INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.

UMI Dissertation Information Service
University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106
Bart, Leslie Ann

THE ROLE OF THE ASSESSSEE IN DEVELOPMENTAL ASSESSMENT
CENTER: A FIELD EXPERIMENT

The Ohio State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
THE ROLE OF THE ASSESSEE IN
DEVELOPMENTAL ASSESSMENT CENTERS:
A FIELD EXPERIMENT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Leslie Ann Bart, B.A., M.A.

August, 1986

Dissertation Committee:

Lyle Schmidt
Keith Smith
Bruce Walsh
Tom Milburn

Approved by

L. J. Schmidt
Adviser
Department of Psychology
I wish to thank Keith Smith and the staff of Ohio Cooperative Extension Service for their time, effort, and continuous support of this project.
VITA

November 8, 1959 ................Born - Los Angeles, Calif.

1982 ..................................B.A., California State University, Northridge

1984 ..................................M.A., Ohio State University

Major Field: Industrial/Organizational Psychology

Minor Field: Organizational Communication
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .......................................................... ii

VITA ........................................................................ iii

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

CHAPTER  PAGE

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

An Overview of Assessment Centers ......................... 3
Developmental Assessment Centers .......................... 10
Self-Assessment ......................................................... 19
Adult Learning .......................................................... 27
Research Objectives ................................................... 33

II. METHOD ................................................................. 39

Participants ............................................................... 39
Design ...................................................................... 39
Procedure ................................................................. 41
Self-Assessment Training ........................................ 45
The Assessment Center ............................................ 48
Measures ................................................................. 54
Analysis Plan ........................................................... 56

III. RESULTS ................................................................. 57

Manipulation Check .................................................. 57
Self-Assessment - 16 Dimensions .............................. 59
Self-Assessment (Pretest) ......................................... 63
Self-Assessment (Posttest) ........................................ 65
Change in Self-Ratings .............................................. 67
Assessor Ratings - 16 Dimensions ............................. 69
Overall Rating ........................................................ 71
Assessor Ratings - By Group .................................... 72
Comparison of Assessor and Self-Ratings ................. 74
Comparison of Assessor and Self-Ratings - By Group 76
Correlation of Assessor and Self-Ratings ................... 79
Correlation of Assessor Ratings ............................... 80
Satisfaction .............................................................. 81
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pretest Self-Ratings - All 16 Dimensions............................</td>
<td>60</td>
</tr>
<tr>
<td>2. Posttest Self-Ratings - All 16 Dimensions.........................</td>
<td>60</td>
</tr>
<tr>
<td>3. Average Ratings by Dimension........................................</td>
<td>62</td>
</tr>
<tr>
<td>4. Average Rankings by Group - Pretest Self-Ratings...................</td>
<td>64</td>
</tr>
<tr>
<td>5. Average Rankings by Group - Posttest Self-Ratings..................</td>
<td>66</td>
</tr>
<tr>
<td>6. Differences Between Pretest and Posttest Ratings..................</td>
<td>69</td>
</tr>
<tr>
<td>7. Assessor Ratings - All 16 Dimensions................................</td>
<td>70</td>
</tr>
<tr>
<td>8. Average Ratings by Dimension - Assessor Ratings....................</td>
<td>71</td>
</tr>
<tr>
<td>9. Average Rankings by Group - Assessor Ratings........................</td>
<td>73</td>
</tr>
<tr>
<td>10. Differences Between Assessor and Pretest Ratings..................</td>
<td>75</td>
</tr>
<tr>
<td>11. Differences Between Assessor and Posttest Ratings..................</td>
<td>76</td>
</tr>
<tr>
<td>12. Differences Between Assessor Ratings and Pretest Ratings - By Group</td>
<td>77</td>
</tr>
<tr>
<td>13. Differences Between Assessor Ratings and Posttest Ratings - By Group</td>
<td>79</td>
</tr>
<tr>
<td>14. Correlation of Assessor Ratings.....................................</td>
<td>81</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

If there is one thing all organizations have in common, it is that they are all interested in having employees improve their performance. Improved employee performance will increase organizational efficiency. But improved performance will not occur automatically. An organization concerned with improving employee performance needs to see to it that the employees participate in training activities which will make them more effective on the job.

The organization can provide extrinsic rewards to motivate its employees. It can hold out the promise of pay raises and promotions to encourage employees to participate in developmental activities. But extrinsic rewards can serve to undermine intrinsic motivation (Deci, 1975). Employees who participate in developmental activities only because they are required by the organization are not likely to feel their abilities are in any way deficient.
Since change and growth cannot be forced on someone, the organization is better off encouraging internal motivation. To the extent that employees see areas where they are not as capable as they could be, and feel these deficiencies may keep them from attaining the things they desire, they will be motivated to initiate a self-development program.

For individuals to see areas where improvement is needed, they need to engage in self-assessment. This self-assessment must take the form of a comparison between the person's current performance levels and some criteria of ideal performance. Although this self-assessment could occur in many situations, one situation where self-assessment may be particularly appropriate is an assessment center. Many organizations that use assessment centers for developmental purposes assume that the feedback assessees receive following the assessment center will motivate them to seek out activities to develop areas where improvement is needed. But, as will be discussed, there is reason to doubt whether information from others in itself motivates self-development. Individuals also need to engage in active self-evaluation, since this information is more likely to be accepted and to motivate action.

Since the focus of this study is on the role of self-assessment in developmental assessment centers, the
following discussion will begin with an overview of assessment centers. The second section will discuss developmental assessment centers. The third section will deal with self-assessment, and the way self-assessments are developed. Since one premise of this paper is that adult learning theory can help explain why self-assessments should be made part of the assessment center process, the fourth section will present an overview of adult learning theory. The final section will state the research objectives and hypotheses for the study.

An Overview of Assessment Centers

Many methods have been used by organizations to try to evaluate or predict the managerial success of individuals. One of the most effective and increasingly popular methods is the use of the assessment center. According to Thornton and Byham (1982):

"An assessment center is a comprehensive, standardized procedure in which multiple assessment techniques such as situational exercises and job simulations (i.e., business games, discussion groups, reports, and presentations) are used to evaluate individual employees for various purposes. A number of trained management evaluators, who are not in direct supervisory capacity over the participants, conduct the assessment and make recommendations regarding the management potential and developmental needs of the participants. The results of the assessment are communicated to higher management and can be used for personnel decisions involving such things as promotion, transfers, and career planning. When the results are communicated to the participants, they form the basis for self-insight and development planning" (p. 1).
According to Thornton and Byham (1982), the use of multiple assessment procedures was first implemented in Germany in the 1930s. The purpose was to help in the selection of military officers. The Germans decided that paper and pencil tests alone did not give a good indication of success as an officer. Rather than assessing a variety of traits individually, German psychologists believed that a person needed to be studied as a whole. They further felt that assessment should take place by observing behavior in naturalistic, everyday situations. Intellectual abilities and personality characteristics were to be examined using tasks of a realistic and serious nature rather than abstract tests. The assessment procedure lasted two or three days and was conducted by two officers, one physician, and three psychologists. Although the simulations were crude and the administration of the exercises was unstandardized, the German military assessment program is notable as being one of the first attempts to use multiple assessment techniques and multiple assessors to evaluate complex behavior.

