one's approach to the work, to exhibit a positive attitude and to set a good example for others.

1. *Describe an <u>achievement</u> which would show that you act as a professional who is dedicated to his/her job. Tell us <u>what you actually did,</u> and include the objective or the problem.*

2. *What was the outcome or result?*

3. *Was this achievement entirely due to you? Yes _____ No _____*

   If no, what is the estimated percentage of this achievement which is due to the efforts of other people? _____%

4. *a) When did this achievement take place (approximate date)?*

   (b) In what district or for what employer?

5. *Please give the name and title of someone who can verify this information.*

APPENDIX O

RATER EVALUATIONS

We appreciate your serving as a rater for the Sanitation District Manager examination. Your answers to the following questions will be used to improve our selection process.

1. *Do you consider it fair to use accomplishment ratings to assess candidates on job dimensions?*

      A. Very fair
      B. Fair
      C. Somewhat fair
      D. Unfair
      E. Very unfair

2. *Do you consider it fair to use accomplishment ratings to determine which applicants will participate in an oral examination?*

      A. Very fair
      B. Fair
      C. Somewhat fair
      D. Unfair
      E. Very unfair

3. *Do you consider it fair to use accomplishment ratings as a weighted part of the exam process?*

      A. Very fair
      B. Fair
      C. Somewhat fair
      D. Unfair
      E. Very unfair

4. *Do you consider it fair to use accomplishment ratings to determine final rankings in the examination process?*

      A. Very fair
      B. Fair
      C. Somewhat fair
      D. Unfair
      E. Very unfair

5. *How effective do you consider accomplishment ratings in assessing candidates on the job dimensions?*

      A. Very effective
      B. Effective
      C. Somewhat effective
      D. Ineffective
      E. Very ineffective

6. *How effective do you consider accomplishment ratings in determining the best qualified applicants to be called for an oral examination?*

      A. Very effective
      B. Effective
      C. Somewhat effective
      D. Ineffective
      E. Very ineffective

7. *How effective do you consider accomplishment ratings as a weighted part of the exam process?*

      A. Very effective
      B. Effective
      C. Somewhat effective
      D. Ineffective
      E. Very ineffective

8. *How effective do you consider accomplishment ratings in determining the best qualified applicants for a job?*

      A. *Very effective*
      B. *Effective*
      C. *Somewhat effective*
      D. *Ineffective*
      E. *Very ineffective*

9. *How difficult did you consider the accomplishments to rate?*

      A. *Very easy*
      B. *Easy*
      C. *Somewhat difficult*
      D. *Difficult*
      E. *Very difficult*

10. *How reasonable do you consider the time you spent rating accomplishments for a position of this level and type?*

      A. *Very reasonable*
      B. *Reasonable*
      C. *Somewhat reasonable*
      D. *Unreasonable*
      E. *Very unreasonable*

RATER EVALUATIONS

We appreciate your serving on the oral examination panel for the position of Sanitation District Manager. As you recall, the examination involved rating candidates on individual dimensions such as Planning & Organizing Skills, Analytical and Decision-Making Skills, Supervisory Skills, etc. Your answers to the following questions will be used to improve our selection process.

