The assessment center process selection of non-managerial talent in the public sector

Beth Perrine

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd-project

Part of the Public Administration Commons

Recommended Citation
https://scholarworks.lib.csusb.edu/etd-project/197

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
THE ASSESSMENT CENTER PROCESS:
SELECTION OF NON-MANAGERIAL TALENT IN THE PUBLIC SECTOR

A Thesis
Presented to the
Faculty of
California State College
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology

by
Beth Perrine
June 1980
THE ASSESSMENT CENTER PROCESS:
SELECTION OF NON-MANAGERIAL TALENT IN THE PUBLIC SECTOR

A Thesis
Presented to the
Faculty of
California State
College, San Bernardino

by
Beth Perrine
June 1980

Approved by:

Chair

Date
June 9, 1980
ABSTRACT

The effectiveness of the assessment center process in selecting candidates who have potential ability for non-managerial positions was evaluated in this study. The research investigated a somewhat unique program, in that assessment center techniques had been applied to the selection of Upward Mobility candidates in a Federally mandated program. In Analysis 1, the relationship between educational levels and assessment ratings of 137 subjects was found to be negligible. Since the assessment center's objective is to assess potential, rather than education or experience, this was a positive finding. The second analysis examined the retest reliability of the assessment center at a one-year interval. The reliability coefficients, based on a sample of 24 subjects, were all nonsignificant, but a clear interpretation of these results was impossible due to intervening factors. A repeated measures analysis of variance on the retest data found significant main effects for the skill areas and an interaction of measurement periods and skill areas. The direction of the effect for two variables is consistent with the developmental emphasis of the program. A third analysis, comparing 29 Upward Mobility "graduates" to a matched group of 29 subjects who were hired through
normal merit procedures, found no difference in job performance as evaluated by the employees' supervisors. Evidence for the validity of the Upward Mobility program's selection procedure can be inferred from the performance ratings; after training, Upward Mobility graduates are able to perform as competently as employees selected on the basis of prior experience and education.
# TABLE OF CONTENTS

LIST OF TABLES ........................................... vi

ACKNOWLEDGMENTS ........................................... vii

Chapter

I. INTRODUCTION ........................................... 1

- Assessment Center History and Research ............. 1
- Recent Developments .................................. 3
- The Upward Mobility Assessment Program .......... 7
- The Scope of the Present Study ....................... 10

II. ANALYSIS 1 ............................................ 13

- Method .................................................. 13
- Subjects ................................................. 13
- Procedure .............................................. 13
- Results and Discussion ................................ 15

III. ANALYSIS 2 ............................................ 21

- Method .................................................. 21
- Subjects ................................................. 21
- Procedure .............................................. 21
- Results and Discussion ................................ 22

IV. ANALYSIS 3 ............................................ 27

- Method .................................................. 27
- Subjects ................................................. 27
- Performance Evaluation Form ......................... 27
- Procedure .............................................. 27
- Results and Discussion ................................ 29

V. DISCUSSION ............................................. 30

REFERENCES .............................................. 34
LIST OF TABLES

1. Means and Intercorrelations Among Skill Areas and Years of Education 16
2. Educational Levels of 1979-1980 Assessment Center Participants Who were Selected for Jobs 19
3. Correlations Between 1978 Assessment Scores and 1979 Assessment Scores 23
5. Supervisory Evaluation Form 28
ACKNOWLEDGMENTS

I would like to acknowledge my advisors, Drs. Warren, Cramer and Cowan, for their commitment, interest and assistance as this thesis evolved. Special thanks to Ron Covert, who introduced me to the computer and stayed as a willing translator as long and as often as needed. My sincerest appreciation goes to Clara Erickson for her thoughtful and timely presentations of opportunities, resources and encouragement. I am indebted to the Personnel Department at the Naval Weapons Center, China Lake and the Psychology Department at California State College, San Bernardino for their generous cooperation and support.
CHAPTER I

INTRODUCTION

Assessment Center History and Research

The assessment center process was initially developed by German military psychologists in the 1930s. During World War II, the idea was utilized by Britain's War Officer Selection Boards and America's Office of Strategic Services (OSS). The OSS used the procedure to select intelligence agents based on their abilities to apply a variety of skills in simulated situational exercises. The trials and errors of the OSS effort are detailed in their report, Assessment of Men (1948). Although the OSS staff was unable to prove that the OSS assessments produced effects which more than balanced the expenditure of time and money, the results and recommendations provided a model for industry to use in the selection of management personnel.