This early work by the Germans led to the use of similar assessment procedures by the British War Office Selection Boards during World War II. The goal of these boards was to identify potential officers for the British Army. Prior to the summer of 1942, officers were selected on the basis of a 20-minute interview of candidates who had
graduated from certain prominent schools that offered officer training programs. The selection panel relied on intuitive judgments based on information believed to be relevant to officer performance. The British military programs made several advances over the German efforts. Effective performance was more clearly defined. The exercises presented more realistic situations to the candidates, and they were better designed. The responsibility for assessment decisions shifted to military personnel, who were more knowledgeable about what it took to succeed as an officer than were psychologists or psychiatrists (Thornton & Byham, 1982)

The first fully developed multiple assessment program for selection and placement in the United States was developed by the Office of Strategic Services (OSS). The purpose of the OSS management program was to evaluate the personalities of candidates for a variety of positions in the OSS, including secret intelligence agents, saboteurs, and propaganda experts. Overlying the whole program of specific exercises was the requirement of a false cover story being maintained by each candidate. Each person assumed a false name and made up a cover story about his or her life and work. Assessors would try to trick the candidates into revealing themselves. In a sense, then, the entire three days of the assessment program comprised a
situational test of the person's ingenuity in making up a story and tenacity in maintaining it. The use of situational and performance exercises by the OSS marked a significant shift away from primary reliance on paper-and-pencil instruments, which had been used previously in the United States (Thornton & Byham, 1982).

Industrial usage of the assessment center began in 1956 when American Telephone and Telegraph (AT&T) began its Management Progress Study. The study was designed to try to "understand what characteristics (cognitive, motivational, and attitudinal) are important to the career progress of young employees from the time they take their first job in the Bell System and as they continue to move into middle and upper-management levels" (Cascio, 1982, p. 251). Candidates were evaluated on 25 variables which were later condensed into seven major areas. The candidates were appraised by the assessment techniques and then their careers were monitored and compared with the assessment data. Feedback of results was not given to the participants, their superiors, or any other company managers in order the minimize the effect of assessment on the individual's progress in the company. The data were so predictive of future success that today more than 1,100 businesses are using some sort of assessment center (Cascio, 1982).
Individuals may be required to participate in an assessment center at a certain point in their careers, they may be nominated to attend by their supervisors, or they may nominate themselves for attendance. Usually, a small group of individuals (6 to 12) is assessed at one time. Group assessment allows for group activities, and is a more efficient use of resources. The ratio of assessees to assessors is low, typically 2:1; this allows the assessors to closely observe the assessees.

During the course of the assessment center (usually 1 to 3 days), the assessees participate in a variety of activities designed to allow them to demonstrate skills and abilities that are necessary for successful on-the-job performance. Some commonly used assessment center exercises are:

1. Non-assigned role group discussion. The entire group of assessees participate in a discussion and make recommendations on how to resolve a problem they had been presented.

2. Assigned role group discussion. Each assessee is assigned a candidate for a job opening and must try to persuade the other assessees to vote for his/her candidate.
3. In-Basket. The assessee deals with a packet of letters, memos, phone messages, etc. which a manager might find in his/her "in-basket."

4. Interview simulation. The assessee takes the role of a manager giving a performance appraisal to a subordinate.

5. Written case study. The assessee is given a written description of a problem and prepares a written report on how he/she would handle the problem.

6. Fact-finding. The assessee asks a role-player questions about a problem which supposedly developed in the office and then gives an oral report on how he/she would deal with the situation.

The exercises are chosen based on an analysis of the characteristics a person needs to be a successful manager. The assessors observe the assessee as they perform the oral activities and evaluate the assessee's written work. Once all activities have been completed, all of the assessors assigned to the assessee individually rate the assessee on every dimension. The assessors then discuss the assessee, reach a consensus decision of the assessee's performance on
each dimension, and make a prediction of overall potential or ability.

Following the assessment center, each assessee has a feedback interview with one or more of the assessors. The assessees are told how they performed in the assessment center exercises, although they are probably not told their exact scores on each dimension. They are told about their areas of relative strength and weakness, and they may be given suggestions about ways to develop themselves.

Assessment Center Research

Most of the research on assessment centers has examined whether assessment center ratings predict future success in the organization. A review of the literature by Klimoski and Strickland (1977) showed assessment center ratings to be consistently valid predictors of number of future promotions. Cohen, Moses, and Byham (1974) also reviewed the assessment center literature and found assessment center ratings were consistently related to future promotions, but not highly predictive of job performance. In a more recent study, Turnage and Muchinsky (1984) also found assessment center evaluations could predict future promotions, but were not related to job performance.

Although many organizations use their assessment centers for developmental purposes, there has been no
published research on the usefulness of assessment centers for employee development.

**Developmental Assessment Centers**

Some assessment centers are used for selection or promotion, others are used to diagnose an employee's developmental needs. Even though an assessment center is supposed to be measuring stable traits, the assumption behind developmental assessment centers is that most motivated people can develop some of their weak areas.

The specific design of an assessment center will depend, in part, on the purpose for the center. Assessment centers used for employee development are usually longer and more complex than centers used for selection or promotion decisions. Developmental assessment centers typically have more dimensions, so a larger number of exercises are needed to obtain the necessary information about an assessee's ability on each dimension. The final overall decision is far less important than the decisions made for each dimension, and in many organizations having diagnostic centers, no overall decision is made. In a diagnostic center, each dimension must be measured with a high degree of reliability and validity because decisions about an individual's developmental needs will be made on each dimension. The reports generated are lengthier and the
feedback given to assessees is more detailed than would be the case in an assessment center used for selection or appraisal of potential. Although these few differences usually exist, assessment center used for developmental purposes are highly similar in structure to those used for selection or promotion. A major premise of this paper is that this is inappropriate; assessment centers used for development need be structured differently than those used for selection or promotion.

A developmental assessment center is the first stage in a developmental program. It is where each individual's developmental needs are determined. Areas of weakness can then be strengthened through appropriate training activities. However, there are several limitations to existing practices: organizational involvement in the training activities is typically low. Individuals are expected to take the initiative for their own development, and the goal of development is the individual's own increased satisfaction and improved job performance (Boehm & Hoyle, 1977).

**The Feedback Interview**

A vital part of the developmental assessment center is the feedback interview in which the judgments of the assessors are made available to the assessees. And if
change is to occur as a result of the assessment center, then the feedback offered by the assessors must be accepted by the assessees.

Ashford and Cummings (1983) have proposed that individuals frequently become very involved in the feedback process, actively seeking out information from various sources with which to evaluate themselves. Individuals will be most motivated to seek out feedback which they perceive to be valuable to them. When they perceive that the feedback will help them achieve desired goals, they will actively monitor their information environment for personally relevant information and directly seek such information from others in their environment.