1. Do you consider it fair to use oral examination ratings to assess candidates on job dimensions?

       A. Very fair
       B. Fair
       C. Somewhat fair
       D. Unfair
       E. Very unfair

2. Do you consider it fair to use oral dimension ratings as a weighted part of the exam process?

       A. Very fair
       B. Fair
       C. Somewhat fair
       D. Unfair
       E. Very unfair

3. Do you consider it fair to use oral dimension ratings to determine final rankings in the examination process?

       A. Very fair
       B. Fair
       C. Somewhat fair
       D. Unfair
       E. Very unfair

4. How effective do you consider oral examination ratings in assessing candidates on the job dimensions?

       A. Very effective
       B. Effective
       C. Somewhat effective
       D. Ineffective
       E. Very ineffective

5. How effective do you consider oral dimension ratings as a weighted part of the exam process?

       A. Very effective
       B. Effective
       C. Somewhat effective
       D. Ineffective
       E. Very ineffective

6. How effective do you consider oral dimension ratings in determining the best qualified applicants for a job?

       A. Very effective
       B. Effective
       C. Somewhat effective
       D. Ineffective
       E. Very ineffective

7. How difficult did you consider the oral dimensions to rate?

      A. Very easy
      B. Easy
      C. Somewhat difficult
      D. Difficult
      E. Very difficult

8. How reasonable do you consider the time you spent rating oral dimensions for a position of this level and type?

      A. Very reasonable
      B. Reasonable
      C. Somewhat reasonable
      D. Unreasonable
      E. Very unreasonable

9. Please add any comments in the space below.

Significance Tests for Differences Between Oral and
Behavioral Consistency Reliability Coefficients for
Dimension and Overall Reliabilities for the Sanitation
District Manager Study

	r_1	r_2	r_{12}	W	t
D1	.88	.80	.58	1.67	1.10
D2	.89	.61	.29	3.55	2.44*
D3	.87	.73	.62	2.08	1.65
D4	.76	.68	.74	1.33	.74
D5	.80	.54	.31	2.30	1.56
D6	.84	.88	.62	.75	-.64
O	.90	.86	.67	1.40	.79

Note. D1 = Planning and Organizing; D2 = Analyzing and
Decision-making; D3 = Oral Communication/Interpersonal
Skill; D4 = Supervision; D5 = Training; D6 = Profession-
alism/Dedication; O = Overall Score; column definitions are
as follows:

r_1 = oral reliability based on average of four raters,

r_2 = behavioral consistency reliability,

r_{12} = correlation between oral and behavioral
consistency ratings,

W = $(1-r_2)/(1-r_1)$,

t_{N-2} = $(W-1)(N-2)^{1/2} / [4W(1-r_{12}^2)]^{1/2}$, and

N = sample size.

* $p < .05$, df = 12, two-tailed.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, C. W. (1960). The relation between speaking times and decision in the employment interview. Journal of Applied Psychology, 44, 267-268.

Ash, R. A. (1984, May). The activity/achievement indicator: A possible alternative to the behavioral consistency method of training and experience evaluation. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Seattle, WA.

Ash, R. A. (1983, May). A comparative study of the behavioral consistency and wholistic judgment methods of job applicant training and work experience evaluation. Paper presented at the meeting of the International Personnel Management Association Assessment Council, Washington, DC.

Ash, R. A., & Levine, E. L. (1982, August). Job applicant training and work experience evaluation: An empirical investigation. Paper presented at the meeting of the American Psychological Association, Washington, DC.

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. Personnel Psychology, 35, 281-322.

Barbee, J. R., & Keil, E. C. (1973). Experimental techniques of job interview training for the disadvantaged: Videotape feedback, behavior modification and microcounseling. Journal of Applied Psychology, 58, 209-213.

Baron, R. A. (1983). "Sweet smell of success"? The impact of pleasant artificial scents on evaluations of job applicants. Journal of Applied Psychology, 68, 709-713.

Baskett, G. D. (1973). Interview decisions as determined by competency and attitude similarity. Journal of Applied Psychology, 57, 343-345.

Blakeney, R. N., & Mac Naughton, J. F. (1971). Effects of temporal placement of unfavorable information on decision making during the selection interview. *Journal of Applied Psychology*, *55*, 138-142.

Bolster, B. I., & Springbett, B. M. (1961). The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology*, *45*, 97-103.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cann, A., Siegfried, W. D., & Pearce, L. (1981). Forced attention to specific applicant qualifications: Impact on physical attractiveness and sex of applicant biases. *Personnel Psychology*, *34*, 65-75.

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.

Carlson, R. E. (1967). Selection interview decisions: The relative influence of appearance and factual written information on an interviewer's final rating. *Journal of Applied Psychology*, *51*, 461-468.

Carlson, R. E. (1970). Effect of applicant sample on ratings of valid information in an employment setting. *Journal of Applied Psychology*, *54*, 217-222.

Carlson, R. E. (1971). Effect of interview information in altering valid impressions. *Journal of Applied Psychology*, *55*, 66-72.

Carlson, R. E., Thayer, P. W., Mayfield, E. C., & Peterson, D. A. (1971, April). Improvements in the selection interview. *Personnel Journal*, 268-317.