The assessment center process as a selection procedure for managerial talent was first implemented by American Telephone and Telegraph (AT&T) in the mid-1950s. A longitudinal research project, the AT&T Management Progress Study (Bray & Grant, 1966), became the model and rationale for assessment center validity. The information from the assessment process was not given to the company or used to
influence the careers of the 422 men who were assessed. Thus, the assessment results did not contaminate subsequent criteria. The relationship between assessment center predictions and management level achieved was quite impressive: \( r = .44 \) for college men and .71 for noncollege men (Bray & Grant, 1966).

Since the 1950s, the use of assessment techniques has been applied to a wide variety of positions. The common element in the use of assessment centers is that the applicant has not had the opportunity to display his or her potential to perform the duties of the position; the assessment center techniques are thus uniquely suited for providing data for the selection process when potential is being assessed. The basic components of an assessment center are simulations of actual work-related situations, multiple exercises, and multiple observations.

A number of studies have supported the belief that assessment centers are highly effective in identifying potential. In a review of the research literature, Byham (1970) found 22 studies indicating that assessment centers are more effective than other approaches such as rating personnel jackets and interviewing candidates, and he found no study indicating that assessment centers were less effective. Correlations in these studies between center prediction and achievement criteria of advancement, salary
level and performance ratings range as high as .64. The predictive accuracy of the assessment ratings has been well established; Cohen, Moses and Byham (1974) also found that 18 research studies in a variety of organizational settings consistently showed assessment ratings related to several performance effectiveness measures.

The question of interrater reliability in the assessment process also arises because multiple observers (assessors) rate the candidates and pool their judgments. Again, accumulated evidence from past research is conclusive in showing that the assessment process is not limited by low reliability. Typical interrater reliabilities for assessment dimensions range from $r = .68$ to $r = .99$ (Huck, 1973; Greenwood & McNamara, 1967). Assessment center reliability is enhanced by the intensive training of assessors and by the standardization incorporated into the assessment procedures.

One of the few studies which deals with the consistency of participants' performance over time was conducted by Michigan Bell (Moses, 1973). Correlations between two assessment centers ranged from .68 to .77, with a one-month interval between assessments.

**Recent Developments**

Two recent developments in the brief history of assessment centers are an increasing use of the process in the public sector and the expansion of the process to include
selection to non-managerial positions. Complex governmental regulations and legislative requirements have made the assessment center very attractive as a selection method compared to traditional methods. The expanding use of the process in the public sector, however, brings a concurrent concern: few studies have been conducted on the effectiveness of public service assessment centers. Ninety-five percent of the validation studies have been conducted in the private sector and not under selection conditions of a civil service system (Ross, 1979). Because the assessment center concept is well-grounded in research, it has been assumed that any assessment center will be successful. However, in her evaluation of assessment center, Howard (1974) stressed the point that each situation requires a somewhat different approach and its own evaluation to be considered valid. Additionally, little research has been done on assessment center effectiveness in the identification of non-managerial potential.

In her review, Ross (1979) suggested that many public sector assessment centers are marginal in terms of meeting the criteria for a reliable and valid center. The commonly used guidelines to define an assessment center are listed in the "Standards and Ethical Considerations for Assessment Center Operations" developed by the Third International Congress on the Assessment Center Method (Moses et al., 1975).
The seven minimal professional requirements of an assessment center are as follows:

1. Multiple assessment techniques must be used. At least one of these techniques must be a simulation.

   A simulation is an exercise or technique designed to elicit behaviors related to dimensions of performance on the job by requiring the participant to respond behaviorally to situational stimuli. The stimuli present in a simulation parallel or resemble stimuli in the work situation. Examples of simulations include group exercises, in-basket exercises and fact finding exercises.

2. Multiple assessors must be used. These assessors must receive training prior to participating in a center.

3. Judgments resulting in an outcome (i.e., recommendation for promotion, specific training or development) must be based on pooling information from assessors and techniques.

4. An overall evaluation of behavior must be made by the assessors at a separate time from observation of behavior.

5. Simulation exercises are used. These exercises are developed to tap a variety of pre-determined behaviors and have been pretested prior to use to insure that the techniques provide reliable, objective and relevant behavioral information for the organization in question.

6. The dimensions, attributes, characteristics or qualities evaluated by the assessment center are determined by an analysis of relevant job behaviors.
7. The techniques used in the assessment center are designed to provide information which is used in evaluating the dimensions, attributes or qualities previously determined (Moses et al., 1975).