The choice of a feedback source will depend upon what sources are easily accessible. Greller and Herold (1975) found that oneself was the source relied on most heavily, followed by information available from the task itself. Individuals will seek out feedback from others only when self-supplied or task-supplied feedback is not readily available. Yet the source of the feedback may be the most important influence on the extent to which recipients accept the feedback (Ilgen, Fisher, & Taylor, 1979). This seems to be due to the degree of credibility which recipients attribute to the source. Although assessees may view the
assessors as credible sources, they often will be more inclined to accept self-supplied or task-supplied feedback.

Although some assesseses will be more motivated than others to participate in an assessment center, most tend to be active information processors. The information gained in the assessment center can help them to achieve valued goals, either becoming more effective in their present jobs, or moving up in the organization. If they know what dimensions are being assessed, they will know what the organization considers important for success.

Self-evaluation is the final stage in any feedback process. Individuals will seek out feedback from various sources; they will tend to accept or reject the feedback depending on the credibility of its source and its congruence with their self-concept (Ilgen, Fisher & Taylor, 1979). They may then use the information they have obtained to make their own judgments about their performance and its implication about their abilities. Even though the assessors' judgments may be highly valid, they will be useless if they do not become a part of an individual's self-evaluations.

Advantages of Developmental Assessment Centers

An assessment center has several advantages which make it useful for employee development. An assessment center is
the first step in the developmental process. It provides a diagnosis of developmental needs so that the person knows which developmental activities would be appropriate.

The assessment center approach is systematic; everyone is presented the same exercises and evaluated according to the same criteria. This is unlike the way employees are typically evaluated in organizations, where each employee is evaluated individually by his/her supervisor. In addition, assessors are trained to evaluate the person's behavior, while supervisors generally are not. Therefore, both the employees and the organization can have more confidence in assessment center evaluations than in a typical performance appraisal, where some supervisors will do much better at evaluating performance than others.

Another advantage to having assessment centers is that following the assessment center, assessees are given feedback about their performance which may help them direct their self-development efforts. Many supervisors do not give their employees much feedback and sometimes employees work independently and their supervisor is unable to observe their performance.

Disadvantages of Developmental Assessment Centers

Assessment centers also have disadvantages which lessen their usefulness for developmental purposes.
First, assessment centers are presumed to measure stable traits that people express in a wide variety of situations. However, people's behavior may be very situation specific (Mischel, 1973; Turnage & Muchinsky, 1982). The way people behave in the assessment center exercises may not be the way they behave in other situations, so it would be inappropriate to make inferences about their ability based solely on observations of their performance in the assessment center setting.

Second, in developmental assessment centers each dimension must be evaluated independently since assesseses are given feedback about specific areas where development is needed. But assessor ratings have been found to show halo error (Turnage & Muchinsky, 1982). Assesseses may receive high ratings on each dimension or low ratings in all areas. This raises doubt about whether it is appropriate to use these ratings for development when they appear to be based on overall perceptions rather than evaluations of performance on specific dimensions.

Third, Klimoski and Brickner (1985) have suggested that, to a large extent, assessment centers are measuring intellectual functioning. Since intelligence is not entirely subject to development, it would be inappropriate to use only intelligence measures for assessment of developmental needs.
Fourth, assessment center ratings are stable. Bray, Campbell, and Grant (1979) found that when people were re-assessed after a period of several years, their ratings changed only slightly. However, this does not imply that people cannot change; only that people are not likely to change if the assessment is not followed by systematic developmental activities.

Fifth, an important part of developmental assessment centers is the feedback interview. Since assessees are expected to take the initiative for their own development, it is vital that the assessees accept the feedback they receive from the assessors. But assessees may be resistant to negative feedback from the assessors. Rather than accepting the feedback, they may counterargue and eventually disagree with the assessor even more than was originally the case (Petty, Ostrom, & Brock, 1981).

Can People Be Developed?

If assessment centers are to be useful for employee development, then they must focus on dimensions that are potentially developable. The training literature suggests that the development of adults is very difficult (Goldstein, 1980). But a distinction needs to be maintained between those skills that could be developed through appropriate
developmental activities, albeit with difficulty, and abilities which are innate and not developable.

Almost every assessment center assesses oral and written communication skills. Although some people have more aptitude in these areas than do others, they are essentially skills that are learned in the classroom, and improve with practice. Therefore, these skills could be improved by a person who is motivated to do so (Boehm & Hoyle, 1977).

Assessment centers also typically measure interpersonal skills, or the ability to deal effectively with others. Several organizations have found that interpersonal skills can be developed through behavior modeling (Goldstein & Sorcher, 1974; Latham & Saari, 1979). Employees participate in a workshop where they watch videotapes of managers successfully handling a variety of interpersonal situations. Then the employees practice their skills using the techniques they saw in the videotape and receive feedback from the trainer and the other members of the group. The concept behind behavior modeling is that people can learn new skills by observing the actions of others.

Another technique for developing interpersonal skills is assertiveness training. People who have trouble being assertive are unable to state their needs and feelings in a positive, forceful, but non-threatening manner.
Assertiveness training involves identifying problem situations, providing examples of appropriate and inappropriate behavior (usually through the use of models), practice of assertive behaviors by the person, and feedback and reinforcement of appropriate behaviors by the trainer (Osipow, Walsh, & Tosi, 1984). Research studies have consistently demonstrated the effectiveness of assertiveness training, and the initial positive effects are maintained over time (Galassi, Kostka, & Galassi, 1975).

A dimension such as leadership is much more problematic since it is unclear what constitutes effective leadership. Much of the research has focused on leadership as a function of the fit between leader behavior and situational characteristics (Fiedler, 1967; House, 1971). A person is not a good or bad leader; rather, a person is effective when his/her personal style matches the requirements of a particular situation. Fiedler, Chermers, and Mahar (1976) developed a leadership training program called Leader Match. Using this program, individuals learn to modify their situation to match their personal leadership style. Fiedler and Mahar (1979) reviewed twelve studies testing the effectiveness of Leader Match and found all twelve studies supported Leader Match training. But Kabanoff (1981) argued that none of the studies used appropriate criteria of leader
effectiveness and that the results of many studies were open to alternative explanations.

If assessment centers are measuring intelligence, as suggested by Klimoski and Brickner (1985), then the effects of developmental activities may be limited. Wechsler (1958) believed that intelligence is "the capacity to act purposefully, to think rationally, and to deal effectively with the environment" (p. 7). Many years later, he suggested that intelligence is "the capacity to understand the world and the resourcefulness to cope with its challenges" (Wechsler, 1975, p. 139). These broad definitions appear to include some of the dimensions commonly measured in assessment centers, such as planning, judgement and decision- making, initiative, adaptability, objectivity, and perception. Although none of these abilities may be entirely a function of intelligence, and some increase in ability levels is attainable through appropriate training, it may be unrealistic to expect substantive change in these areas.

**Self-Assessment**

Self-assessment has not been widely used in industrial settings. The reasons for this lack of interest has been the perception that 1) self-assessment typically results in an overestimation of the individual's abilities, and 2)
individuals cannot analyze themselves objectively or reliably enough to make accurate appraisals (Levine, Flory, & Ash, 1977).