Cash, T. F., Gillen, B., & Burns, D. S. (1977). Sexism and "Beautyism" in personnel consultant decision making. *Journal of Applied Psychology*, *62*, 301-310.

Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. *Journal of Applied Psychology*, *60*, 566-572.

Davey, B. (1984, May). *Are All Oral Panels Created Equal? A Study of Differential Validity Across Oral Panels*. Paper presented to the International Personnel Management Association Assessment Council, Seattle, WA.

Dipboye, R. L., Fontenelle, G. A., & Garner, K. (1984). Effects of previewing the application on interview process and outcomes. Journal of Applied Psychology, 69, 118-128.

Dipboye, R. L., Fromkin, H. L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 60, 39-43.

Dunnette, M. D. (1962). Personnel management. Annual Review of Psychology, 13, 285-314.

Ebel, R. L. (1951). Estimation of the reliability of ratings. Psychometrika, 16, 407-424.

Farr, J. L. (1973). Response requirements and primacy-recency effects in a simulated selection interview. Journal of Applied Psychology, 57, 228-232.

Farr, J. L., & York, C. M. (1975). Amount of information and primacy-recency effects in recruitment decisions. Personnel Psychology, 28, 233-238.

Fay, C. H., & Latham, G. P. (1982). Effects of training and rating scales on rating errors. Personnel Psychology, 35, 105-116.

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. Psychometrika, 45, 99-105.

Ferris, G. R., & Gilmore, D. C. (1977). Effects of mode of information presentation, sex of applicant, and sex of interviewer on simulated interview decisions. Psychological Reports, 40, 566.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Forsythe, S., Drake, M. F., & Cox, C. F. (1985). Influence of applicant's dress on interviewer's selection decisions. Journal of Applied Psychology, 70, 374-378.

Frank, L. L., & Hackman, J. R. (1975). Effects of interviewer-interviewee similarity on interviewer objectivity in college admissions interviews. Journal of Applied Psychology, 60, 356-360.

Hakel, M. D. (1971). Similarity of post-interview trait rating intercorrelations as a contributor to interrater agreement in a structured employment interview. Journal of Applied Psychology, 55, 443-448.

Hakel, M. D., Dobmeyer, T. W., & Dunnette, M. D. (1970). Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. Journal of Applied Psychology, 54, 65-71.

Hakel, M. D., Ohnesorge, J. P., & Dunnette, M. D. (1970). Interviewer evaluations of job applicants' resumes as a function of the qualifications of the immediately preceding applicants: An examination of contrast effects. Journal of Applied Psychology, 54, 27-30.

Heilman, M. E. (1980). The impact of situational factors on personnel decisions concerning women: Varying the sex composition of the applicant pool. Organizational Behavior and Human Performance, 26, 386-395.

Heilman, M. E., & Saruwatari, L. R. (1979). When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and nonmanagerial jobs. Organizational Behavior and Human Performance, 23, 360-372.

Heneman, H. G. III, Schwab, D. P., Huett, D. L., & Ford, J. L. (1975). Interviewer validity as a function of interview structure, biographical data, and interviewee order. Journal of Applied Psychology, 60, 748-753.

Hollandsworth, J. G., Jr., Dressel, M. E., & Stevens, J. (1977). Use of behavioral versus traditional procedures for increasing job interview skills. Journal of Applied Psychology, 24, 502-510.

Hollandsworth, J. G., Jr., Kazelskis, R., Stevens, J., & Dressel, M. E. (1979). Relative contributions of verbal, articulative and nonverbal communication to employment decisions in the job interview setting. Personnel Psychology, 32, 359-367.

Hough, L. M. (1984). Development and evaluation of the "Accomplishment Record" method of selecting and promoting professionals. Journal of Applied Psychology, 69, 135-146.

Imada, A. S., & Hakel, M. D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. Journal of Applied Psychology, 62, 295-300.

Ivancevich, J. M., & Smith, S. V. (1981). Goal setting interview skills training: Simulated and on-the-job analyses Journal of Applied Psychology, 66, 697-705.