Specific areas of concern delineated by Ross (1979) included the overuse of assessment dimensions and exercises, a lack of assessor training, and a need for job analyses to be situation specific. Ross challenged public sector organizations to evaluate critically the conduct of their assessment centers or risk the loss of an effective selection device.

Alexander, Buck and McCarthy (1975) addressed both the issues of non-managerial selection and public sector utilization of an assessment center by developing an assessment center process to select Upward Mobility candidates for the Federal Aviation Administration (FAA). The goal of formal Upward Mobility Programs is to provide advancement opportunities to employees based on their potential to perform the duties of the target position rather than on their past work experience and background. Upward Mobility selection is complicated by three factors: (a) the candidates generally have no prior experience with the positions for which they are being selected, (b) applicants are applying for several unrelated target positions, and (c) the large numbers of applicants cannot be equitably screened with traditional procedures. In their comparison of supervisors' ratings and assessment scores, Alexander et al. found that
the assessment process may be more fair and accurate than normal merit procedures. Their results indicated that the assessment center process identified a different group of people than supervisory ratings would have; the data also indicated that supervisory ranking does not provide as much discrimination among candidates. Of the 111 people evaluated, only 19% of those ranked highest by a supervisory rating method were also ranked highest by the assessment center method; no more than 40% of those selected by one method would have been selected by the other. The correlation between the two methods was only .23, whereas internal correlations between assessment center exercises were much higher. For example, the correlation between a patterned interview ranking and the ranking of an analytical exercise was .65. The data suggested that supervisory ratings are not necessarily an accurate basis for selection because the supervisors lack training in objective observation and standardized rating. Alexander et al. concluded that the assessment center method is particularly effective in the selection of Upward Mobility candidates.

The Upward Mobility Assessment Program

The success reported in the Alexander study prompted the Naval Weapons Center (NAVWPNCEN) to adapt the FAA assessment center model to the NAVWPNCEN Upward Mobility Program. Since the first NAVWPNCEN Assessment Center in 1975, about 500 Upward Mobility applicants have been
assessed. Currently, 30 selected candidates have completed their two-year training programs and are established in target positions. An additional 14 selectees are in training programs, while 12 have dropped out of training due to promotions, resignations, and reassignments. NAVWPNCCEN reserves 20 billets to be used only for Upward Mobility, and the Upward Mobility training plans designed for each selected candidate can require up to two years of training before the target position is achieved. Thus, the Upward Mobility register is infrequently used outside of the 20 allocated billets, even though it is available as an alternative staffing option.

The two-day assessment center program at NAVWPNCCEN consists of ten hours of assessment exercises and feedback and six hours of career development and goal setting activities. The assessment exercises used are (a) a one-on-one interview based on the Background Questionnaire, (b) a six-participant group discussion, and (c) an analysis exercise and one-on-one interview based on the analysis exercise. These exercises are tailored to provide opportunity for the candidates to demonstrate their potential in the skill areas previously identified as critical to the target positions covered by the register. The skill areas were identified through a task-analysis of technical/electronic jobs and a review of commonly used dimensions in similar contexts for the nontechnical job clusters.
The participants are also given training in how to give and receive feedback, and receive a 45 minute one-on-one feedback interview concerning their performance in the assessment exercises. In addition, the career development portion of the program gives participants feedback on their occupational interests, interpersonal style and values. The program concludes with career goal setting based on all the feedback received during the two-day program.

The assessment is performed by about 20 persons in the organization who have supervisory or administrative experience. These assessors receive 16 hours of training in the assessment exercises as well as interviewing, observation and feedback skills.

The participants' scores for each of nine skill areas are derived from the combined judgments of three assessors. A tenth skill area, Dependability, is rated by the participants' supervisors. This program does not utilize an overall assessment rating for each participant; a participant's final score in each skill area is maintained in the register for selection by potential employers. Scoring on the following skill area dimensions ranges from 0 to 4, with 0 being "no ability" and 4 being "superior ability".