According to Thornton (1980), the finding of the majority of studies on self-assessment has been that individuals rate themselves higher than they are rated by comparison groups such as supervisors or peers. This finding has occurred across different types of jobs (i.e., from clerks to executives), and with different types of rating scales. Self-appraisals do not show much variability in ratings. The majority of self-raters place themselves on the high end of the scale, and it is rare to find anyone rating themself as less than average (DeNisi & Shaw, 1977; Meyer, 1980). Individuals seem to have unrealistically favorable perceptions of their own abilities. They see themselves as at least average and usually better than average.

Do self-ratings correlate with the ratings of others? The findings are inconsistent. Holzbach (1978) found sporadic agreement between self and superior ratings, but virtually no agreement between self and peer ratings. Brief, Aldag, and Van Sell (1977) reviewed nine studies which reported correlations between superior and subordinate ratings of job performance. Correlations were generally modest, but varied widely, from a low of .05 (Klimoski &
London, 1974) to a high of .60 (Williams & Seiler, 1973). Stronger relationships between self and superior ratings have been found when the criteria was effort (Porter & Lawler, 1968; Williams & Seiler, 1973). Individuals and their supervisors seem to have similar perceptions of how much effort the individual is expending but dissimilar opinions about whether that effort leads to effective performance.

Self-appraisals can also be compared to objective tests of performance. Levine, Flory, and Ash (1977) found that individuals could rate their future performance fairly accurately on various clerical tasks. Ash (1980) obtained modest correlations between self-assessments and typing test scores. DeNisi and Shaw (1977) compared self-ratings of ability with test measures of visual pursuit, manual speed and accuracy, spatial orientation, verbal and numerical ability, intelligence, and mechanical comprehension. Of the eleven correlations, ten were statistically significant, but too low to have any practical value.

There are at least two reasons why self-ratings differ from the ratings of others. First, individuals may need to distort their evaluations to maintain their self-esteem (Bem, 1972). They will be tempted to blame environmental factors when performance is poor, rather than attribute their performance to stable traits or abilities. They may
also be resistant to information from others that they are lacking in ability. They may see the person rating them as biased, as lacking in understanding of what it takes to succeed, or as not having had an adequate opportunity to observe the workers performance. They may be reluctant to accept feedback from others which is discrepant with their own self-evaluation.

Second, individuals and those observing them may have different ideas of what constitutes good and poor performance. Workers may be rating themselves highly because they have met their own standards of performance. Their supervisors may rate them lower because the supervisors are assessing the workers using different standards (Latham & Wexley, 1981).

Towards a Theory of Self-Assessment

The research in self-assessment is largely atheoretical (Heneman, 1980). But Bandura (1978) has made a start in this area by formulating a theory which could help explain how self-assessments are developed.

First, people derive much of their information about themselves from direct experience of the effects produced by their actions. Each of us has the opportunity to become quite knowledgeable about our own attributes under a variety of conditions. Second, people learn from vicarious
experience as well as direct experience. Knowledge can be accrued by observing the actions of others and the reactions they receive. A third way in which we learn about ourselves is from the judgments that others express about us. Their positive and negative feedback about our behavior tells us how we are viewed by the world. And fourth, we are able to use our powers of reasoning to derive new knowledge about things that go beyond our direct experiences (Bandura, 1978).

How will these processes affect the self-ratings made by assesses in assessment centers? Since people use their experiences to make their evaluations, they could be expected to use their behaviors in the assessment center exercises as a source of information. However, they probably will not limit their judgments to what happened during the assessment center. They will also use their past experiences, which may or may not lead to the same conclusions as a judgment based only on behavior in assessment center exercises.

Individuals' evaluations are also shaped by their observations of others. Although there are group activities in most assessment centers, assesses do not have much opportunity to compare their own performance to those of others during the assessment center.
Since people use others' evaluations of themselves in making their own evaluations, the information they receive from assessors will affect their self-evaluations. However, the judgments voiced by others may not be as important a factor as a person's direct experiences.

Self-assessment can also lead to self-selection, and this could be a function of the final process proposed by Bandura (1978), making inferences that go beyond direct experience. The assessment center could act as a realistic job preview, giving people enough information about themselves and about the characteristics of the job that they can then infer their suitability for a particular position (Wanous, 1980). However, this would more likely occur in assessment centers intended for selection or promotion purposes.

**Self-Presentation and Defensive Attributions**

A repeated finding in the literature on self-assessment has been that individuals rate themselves higher than they are rated by others (Thornton, 1980). Studies have also shown there is a tendency for individuals to make self-attributions for their own positive behaviors and external attributions for negative behavior (Kelley, 1973). Kelley's (1973) explanation for these results was that people have a need to protect and enhance their self-esteem,
and they do this by taking credit for good acts, denying blame for bad outcomes, and giving themselves overall positive evaluations.

However, Bradley (1978) suggested that under certain conditions, esteem needs may best be served by accepting responsibility for negative outcomes. If your self-evaluation will be made public, and if others will have the opportunity to view your behavior in areas relevant to the dimensions on which you have evaluated yourself, the best strategy may be modesty. It would be embarrassing (and could cause loss of self-esteem) to give yourself a positive self-evaluation and then have others watch you perform poorly.

This notion has particular relevance to the type of self-assessments which assesses may make. If they are aware that the assessors will see their self-ratings, they are likely to be fairly modest in their evaluations of themselves. After all, the assessors will have an opportunity to view the assesses' performance on the various dimensions. Although assesses may revise their public self-assessments to bring them in line with what they expect to receive from the assessors, this does not mean they will actually believe they deserve such a modest evaluation. While people may respond to the social demands of a particular situation, they will not readily change
their self-concept (Baumeister & Jones, 1978). This indicates a need for caution when comparing self and assessor ratings of assessee performance. If they are similar, this does not necessarily mean that assessee and assessors are in agreement about the assessee's abilities. The assessee may merely be mimicking what they think the assessors will say.

Facilitating Self-Assessment

Schrauger and Osberg (1981) made some suggestions for facilitating self-assessments. Their remarks can be adapted for use in assessment centers.

First, assesses need to be given clear, unambiguous definitions of the dimensions on which they are to evaluate themselves. Their ability to accurately assess their abilities will depend, in part, on whether they understand exactly what they are to be assessing.

Second, assesses should be encouraged and motivated to be accurate. If the assessment center is used for selection or promotion decisions as well as for development, the assesses will be reluctant to give themselves low ratings. An assessment center intended for developmental purposes should be used solely for that purpose, and assesses should be told that no one rates highly on every dimension and that it is in their own best interests to note areas where they
are weak, so that they can be given assistance in developing these areas. All of this, of course, assumes that a high degree of trust exists between the asseesees and the people in charge of the assessment center.

Third, distinctions need to be maintained between an evaluation of behavior in the assessment center and an evaluation of behavior in general. The assessors have only the assessment center exercises with which to make their evaluations. The asseesees have two sources of information--their performance in the assessment center and their entire past life history. Self-evaluations based on these two different sources may or may not be congruent.

**Adult Learning**

A developmental assessment center can be considered a learning activity since it is expected that asseesees will learn new things about themselves through increased self-insight (Thornton & Byham, 1982). Therefore, the assessment center needs to be structured so that learning will occur.