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

James, S. P., Campbell, I. M., & Lovegrove, S. A. (1984). Personality differentiation in a police-selection interview. Journal of Applied Psychology, 69, 129-134.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. Journal of Applied Psychology, 67, 577-580.

Johnson, J. C., Guffey, W. L., & Perry, R. A. (1980, July). When is a T and E rating valid? Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Boston, Ma.

Keenan, A. (1978). The selection interview: Candidates' reactions and interviewers' judgements. British Journal of Social and Clinical Psychology, 17, 201-209.

King, M. R., & Manaster, G. J. (1977). Body image, self-esteem, expectations, self-assessments and actual success in a simulated job interview. Journal of Applied Psychology, 62, 589-594.

Kopelman, M. D. (1975). The contrast effect in the selection interview. British Journal of Educational Psychology, 45, 333-336.

Landy, F. J., & Bates, F. (1973). Another look at contrast effects in the employment interview. Journal of Applied Psychology, 58, 141-144.

Langdale, J. A., & Weitz, J. (1973). Estimating the influence of job information on interviewer agreement. Journal of Applied Psychology, 57, 23-27.

Latham, G. P., Fay, C. H., & Saari, L. M. (1979). The development of behavioral observation scales for appraising the performance of foremen. Personnel Psychology, 32, 299-311.

Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. Journal of Applied Psychology, 69, 569-573.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. Journal of Applied Psychology, 65, 422-427.

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.

Leonard, R. L., Jr. (1974). Relevance and reliability in the interview. Psychological Reports, 34, 1331-1334.

London, M., & Hakel, M. D. (1974). Effects of applicant stereotypes, order and information on interview impressions. Journal of Applied Psychology, 59, 157-162.

Maas, J. B. (1965). Patterned scaled expectation interview: Reliability studies on a new technique. Journal of Applied Psychology, 49, 431-433.

Mayfield, E. C. (1964). The selection interview--A re-evaluation of published research. Personnel Psychology, 17, 239-260.

Mayfield, E. C., Brown, S. H., & Hamstra, B. W. (1980). Selection interviewing in the life insurance industry: An update of research and practice. Personnel Psychology, 33, 725-739.

McDonald, T., & Hakel, M. D. (1985). Effects of applicant race, sex, suitability, and answers on interviewer's questioning strategy and ratings. Personnel Psychology, 38, 321-334.

McIntyre, S., Moberg, D. J., & Posner, B. Z. (1980). Preferential treatment in preselection decisions according to sex and race. Academy of Management Journal, 23, 738-749.

Moore, L. F., & Lee, A. J. (1974). Comparability of interviewer, group and individual interview ratings. Journal of Applied Psychology, 59, 163-169.

Mullins, T. W. (1982). Interviewer decisions as a function of applicant race, applicant quality and interviewer prejudice. Personnel Psychology, 35, 163-174.

Okanes, M. M., & Tschirgi, H. (1978). Impact of the face-to-face interview on prior judgments of a candidate. Perceptual and Motor Skills, 46, 322.

Osburn, H. G., Timmreck, C., & Bigby, D. (1981). Effect of dimensional relevance on accuracy of simulated hiring decisions by employment interviewers. Journal of Applied Psychology, 66, 159-165.

Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening applicants. Personnel Psychology, 37, 507-514.

Parsons, C. K., & Liden, R. C. (1984). Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. Journal of Applied Psychology, 69, 557-568.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.

Rand, T. M., & Wexley, K. N. (1975). Demonstration of the effect, "Similar to Me," in simulated employment interviews. Psychological Reports, 36, 535-544.

Rasmussen, K. G., Jr. (1984). Nonverbal behavior, verbal behavior, resume credentials, and selection interview outcomes. Journal of Applied Psychology, 69, 551-556.

Reynolds, A. H. (1979). The reliability of a scored oral interview for police officers. Public Personnel Management, Sep-Oct, 324-328.

Ricchiute, D. N. (1985). Presentation mode, task importance, and cue order in experimental research on expert judges. Journal of Applied Psychology, 70, 367-373.

Rosen, B., & Jerdee, T. H. (1976). The influence of age stereotypes on managerial decisions. Journal of Applied Psychology, 61, 428-432.