Oral Communications Skills
Ability to Express Oneself in Writing
Ability to Interpret and Apply Instructions and Procedures
Initiative
Analytical Ability
Ability to Work Well with Others
Flexibility/Adaptability
Ability to Work Independently
Interest in Electronics
Dependability (rated by supervisors, not assessors)

The Scope of the Present Study

Given the dearth of empirical data on public sector assessment centers and their increasing use in prediction of non-managerial potential, the present study is designed to evaluate aspects of reliability and validity of the NAVWPNCEN Upward Mobility Assessment/Development Program (UMA/D). A critical problem in this type of evaluation research is to obtain a reliable estimate of performance effectiveness. Performance effectiveness measures such as salary growth and promotion rate are confounded when used as criteria, since the assessment ratings are used as a basis for initial promotion. In the present instance, it was not feasible to wait several years while developing a long-term predictive validation strategy, such as the AT&T longitudinal research project; a more immediate evaluation of the program's effectiveness was required. The major questions addressed by this study are: (a) Is there a relationship between educational level and assessment center scores, and which variable, skill area or education level, contributes
the most to the selection procedure? (b) Are assessment ratings relatively stable in a test-retest situation? (c) How comparable is the assessment center selection to normal merit procedures at full performance levels?

(a) It was expected that a positive relationship exists between educational levels and the ratings of several skill areas; highly significant positive correlations would suggest that assessment ratings are not independent of educational levels. If the variance in a given skill area can be well explained by years of completed education, then the question arises: Is education rather than skill potential actually being measured in the assessment process? For example, a rating on Oral Communications may be accurately predicted by years of school completed; the utility of assessing this skill area would then be questionable. A related question addressed in this study was: What percentage of subjects selected by the assessment process would have been selected by education alone?

(b) Because several changes were made in the 1979 assessment exercises, 1978 participants were provided the opportunity of reapplying to the 1979 assessment center. Thus, a group of 24 participants were retested; the group's two sets of ratings were analyzed to provide information on retest stability, developmental progress, and effects of the program changes. It was expected that the 1979 group mean would be equal to or greater than the 1978 group mean.
(c) The relationship between selection decisions based on the assessment ratings and subsequent job performance was addressed inferentially through the comparison of Upward Mobility "graduates" matched to a comparison group who achieved equivalent positions through normal merit procedures. The comparison criterion was a supervisory performance evaluation form designed for this study. The null hypothesis was tested with the expectation that the two groups would not be significantly different. Evidence for the predictive validity of the assessment process can be inferred if there is no difference between the two groups' performance ratings.
CHAPTER II

ANALYSIS 1

Method

Subjects. The subjects for the analysis of assessment scores were 137 self-selected Upward Mobility Assessment/Development Program applicants of the August 1979 and the January 1980 NAVWPNCEN assessment centers. The eligibility requirements for Upward Mobility are that the individual be at or below General Schedule (GS) 9 or Wage Grade (WG) 9 levels. Applicants must also have been employed for a minimum of 90 days by a Federal agency. There were 108 females, 29 males, and 10 minorities in this group of subjects. The subjects ranged in age from 22 to 67 years old. The education completed by the subjects varied from a high school diploma to a graduate degree with 82 subjects having completed high school, 48 having completed 13 to 14 years, 6 having completed a Bachelor's degree, and one a Master's degree.

Procedure. The data for the analysis of assessment center scores consisted of eight skill area scores and level of education for each of the 137 subjects. Descriptive statistics on the assessment scores were computed on two groups (a) all 1979-1980 participants' scores (n = 137) and
(b) 1979-1980 scores without the retest group (n = 114).

The frequency distributions were of particular interest because a retest group was included in 1979 and because the overall distribution was negatively skewed. There was thus some possibility that test familiarity effects of the retest group could have skewed the assessment ratings in 1979.

Relationships among skill areas and educational levels were assessed by computing the Pearson product-moment correlations between the skill areas and education levels. Two of the original ten skill areas assessed were omitted from the analysis; Dependability because it had not been rated by the assessors and Interest in Electronics because it had been rated on the participants' expressed interest as well as the assessors' judgments of potential. The eight skill areas retained as variable were:

- Oral Communication
- Written Expression
- Ability to Interpret and Apply Instructions
- Initiative
- Analytical Ability
- Ability to Work Well with Others
- Flexibility/Adaptability
- Ability to Work Independently

The variable "years of completed education" was coded into these seven levels:

1 = 9 to 11 years completed
A series of independent \( t \) tests was also computed to compare the means of subpopulation groups on selected variables. In the first eight \( t \) tests, Group 1 was formed by the selection of all subjects whose education was greater than or equal to one-year of college \((n = 54)\) and Group 2 was composed of all remaining subjects \((n = 83)\). These two groups were then compared on each of the eight skill area dimensions.

Results and Discussion

An examination of the frequency distributions indicated that the inclusion of the retest group did not substantially alter the frequency distributions of the 1979-1980 assessment scores; the data base of all 137 subjects was thus used for the subsequent correlations and \( t \) tests. As can be seen in Table 1, scores for the eight skill areas were found to be significantly intercorrelated, with coefficients ranging from \( r = .21 \) to \( r = .61, p < .05 \), corrected for Type I error.