Lindeman (1926) identified some key assumptions about adult learners; these assumptions still constitute the cornerstone of adult learning theory:
1. Adults are motivated to learn as they experience needs and interests that learning will satisfy.

2. Experience is the richest resource for adult learning; therefore, the core methodology of adult education is the analysis of experience.

3. Adults need to be self-directing; therefore, teachers of adults should assist with, but not direct, the learning process.

Adult learning (or andragogy) operates on a different base of assumptions than those governing learning by children (pedagogy) (Knowles, 1978). And in several ways the structure of most assessment centers facilitates learning by adults.

As individuals mature, they accumulate a reservoir of experiences which provides them with a base with which to relate new learnings. Discussions, simulations, and projects are the best techniques with which to tap into these experiences, unlike the use of lectures and readings which characterize traditional teaching techniques (Knowles, 1978). All assessment center exercises (unless personality tests are used) are of the action/doing type so they are suited for use by adults.
Learning experiences need to be relevant to the types of problems individuals face everyday. They may have little patience with topics for which they cannot see an immediate application (Knowles, 1978). Assessment center exercises are designed to simulate the type of problems managers might find on the job. Therefore, the activities may be seen as relevant.

Individuals must be ready to learn what they are being taught. They must have had appropriate past experiences, and must be in a position where they "need" the information they are receiving (Knowles, 1978). Insofar as developmental assessment centers provide assessees with information which will help them to perform better in their jobs, it may be construed as "needed" information.

Where typical assessment centers may fall short is in the assessors' direction of the learning process. The assessees are put through the exercises, evaluated by the assessors, and given feedback about their performance, with little or no opportunity to play anything but a passive role in the entire process.

This lack of self-direction may limit the usefulness of the assessment center as a learning experience. There are several reasons why assessees need to be given a more active role in the assessment center process.
First, when individuals find themselves in a situation where they are not allowed to be self-directing, tensions occur between the situation and the person's concept of himself or herself as an autonomous, independent individual. This tension is likely to lead to resentment and resistance, which means assessor feedback will not be accepted and assessees may have negative feelings about the entire assessment process (Knowles, 1978).

Second, one thing that enhances adult learning is learner involvement in planning and conducting the learning activity. When learners are given an active role in the learning process, they report greater satisfaction with the learning activity (Long, 1983). And, in general, the more active the learners are in a learning situation, the more they will learn (Knowles, 1950).

Third, an individual's motivation to learn and change is an extremely important consideration. One repeated finding in the adult learning literature is that a self-diagnosed need for learning produces much greater motivation to learn than an externally diagnosed need (Knowles, 1980).

In sum, then, if assessees were given a more active role in the assessment center process, they would be more satisfied with the assessment center, more motivated to change, and would learn more about themselves.
Person-Centered Learning

Another way to look at the issue of learner involvement is through Carl Rogers' concept of person-centered or experiential learning. According to Rogers' (1983):

"Experiential learning...has a quality of personal involvement - the whole person in both feeling and cognitive aspects being in the learning event. It is self-initiated. Even when the impetus or stimulus comes from the outside, the sense of discovery, of reaching out, of grasping and comprehending, comes from within. It is pervasive. It makes a difference in the behavior, the attitudes, perhaps even the personality of the learner. It is evaluated by the learner. She knows whether it is meeting her needs, whether it leads toward what she wants to know, whether it illuminates the dark area of ignorance she is experiencing. The locus of evaluation, we might say, resides definitely in the learner. Its essence is meaning. When such learning takes place, the element of meaning to the learner is built into the whole experience" (p. 20).

Under the traditional learning mode, the teacher decides what the learner should know, takes responsibility for imparting the knowledge, then evaluates whether the student has learned what he/she has been told. In contrast, with the person-centered approach, the students decide what they wish to learn, and evaluate their own progress. Rogers (1983) feels that in this type of environment the learning tends to be deeper, proceeds at a more rapid rate, and is more pervasive in the life and behavior of the student than is learning acquired in the traditional matter. This is because the learning is self-initiated, and learners have invested a good deal of themselves in the process.
Boehm and Hoyle (1977) proposed that in diagnostic assessment centers, the assessor's role should become more that of a counselor and less that of an evaluator. The evaluative function should be placed primarily in the hands of the assessee. This could be accomplished by providing assesses with the same type of information about the evaluation process that is normally only given to the assessors. Thornton and Byham (1982) reported that assesses lowered their self-evaluations following assessor training. The information they gained in assessor training gave them added insight into their own abilities by showing them the discrepancy between their own typical behavior and what would be considered ideal behavior. If assesses received similar information prior to participating in the assessment center, they, too, would gain in self-awareness.

There is an additional reason for encouraging self-diagnosis. One problem with assessor ratings mentioned previously is that they tend to involve halo. The assessors do not seem to evaluate each dimension independently. When assessment centers are used to predict overall potential, halo may not be a problem as long as the overall prediction is accurate. But when the purpose is an assessment of an individual's developmental needs, then each dimension needs to be evaluated separately. Self-appraisals, on the other hand, tend to show relatively low inter-scale correlations
(Thornton, 1980). Less halo may suggest that intra-individual strengths and weaknesses are being diagnosed more completely. This makes self-appraisals particularly well-suited to the identification of areas where development is needed.

Research Objectives

The purpose of a developmental assessment center is to evaluate an employee's developmental needs. Once these needs have been diagnosed, employees can engage in appropriate developmental activities which will make them more effective on the job.

If the current structure of assessment centers is an effective method of assessing employees' developmental needs, then various outcomes should result from an employee's participation in the assessment center:

1. Employees will change their own evaluations of their abilities following the assessment center, particularly after the feedback interview with the asessors.

2. Once they have been told what areas need developing, employees will seek out appropriate developmental activities, such as courses or workshops.
3. Employees will become more effective at their jobs as a result of participating in these developmental activities.

If these outcomes do not occur, then the assessment center needs to be restructured to make it more effective. One way this could be done would be by letting the assesses participate more in the assessment center process. Increased participation leads to more learning, greater motivation to change, and greater satisfaction with the learning process (Knowles, 1978; Knowles, 1980; Long, 1983).

In many adult learning situations, involvement is increased by allowing the learners to participate in the planning or development of the learning activity. Involvement can also be increased by allowing learners to evaluate their own needs and their own performance in the learning activity (Knowles, 1978). In an assessment center situation, certain limitations on the nature and amount of assessee involvement exist. Since the development of the exercises and the training of the assessors is expensive and time-consuming, and the assessment center may be used for selection and promotion as well as development, the assessment center exercises will need to be left intact. Changes can be made prior to the assessment center. Typically, an assessee comes to the assessment center naive
as to what will occur or how they will be evaluated. One way to increase assessesee involvement would be to explain how they will be assessed and encourage them to assess themselves. Although assessesees are likely to be self-evaluative during an assessment center, they should evaluate themselves on the same dimensions, using the same criteria as the assessors for their self-evaluations to be useful.

According to Websters Third New International Dictionary of the English Language: Unabridged (1961), involve is defined as "to draw in as a participant...to occupy (oneself) absorbingly." These definitions are consistent with the idea of active participation presented earlier.