Rosen, B., & Mericle, M. F. (1979). Influence of strong versus weak fair employment policies and applicant's sex on selection decisions and salary recommendations in a management simulation. Journal of Applied Psychology, 64, 435-439.

Rothstein, M., & Jackson, D. N. (1980). Decision making in the employment interview: An experimental approach. Journal of Applied Psychology, 65, 271-283.

Rowe, P. M. (1963). Individual defferences in selection decisions. Journal of Applied Psychology, 47, 304-307.

Rozelle, R. M., & Baxter, J. C. (1981). Influence of role pressures on the perceiver: Judgments of videotaped interviews varying judge accountability and responsibility. Journal of Applied Psychology, 66, 437-441.

Sackett, P. R. (1982). The interviewer as hypothesis tester: the effects of impressions of an applicant on interviewer questioning strategy. Personnel Psychology, 35, 789-804.

Schmidt, F. L., Caplan, J. R., Bemis, S. E., Decuir, D., Dunn, L., & Antone, L. (1979). The behavioral consistency method of unassembled examining. (TM-79-21) Washington, DC: Office of Personnel Management. (NTIS No. PB80-139942).

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 29, 79-101.

Schwab, D. P., & Heneman, H. G. III. (1969). Relationship between interview structure and interinterviewer reliability in an employment situation. Journal of Applied Psychology, 53, 214-217.

Schwab, D. P., Heneman, H. G. III, & DeCotiis, T.A. (1975). Behaviorally anchored rating scales: A review of the literature. Personnel Psychology, 28, 549-562.

Simas, K., & McCarrey, M. (1979). Impact of recruiter authoritarianism and applicant sex on evaluation and selection decisions in a recruitment interview analogue study. Journal of Applied Psychology, 64, 483-491.

Smith, P., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

State of Alabama Personnel Office. (1984). Summary of T and E Examinations Data Reported by State and Municipal Jurisdictions. Montgomery, AL.

Sterrett, J. (1978). The job interview: Body language and perceptions of potential effectiveness. Journal of Applied Psychology, 63, 388-390.

Tengler, C. D., & Jablin, F. M. (1983). Effects of question type, orientation, and sequencing in the employment screening interview. Communication Monographs, 50, 245-263.

Tubiana, J. H., & Ben-Shakhar, G. (1982). An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. Personnel Psychology, 35, 349-357.

Tucker, D. H., & Rowe, P. M. (1977). Consulting the application form prior to the interview: An essential step in the selection process. Journal of Applied Psychology, 62, 283-287.

Tucker, D. H., & Rowe, P. M. (1979). Relationship between expectancy, causal attributions and final hiring decisions in the employment interview. Journal of Applied Psychology, 64, 27-34.

Tullar, L., Mullins, T., & Caldwell, S. (1979). Effects of interview length and applicant quality on interview decision time. Journal of Applied Psychology, 64, 669-674.

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin, 63, 100-116.

Valenzi, E., & Andrews, I. R. (1973). Individual differences in the decision process of employment interviewers. Journal of Applied Psychology, 58, 49-53.

Vance, R. J., Kuhnert, K. W., & Farr, J. L. (1978). Interview judgments: Using external criteria to compare behavioral and graphic scale ratings. Organizational Behavior and Human Performance, 22, 279-294.

Wagner, R. (1949). The employment interview: A critical summary. Personnel Psychology, 2, 17-46.

Washburn, P. V., & Hakel, M. D. (1973). Visual cues and verbal content as influences on impressions formed after simulated employment interviews. Journal of Applied Psychology, 58, 137-140.

Wexley, K., Sanders, R., & Yukl, G. (1973). Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 57, 233-236.

Wexley, K., Yukl, G., Kovacs, S., & Sanders, R. (1972). Importance of contrast effects in employment interviews. Journal of Applied Psychology, 56, 45-48.

Wiener, Y., & Schneiderman, M. L. (1974). Use of job information as a criterion in employment decisions of interviewers. Journal of Applied Psychology, 59, 699-704.

Wright, O. R. Jr. (1969). Summary of research on the selection interview since 1964. <u>Personnel Psychology</u>, <u>22</u>, 391-413.