The intercorrelations among skill areas may be indicative of a general learning ability or, alternatively, a lack of discriminatory power in the measures used. In either
Table 1
Means and Intercorrelations among Skill Areas and Years of Education

Analysis 1
(_n = 137_)

<table>
<thead>
<tr>
<th>Skill Area</th>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>1</td>
<td>10.5</td>
<td>.50*</td>
<td>.59*</td>
<td>.52*</td>
<td>.58*</td>
<td>.40*</td>
<td>.44*</td>
<td>.51*</td>
<td>-.05</td>
</tr>
<tr>
<td>Written Expression</td>
<td>2</td>
<td>6.3</td>
<td>.61*</td>
<td>.24*</td>
<td>.49*</td>
<td>.17*</td>
<td>.32*</td>
<td>.56*</td>
<td>.19*</td>
<td></td>
</tr>
<tr>
<td>Ability to Interpret &amp; Apply Instructions</td>
<td>3</td>
<td>10.3</td>
<td>.45*</td>
<td>.59*</td>
<td>.35*</td>
<td>.41*</td>
<td>.61*</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiative</td>
<td>4</td>
<td>6.8</td>
<td>.55*</td>
<td>.49*</td>
<td>.55*</td>
<td>.35*</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytical Ability</td>
<td>5</td>
<td>9.8</td>
<td>.37*</td>
<td>.46*</td>
<td>.58*</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to Work Well With Others</td>
<td>6</td>
<td>6.9</td>
<td>.54*</td>
<td>.21*</td>
<td>-.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexibility/Adaptability</td>
<td>7</td>
<td>7.0</td>
<td>.32*</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to Work Independently</td>
<td>8</td>
<td>7.0</td>
<td>.17*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Completed Education</td>
<td>9</td>
<td>12.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Values</td>
<td>12</td>
<td>8</td>
<td>12</td>
<td>8</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

Note: Means are located on the diagonal

* p < .05, corrected for Type I error
case, the global qualities of these particular skill areas may be a necessary function of this assessment center's objectives. The elements used to assess Upward Mobility potential must encompass job-relevant criteria for a broad range of positions as well as being relevant to potential ability, rather than education or experience.

In contrast to the intercorrelations among skill areas, education was found to be significantly correlated with only two skill areas: Written Communications ($r = .19$, $p < .05$) and Ability to Work Independently ($r = .17$, $p < .05$). Although these correlations were significant, it should be noted that the proportion of variance in one variable that was explained by the other was negligible ($r^2 = .04$ and .03, respectively). Overall, the correlations indicated that the relationship between educational level and assessment ratings is quite weak; the coefficients of determination were so minimal that, in effect, the assessment scores can be considered independent of education levels.

Another way to examine the relationship between education and skill area scores was by means of $t$ tests comparing the two groups of more- and less-educated subjects. As would be expected given the results of the correlational analysis just presented, only two comparisons were significant: Written Communications and Ability to Work Independently, $t (135) = 2.05$, $p = .042$ and $t (131.71) = 2.60$, $p = .01$, respectively. The $t$ tests again indicated that
these two skill areas were related to years of completed education. Subjects with higher levels of education (greater than or equal to one year of college) were more likely to score well on these two dimensions.

A final way to examine whether educational level, irrespective of skill area scores, was related to selection is to examine the current hiring pattern of supervisors who are using this register of scores and the educational levels of selected candidates (see Table 2). Out of the 137 subjects, only seven subjects had a Bachelor's degree or higher and 54 had one year of college or more; of the ten persons selected for jobs from the 1979-1980 register, only one has had a Bachelor's degree. Five of the ten selected participants had completed high school only, and the other four selectees had one to two years of college. Thus, out of a field of 137 competitors, 50% of the selected candidates had no education beyond high school. Subjects with higher levels of education were not more likely, as a group, to be selected for jobs.

Since the rationale for the use of assessment techniques is to assess potential ability, it is somewhat surprising that the relationship between education and skill areas has not been a focus of previous research. The data presented here suggest that "skills" might more accurately be interpreted as abilities since skills should logically be more influenced by formal learning experience and yet little
Given the high intercorrelations among skill areas, it is puzzling that only a weak relationship between educational levels and skill areas was found. The unexpectedly low correlations between ability and education may reflect limitations of the sample; in this group of subjects the distribution of the assessment ratings was negatively skewed and the distribution of educational levels was positively skewed. The correlations very likely underestimated the magnitude of the relationship because the variables were skewed and the range of the skill areas scores was truncated, as well.