For the purpose of this study, then, working definition of involvement will be "the extent to which assessesees are encouraged to assess themselves, and trained in self-assessment." Although this definition may seem limited, since the purpose of an assessment center is an evaluation of the assessesees' abilities, it is an appropriate one. The independent variable will be level of involvement, which will be manipulated by varying the amount of information assessesees are given about how to assess themselves, and the extent to which they are encouraged to assess themselves. The dependent variables will be accuracy
of self-ratings (which will be determined by comparing self-ratings to assessor ratings), their satisfaction, the amount they feel they learned about themselves, the perceived validity of the assessment center process, and their intentions to engage in developmental activities.

The finding of the majority of studies on self-assessment has been that individuals rate themselves higher than they are rated by others (Thornton, 1980). However, training in self-assessment can give people a more realistic view of their own abilities, causing them to make more accurate self-assessments. In addition, people should be encouraged and motivated to be accurate (Schrauger & Osberg, 1981). Therefore, individuals who are trained in self-assessment, and encouraged to be accurate, will make self-assessments which are closer to the ratings made by others. Since the assesses will be told their self-assessments will be kept confidential, self-presentation (Bradley, 1978) will not be a factor.

Hypothesis 1: Assessees in the "high" involvement condition will make self-ratings which are closer to the assessor ratings than will assesses in the "moderate" involvement condition, and assesses in the "moderate" involvement condition will make self-ratings which are closer to the assessor's ratings than assesses in the "low" involvement condition.

Adult learning theory states that learning by adults is different than learning by children. Adults need to be
active participants in the learning process, not passive recipients of other peoples attempts to teach them. When adults are actively involved in the learning process, the learning activities are more useful and enjoyable than if they played only a passive role (Knowles, 1978; Knowles, 1980; Long, 1983). The next four hypotheses deal with the positive outcomes which result from a person's active involvement in a learning activity.

According to Long (1983), the more active the role learners are given, the greater their reported satisfaction with the learning process.

Hypothesis 2: Assessees in the "high" involvement condition will report greater satisfaction than assessees in the "low" involvement condition. Reports of assessees in the "moderate" involvement condition will fall in between the other two.

The more active the learners are in a learning situation, the more they will learn (Knowles, 1950).

Hypothesis 3: Assessees in the "high" involvement condition will report that they learned more about themselves than assessees in the "low" involvement condition. Reports of assessees in the "moderate" involvement condition will fall in between the other two.

Giving assessees prior training on assessment centers and self-assessment can also be a public relations tool. When learners are involved in the process, they are more likely to accept the process as valid (Knowles, 1978).

Hypothesis 4: Assessees in the "high" involvement condition will be more likely to report that the assessment center is a valid way to assess abilities
than will asseseees in the "low" involvement condition. Reports of Assessees in the "moderate" involvement condition will fall in between the other two.

When learners are involved in the learning process, and evaluate their own progress, they are more motivated to try and improve themselves (Knowles, 1980).

Hypothesis 5: Assessees in the "high" involvement condition will report more intentions to engage in developmental activities than assessees in the "low" involvement condition. Reports of assessees in the "moderate" involvement condition will fall in between the other two.
CHAPTER II

METHOD

Participants

The assessment center used in this study was developed by Ohio Cooperative Extension Service for the purpose of assessing the developmental needs of its employees. All County Chairpersons were required to attend the assessment center during the summer of 1985. All Supervisors-in-Training were also required to attend. This study included five Supervisors-in-Training. Once the Supervisors-in-Training were assessed, the assessment center was opened to County agents. Thirteen County agents participated in this study. They were all recommended by their supervisors because they were believed to have managerial abilities.

Design

The independent variable was level of involvement in the assessment center. Participants were assigned to either the "low", "moderate", or "high" involvement conditions depending on the date of the assessment center they attended.
All three groups were given a pretest in which they were given the list of 16 dimensions on which they would be assessed in the assessment center and were asked to rate themselves on each dimension.

The "low" involvement group was given no further instructions on self-assessment. The "moderate" involvement group was encouraged to assess themselves but was given no training in self-assessment. The "high" involvement group was encouraged to assess themselves and received training in self-assessment prior to the assessment center. Immediately after completion of the assessment center exercises, participants were given a posttest on which they were asked to rate themselves again on the assessment center dimensions. All assesses were also rated by the assessors on the same dimensions the assesses were using to rate themselves.

The dependent variables were 1) differences between posttest ratings and assessor ratings, 2) satisfaction, 3) amount of learning, 4) perceived validity of the assessment center, and 5) intentions to engage in developmental activities. The 2nd through 5th of these dependent variables were assessed with two questionnaires. One questionnaire was given immediately after completion of the assessment center exercises (see Appendix A); and the other questionnaire was sent to assesses four to eight weeks
after the assessment center (see Appendix B). The "moderate" and "high" groups received both questionnaires; assesses in the "low" group did not receive the first questionnaire.

Therefore, the design of the study was as follows:

<table>
<thead>
<tr>
<th>Level of Involvement</th>
<th>X1</th>
<th>AC</th>
<th>X2</th>
<th>Y</th>
<th>Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>X1</td>
<td>E</td>
<td>AC</td>
<td>X2-Q1</td>
<td>Y</td>
</tr>
<tr>
<td>Moderate</td>
<td>X1</td>
<td>T</td>
<td>AC</td>
<td>X2-Q1</td>
<td>Y</td>
</tr>
<tr>
<td>High</td>
<td>X1</td>
<td>T</td>
<td>AC</td>
<td>X2-Q1</td>
<td>Y</td>
</tr>
</tbody>
</table>

X1 = pretest (self-assessment)  
E = encouraged to self-assess  
T = assessee training  
AC = assessment center  
X2 = posttest (self-assessment)  
Q1 = questionnaire 1  
Y = assessor ratings  
Q2 = questionnaire 2  

Procedure

There were 18 participants in this study. The "low" involvement group consisted of the five Supervisors-in-Training and one county agent who participated in the February, 1986 assessment center. The other 12 participants were county agents who participated in one of two assessment centers held in the spring of 1986. Those who participated in the April assessment center were in the "moderate" group, and those who came to the May assessment center were in the "high" group. The agents were
given a choice of which assessment center they would attend, and came on the date that was most convenient for them.

About a month before the assessment center, assesses in the "low" involvement group were sent the list of assessment center dimensions and were asked to rate their ability on each dimension on a five-point scale (1=low, 5=high). Immediately following the assessment center, they were asked to rate their ability on each dimension once again, this time taking into consideration their performance in the assessment center exercises. Since the assessors were also using performance in the assessment center exercises to infer ability, it was appropriate for the assesses to evaluate themselves in this way. As soon as possible after the feedback interview, which was four to eight weeks after the assessment center, they were sent a questionnaire in which they were asked to rate their satisfaction with the assessment center, how much they felt they learned about themselves during the assessment center, their beliefs about whether the assessment center exercises were an adequate representation of their abilities, and their intentions to engage in developmental activities. There were also a series of questions as to what they had been told about the assessment center prior to attending, and how they felt about the assessment center while
participating; these questions acted as a manipulation check (see Appendix C).