Additionally, this sample of subjects is largely composed of women who have returned to the work force
after establishing a home and family. Potentially having a great deal of ability and yet lacking formal education and experience, these women are "qualified" only for clerical or secretarial positions. Given the background variables which characterize the participants, the weak relationship between ability and education becomes more comprehensible.
Method

Subjects. The subjects for the second analysis were 24 persons who were first assessed in 1978 and then reassessed in the August 1979 assessment center. These subjects were identified as a subgroup for a test-retest analysis. There were 23 females and one male with an age range from 22 to 47 years.

Procedure. Seven skill area dimensions were rated in both the 1978 and the 1979 assessment centers; an eighth skill area, Flexibility/Adaptability, was added to the process in 1979, so there was no measure for this variable in 1978. The seven variables used for this analysis are as listed:

Oral Expression
Written Expression
Analytical Ability
Ability to Work Well with Others
Initiative
Ability to Work Independently
Ability to Interpret and Apply Written Instructions
The 1978 and 1979 scores were correlated by computing Pearson product-moment correlations for each dimension. A 2 x 7 repeated-measures analysis of variance (ANOVA) for the two factors of measurement periods and skill areas was also computed to test the differences between means of the 1978 and 1979 groups.

Results and Discussion

The test-retest correlation coefficients of the skill area scores were all nonsignificant as can be seen in Table 3. The repeated-measures ANOVA (Table 4), however, provided additional information on the differences between the two measurement periods.

The ANOVA results included a significant main effect for skill areas, $F(23, 6) = 4.3509, p < .01$, and a significant interaction between measurement periods and skill areas, $F(6, 138) = 6.2462, p < .01$.

Since the interaction was significant, nine simple main effects tests were also computed, and three of the tests were significant. The tests in this series were Dunnized to protect Type I error rate. The 1979 scores for the skill areas Initiative and Analytical Ability were significantly higher than the 1978 scores; $F(1, 161) = 16.0904, p < .01$ and $F(1, 161) = 8.7837, p < .01$, respectively. Although only two were significant, the direction of the difference between the 1978 and 1979 scores for six of the seven skill areas is consistent with the developmental emphasis of the
Table 3
Correlations Between 1978 Assessment Scores and 1979 Assessment Scores
Analysis 2
(n = 24)

<table>
<thead>
<tr>
<th>Skill Areas</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Expression</td>
<td>.22</td>
</tr>
<tr>
<td>Written Expression</td>
<td>.23</td>
</tr>
<tr>
<td>Analytical Ability</td>
<td>-.13</td>
</tr>
<tr>
<td>Ability to Work Well with Others</td>
<td>-.06</td>
</tr>
<tr>
<td>Initiative</td>
<td>-.14</td>
</tr>
<tr>
<td>Ability to Work Independently</td>
<td>.20</td>
</tr>
<tr>
<td>Ability to Interpret and Apply</td>
<td>.09</td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
</tr>
</tbody>
</table>

Note: \( p > .05 \), for all variables

Upward Mobility Assessment/Development Program; 1978 participants would be expected to improve their skills in the one-year interval between tests.

The third simple main effects test was computed on skill areas at measurement period 1978; \( F (6, 276) = 7.2421, p < .01 \). Dunn's procedure for individual t tests was used to compare the seven variables scored in 1978 and eight significant differences between skill area means were found. For example, in 1978 subjects were rated significantly higher
Table 4
Analyses of Variance:
Skill Areas Measured in 1978 and 1979

Analysis 2

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>80.068</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>14.584</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement Periods</td>
<td>2.384</td>
<td>1</td>
<td>2.384</td>
<td>3.1242</td>
</tr>
<tr>
<td>Error for MP</td>
<td>17.551</td>
<td>23</td>
<td>.763</td>
<td></td>
</tr>
<tr>
<td>Skill Areas</td>
<td>5.602</td>
<td>6</td>
<td>.934</td>
<td>4.3509*</td>
</tr>
<tr>
<td>Error for SA</td>
<td>29.617</td>
<td>138</td>
<td>.215</td>
<td></td>
</tr>
<tr>
<td>MP x SA</td>
<td>2.207</td>
<td>6</td>
<td>.368</td>
<td>6.2462*</td>
</tr>
<tr>
<td>Error for MP x SA</td>
<td>8.123</td>
<td>138</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>Measurement Period at SA₁</td>
<td>.130</td>
<td>1</td>
<td>.130</td>
<td>.8150</td>
</tr>
<tr>
<td>SA₂</td>
<td>.255</td>
<td>1</td>
<td>.255</td>
<td>1.5987</td>
</tr>
<tr>
<td>SA₃</td>
<td>.017</td>
<td>1</td>
<td>.017</td>
<td>.1066</td>
</tr>
<tr>
<td>SA₄</td>
<td>2.567</td>
<td>1</td>
<td>2.567</td>
<td>16.0940*</td>
</tr>
<tr>
<td>SA₅</td>
<td>1.401</td>
<td>1</td>
<td>1.401</td>
<td>8.7837*</td>
</tr>
<tr>
<td>SA₆</td>
<td>.200</td>
<td>1</td>
<td>.200</td>
<td>1.2539</td>
</tr>
<tr>
<td>SA₇</td>
<td>.021</td>
<td>1</td>
<td>.021</td>
<td>.1317</td>
</tr>
<tr>
<td>Pooled Error MP + (MP x SA)</td>
<td>25.674</td>
<td>161</td>
<td>.160</td>
<td></td>
</tr>
<tr>
<td>Skill Area at MP₁</td>
<td>5.965</td>
<td>6</td>
<td>.990</td>
<td>7.2421*</td>
</tr>
<tr>
<td>Skill Area at MP₂</td>
<td>1.844</td>
<td>6</td>
<td>.310</td>
<td>2.2677</td>
</tr>
<tr>
<td>Pooled Error SA + (MP x SA)</td>
<td>37.739</td>
<td>276</td>
<td>.137</td>
<td></td>
</tr>
</tbody>
</table>

* p < .01
on Ability to Work Independently than on Analytical Ability. This pattern of scoring was not repeated in 1979, thus explaining the interaction between measurement periods and skill areas.

It is interesting that the two variables with the greatest increase in the one-year interval were Initiative and Analytical Ability. Initiative would logically appear to be susceptible to social or motivational effects in a retest situation; once the subject has learned what the expected behaviors are, the appropriate behaviors can be produced. The measurement differences for Analytical Ability, however, are not as easily explained. The rating of this dimension requires a demonstration of reasoning ability in the Analytical Exercise and an interview with the assessor to clarify the subject's reasoning in working the problem. The explanation of developmental effects for the improved ratings in 1979 gains credibility when the background characteristics of the subjects are considered. For example, all participants of the 1978 assessment center were provided the opportunity of reapplying to the 1979 assessment center, but only 24 of the 136 participants did reapply. This pool of subjects, then, seems to be highly motivated toward career development; their participation in the 1979 assessment center demonstrates such a commitment.

The nonsignificant reliability coefficients are difficult, if not impossible, to interpret because several
factors could have intervened. The unexpectedly weak relationship between the two measurement periods could reflect changes in the assessment process, historical/developmental changes in the participants' abilities, test familiarity effects, as well as statistical instability due to the small sample of 24 subjects. The major influencing factor is most likely to have been an insufficient range for the skill area scores. The restricted score range contributes to an attenuation of the correlation so that the small correlations may reflect a lack of relationship or the fact that there was little variation in skill area scores. Given the possible intervening variables and the truncated range of the data, no definitive statement about the retest stability of this assessment center can be made.
CHAPTER IV
ANALYSIS 3

Method

Subjects. Two groups of subjects were used for the performance evaluation comparison. Twenty-nine graduates of the Upward Mobility program who are established in a target position were matched to a comparison group of 29 employees who were selected for their positions via normal merit promotion procedures. The matching variables were job category and grade level. There were 11 females and 18 males in the matched group and 24 females and 5 males in the Upward Mobility group.

Performance Evaluation Form. The criterion for this analysis consisted of a current supervisor's performance evaluation on each employee. The evaluation form of 14 items was designed specifically for this study (see Table 5).

Procedure. Out of an original 30 pairs of subjects, 29 pairs were actually used in Analysis 3 because an evaluation could not be obtained for one pair member. The comparison of job selection through Upward Mobility versus normal merit procedures was assessed with a two-sample dependent $t$ test, comparing the summed ratings of the two groups. The group comparison results were then checked with
Table 5

SUPERVISORY EVALUATION OF DESIGNATED NAVWPNCEN EMPLOYEES

INSTRUCTIONS: The employee's immediate supervisor should evaluate the employee's work performance, using the following rating scale. Judgments about the employee should be based on demonstrated ability, as compared to other personnel assigned to the same or similar work.