The "moderate" involvement group also was sent the list of assessment center dimensions and were asked to rate themselves on each dimension and bring their ratings with them to the assessment center. Right before the start of the assessment center, this group was encouraged to observe and assess themselves during the assessment center. The assesses were told the following:

"Let me explain to you why we asked you to rate yourselves. These are the sixteen dimensions which were determined to be important to the job of County Chair. What the assessors are going to be doing is observing your behavior during each of the exercises. This behavior can be what you say, what you write, the decisions you make, etc. Then they will use these examples of behavior to make inferences about your ability. Well, we'd like you to be doing essentially the same thing. We've given you the list of dimensions and definitions for each dimension. These are the same definitions the assessors will be using. So, while you're participating in each of the exercises, we'd like you to be analyzing yourselves, focusing on your performance in the exercises. Then, once you've finished all the exercises we'll ask you to rate yourselves once again. So the key here is to use your behavior at the assessment center to make your self-ratings because that's what the assessors will be using. They don't know how you act in the office, or what other people always say about you. The only information they will have is what they are going to be seeing in the next two days.

They were then given a questionnaire which asked how involved they feel in the assessment center process (see Appendix D). They were told the assessment center staff was interested in knowing what people had heard about the assessment center or what their expectations were of the
assessment center. This questionnaire acted as a check on the level of involvement manipulation. The assessee then participated in the assessment center. Immediately following the assessment center, they were asked to rate themselves once again and were given the first questionnaire. Four to eight weeks later, they were sent the second questionnaire. These two questionnaires gave information both about how the assessee felt immediately after the assessment center, and how they felt about the assessment center after having had the feedback interview.

The "high" involvement group was sent the list of dimensions and asked to rate themselves on each dimension and send their self-ratings back to the assessment center coordinator. All of the group members complied with the request. Immediately prior to the start of the assessment center, these assessee participated in an hour-long training session. This training session is described in the next section. Following the training session, they were given the manipulation check questionnaire. The assessee then participated in the assessment center. Immediately following the assessment center, they were asked to rate themselves on the 16 assessment center dimensions once again, and were given the first evaluation questionnaire. Four to eight weeks after
the assessment center, they were sent the second questionnaire.

A possible effect of the training of the "high" involvement group was that they would be treated differently by the assessors, who may have felt they had been told how to act in each of the exercises. In order to minimize this possibility, assessors were not told which assesseees were in what group, and were not told that a training session had taken place.

In addition, as will be described next, the training focused on self-assessment, not on the assessment center exercises.

Self-Assessment Training

I. Start of Training

The training session was held at 9:00 and the assessment center started at 10:00.

II. Introduction

"The primary purpose for this assessment center is for you to gain a better understanding of yourself and your own abilities. One way this will be accomplished is by having the assessors evaluate you and then give you feedback on how
they feel you performed. However, it's also important for you to evaluate yourselves. Since you're the one who is going to seek out ways to develop yourselves, you have to see where development is needed. For example, the assessor can tell that you should work on becoming a little more assertive. But that's probably not going to mean too much to you until you see for yourself that there are certain situations where you don't speak up for yourself or express your opinions as forcefully as you should. Once you have this knowledge about yourself you can start to figure out ways of dealing with it."

"The overall point I'm trying to make is that we don't want you to sit back and say 'okay, assess me. Tell me what I need to know.' We want you to take as active a role as possible. Ideally, you would be in full control of the assessment center. You could decide which exercises you wanted to do and when, and you could redo some and change others; whatever would help you the most in analyzing your own abilities. Unfortunately, we have limited time and resources so the structure of the assessment center is essentially set. But within that structure we'd like you to be examining and analyzing yourself as much as possible.

III. Reason for Using Assessment Centers

A. History of assessment centers
   1. German military assessment program
2. Office of Strategic Services
3. AT&T Management Progress study

B. Improvement on supervisory performance appraisals
   1. assessors are trained in assessment
   2. procedures are standardized
   3. explicit feedback is given

IV. How We Develop Our Self-Assessments

A) Direct experience of the effects produced by our actions

B) Learn by vicarious experience, observing the actions of others

C) Learn from the judgments that others express about us - tells us how our behavior is viewed by the world

D) Use our powers of reasonings to derive new knowledge that goes beyond our direct experiences

V. Self-Assessments vs. Ratings of Others

A) Self-ratings are higher than ratings by others
   1. individuals need to maintain self-esteem
   2. may have different ideas of what is good and poor performance

B) Insider vs. Outsider
   We have a great deal of internal information which is available only to us

C) Intimate vs. Stranger
   We have knowledge about our past behavior which guides our attributions

D) Self vs. Other
   Occurs when individuals may distort their self-attributions to maintain their self-esteem

E) Actor vs. Observer
   Actor's attention is focused outward toward situational cues while observer's attention is focused on the actor
VI. Discussion of Dimensions

"The way the assessors will be evaluating you is by using your behavior in the various exercises to make inferences about your abilities. They were trained to focus on behavior. What we'd like you to do is to also evaluate yourself using your behavior in the exercises. So what I'd like to do now is to have a group discussion where we can come up with some ideas of what behaviors would indicate ability on each of the dimensions."

"Since some of the dimensions are more concrete than others, let's start with the easier ones."

The discussion of dimensions proceeded in the following order:

1. Oral communication
2. Written communication
3. Sensitivity
4. Initiative
5. Leadership/Persuasiveness
6. Objectivity
7. Assertiveness
8. Perception
10. Planning/Organizing
11. Behavioral flexibility/adaptability
12. Development of coworkers
13. Collaborativeness
14. Organizational sensitivity
15. Management control
16. Evaluation

For each dimension, the asessees came up with a few examples of behaviors which they felt would show a person was skilled in that dimension.

The Assessment Center

The asessees participated in seven exercises over a two-day period. The exercises were an In-Basket, two Leaderless Group Discussions (one with assigned roles and one with non-assigned roles), a Case Study analysis, a Background Interview (where the assessors interview the
assessees), a Fact-Finding exercise, and an Interview Simulation.

Sixteen dimensions were assessed in the assessment center. These dimensions were developed based on a job analysis of the County Chairperson's job. A committee of 21 people who were very familiar with the job of Chairperson assisted with the job analysis. For each of the chosen dimensions a definition was developed by the job analysis committee with the assistance of an assessment center consultant.

The dimensions on which the subjects were assessed and the definitions which were used by the assessors are:

1. Oral communication - The extent to which one can give an oral presentation and communicate on a one-to-one basis by listening and responding

2. Leadership/Delegation/Persuasiveness - The ability to influence others to move toward the attainment of a specific goal as efficiently as possible

3. Sensitivity - The ability to respond/react to a problem considering the feelings, emotions, and needs of others

4. Initiative - The capacity to see courses of action and the ability to begin actions without stimulation and support from others

5. Planning/Organizing - The process of establishing a course of action for self and/or others to accomplish a specific goal

6. Development of coworkers - The extent to which one develops and/or assists in developing the skills and competencies of coworkers through training and development activities, counseling, and delegating the duties related to current and future jobs
7. Judgment/Decision-making/Decisiveness - The process of identifying problems, securing relevant information, developing alternative courses of action, and the readiness of making a decision from the information gathered.