THE INFORMATION PROVIDED WILL REMAIN CONFIDENTIAL, WILL NOT BE ENTERED INTO THE RECORDS OF THE EMPLOYEE, AND WILL BE USED ONLY FOR RESEARCH PURPOSES.

<table>
<thead>
<tr>
<th>Rating</th>
<th>1-Outstanding</th>
<th>2-Above Average</th>
<th>3-Average</th>
<th>4-Marginal</th>
<th>5- Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepts and follows directions</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Works as a team member</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wins the respect and confidence of others in performance of assigned tasks</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is self-starting on work assignments</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizes and completes assignments in a timely manner</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is productive</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrates interest &amp; enthusiasm in career field</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicates in writing</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicates orally</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performs problem analysis</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deals with new and different tasks</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Makes sound decisions and recommendations</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular &amp; punctual work attendance</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall evaluation of employee's work performance</td>
<td>( ) ( ) ( ) ( ) ( )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
a series of 14 individual t tests for each evaluation item. 

**Results and Discussion**

The dependent t test found no difference between the supervisors' evaluations of the two groups, t (28) = .2309, p > .05. Additionally, the series of 14 individual tests found no differences between items of the performance evaluation. These results were Dunnized to protect Type I error rate. The data indicate that the job performance of Upward Mobility employees at NAVWPENCEN is very similar to the performance of employees selected through normal merit procedures, when employee performance is rated by the supervisors.
CHAPTER V

DISCUSSION

The results did provide information concerning the effectiveness of the Upward Mobility Assessment/Development Program. Analysis 1 indicated that educational level contributed little to the selection process, certainly less than expected. Of the eight skill areas, all significantly intercorrelated, it is interesting that only two were related to education, Written Expression ($r = .19$) and Ability to Work Independently ($r = .17$). These results should dispel any question that the process is selecting on the basis of education rather than potential.

The results of Analysis 2 assessing retest reliability were ambiguous, due to the particular circumstances of this study. This may not be an uncommon finding, however, because the assessment center literature also reveals a surprising lack of information on retest stability, particularly for intervals longer than five months. A direction for future research would be to incorporate retest situations in the designs of ongoing assessment centers, because skill measure stability is relevant to the general reliability and validity of the skill measures.

Related to the question of retest stability is the
effect of a developmental emphasis in conjunction with the assessment process. The NAVWPNCEN assessment center does incorporate career development sessions in the two-day assessment center. An individualized guide for career development is prepared by the assessors for each participant to use as a basis for developing specific goals. With the feedback, clarification and counseling provided, participants can accomplish a thoughtful and comprehensive self-analysis of his/her own potential and skills. If participants do individually progress in skill refinement and development, a retest analysis might produce ambiguous results similar to those presented here. In other words, the assumption of the test-retest model that the extent of measurement stability over time is true variance may not be consistent with a developmental model. Further research is needed to clarify this issue.

The results of Analysis 3 indicated very clearly that the assessment center selection process is comparable to normal merit procedures at full performance levels. Evidence for the validity of Upward Mobility selection procedures can thus be inferred from the performance ratings; employees selected for potential ability are able, after training, to perform as competently as employees selected on the basis of prior experience and education.

The minimum requirements set by the "Standards and Ethical Considerations for Assessment Centers" (Moses et al.,
1975) are easily met by the NAVWPNCEN assessment process and are far exceeded in some areas. A particular strength of the Upward Mobility assessment center derives from its two days of assessor training. For example, each assessor gives a final feedback interview to an assigned participant; this interview is tape-recorded and the recording is given to the participant as a future resource. The assessor's awareness that the sessions will culminate in this feedback interview inspires a focused attention during training and an objective, professional attitude throughout the session. (Awareness of the feedback interview may also be reflected in the negatively skewed ratings.)

The assessors report in the end-of-session evaluations that their skills in communication, interviewing and objective observation have improved as a result of their experience as assessors. This "hidden benefit" could be having a positive impact on the organization; future evaluation research could address this aspect of the assessment process from a management training perspective.

Overall, the NAVWPNCEN Upward Mobility Assessment/Development Program offers an effective alternative to traditional selection procedures. Women, in particular, have benefited greatly from the NAVWPNCEN program; careers once unavailable to a dead-ended secretary have become real opportunities for change and growth. The Upward Mobility program at NAVWPNCEN has demonstrated that selection of
non-managerial talent in the public sector is well served by the assessment center process.
REFERENCES