8. Behavioral flexibility/Adaptability - The extent to which one's behavior is flexible, adaptable, and effective when confronted with different situations, circumstances, or personalities.

9. Assertiveness - The degree to which one can effectively state his/her positively and forcefully without being hostile or destructive.

10. Organizational sensitivity - The degree of knowledge or awareness one has of formal and informal organizational policies and procedures.

11. Objectivity - The extent to which one can analyze, judge and make a fair decision about a person or situation regardless of one's own attitudes or feelings.

12. Written communication - The extent to which one can express effectively his/her ideas in writing.

13. Perception - The ability to identify or recognize a problem or potential problem.

14. Management control - The extent to which one makes the most efficient use of all resources (personnel, office, committee, etc.) to obtain effective outcomes.

15. Collaborativeness - The degree to which one is willing to work cooperatively with others in making decisions.

16. Evaluation - Assesses and appraises proposal of reported or observed performance; conducts performance appraisal; judges outcomes of programs; judges individual proposals and suggestions.

Some of these dimensions were seen as slightly more important to the job of County Chairperson than other dimensions. The highest priority dimensions were oral communication, leadership, planning/organizing, and decision...
making/judgement. The lowest priority dimensions were behavioral flexibility, assertiveness, organizational sensitivity, and written communication.

All assessors participated in a four-and-a-half day assessor training program led by an experienced assessment center consultant. During the training sessions, each of the seven exercises were described to the assessors and they observed role players performing each of the exercises. After each exercise, the assessors discussed how they felt the role-players should be evaluated. The consultant who was leading the training program facilitated this discussion. The assessors also practiced writing reports, in which they described the role players behavior in one of the exercises. These written reports are necessary since, due to the logistics of getting all the assesseees through all seven exercises in a two-day period, not every assessor has the opportunity to watch the assessee perform in every exercise. The written reports are not intended to be evaluative. Their intent is to inform assessors how the assessee behaved in any exercise they were not able to observe personally.

Throughout the assessor training an emphasis was placed on observation of behavior, and drawing appropriate inferences from these observations about a person's ability levels on each of the dimensions.
No evaluations were made until the assesses had completed all seven of the exercises. Once the exercises were finished, the three assessors who had been assigned to observe and evaluate each assessee independently rated the assessee on each of the sixteen dimensions on a five-point scale (1=low, 5=high). The three assessors then met together to discuss their ratings. Where there was disagreement among the assessors, they continued to discuss the dimension until they had reached agreement. Once the assessors had reached consensus on all dimensions, they formed a rating of the assessee's overall ability:

1=Should exceed normal management expectations
2=Should meet normal management expectations
3=Could meet normal management expectations with development
4=Meeting normal management expectations is questionable

All assessors were given instructions by the assessment center coordinator on how to conduct the feedback interview. The lead assessor was to contact the assessee and set a time and place for the feedback interview. The feedback interview was to be scheduled as soon as feasible following the assessment center, and should last about an hour to an hour and a half. During the feedback interview, the assessor was to review each of the seven assessment center exercises with the assessee. The assessee was to be told
what were the good and bad points about his/her performance in each exercise, and he/she was also to be given a written report which covered the same information. The assessee was not to be told what rating he/she received on any of the dimensions. Since the feedback interviews were confidential, it is not known to what extent the assessors actually followed this procedure. Within a few weeks after the feedback interview, the assessee was to be sent a staff development letter which offered suggestions for courses, workshops, and readings which would assist them in identifying appropriate developmental activities. Since this letter was sent out after the evaluation questionnaire used in this study, it did not influence the way participants responded to the questionnaire.

This assessment center was originated for developmental purposes, and all assessees were told that that was the main purpose, although in some instances later on it might be used as one part of the promotional process. The assessment center was fairly long (two days), had a large number of dimensions (16), and was followed by a feedback interview. Therefore, this assessment center had the characteristics which typify developmental assessment centers.
Measures

Manipulation check

Two questionnaires were used to assess the effectiveness of the level of involvement manipulation. The first questionnaire was given immediately prior to the start of the assessment center exercises. The first five questions were rated on a 1 to 5 scale (strongly agree to strongly disagree) and were designed to check on how the assessee perceived the information they were given on self-assessment. There were also nine semantic differentials which asked about their expectations of the assessment center. Two of the questions (involved vs. uninvolved, self-directed vs. directed by others) should discriminate between the three groups, while the other seven should not discriminate between the three groups. The second questionnaire was sent out four to eight weeks after the assessment center. The questions were the same, but the assesses were asked to recall how they felt during the assessment center.

Satisfaction, learning

These two variables were assessed with two questionnaires, one given immediately after completion of the assessment center exercises, and one sent out several weeks after the assessment center. There were five questions which asked about the assesses satisfaction with
and enjoyment of the assessment center and their satisfactions with their performance in the assessment center exercises. There were three questions which asked how much the assesses learned about themselves and their abilities during the assessment center. All the questions were asked on both questionnaires.

Perceived validity

Perceived validity had to do with whether the assesses felt the assessment center was an appropriate way to assess a person's skills and abilities. On the questionnaire given just after the assessment center, there were four questions designed to assess perceived validity. These four questions were repeated on the second questionnaire; in addition, there were four questions which asked about the feedback interview and the perceived accuracy of the assessor ratings.

Intentions to engage in developmental activities

This variable was assessed on the questionnaire sent out after the assessment center. Assesees were asked how many developmental activities they plan to engage in during the next 12 months, and were asked to list the topic areas of these activities.
Analysis Plan

Due to the small sample size, non-parametric statistics are used, where appropriate. Non-parametric procedures rank each response and then test whether the distribution of rank scores is equivalent across groups. To test for differences between two groups, the Wilcoxon rank sum test is used; the Kruskal-Wallis test will be used where there are more than two groups (Marascuilo & McSweeney, 1977).

A small sample size does not rule out the use of parametric statistics. It does mean that significance will be more difficult to obtain, since there will have to be a larger difference between groups for the difference to be considered significant. So t-tests will be used to test for differences between pretest and posttest ratings, and between self-ratings and assessor ratings.
CHAPTER III
RESULTS

All assessees and assessors returned all the materials they were asked to fill out, for a response rate of 100%. All questionnaires were complete; there were no missing data.

Manipulation Check

Level of involvement in self-assessment was manipulated. Assessees were assigned to either the Low, Moderate, or High involvement groups.

The check on the effectiveness of the level of involvement manipulation occurred at two points in time. The Moderate and High groups received a questionnaire immediately prior to the start of the assessment center; all three groups received a questionnaire several weeks after the assessment center. The lower the score received, the higher the level of involvement felt by the respondent.

There were two types of manipulation check questions. There were five Likert-scaled items which were designed to check on whether they had understood what they were told in the self-assessment training session and/or whether they had
paid attention when they were encouraged to self-assess. The other two questions were semantic differentials and asked about feelings of involvement.

A Wilcoxon rank sum test was used to test for differences between the Moderate and High groups on the first manipulation check questionnaire. For the Likert-scaled questions, the Moderate group had a mean rank of 9.42 (M=16.67, S.D.=1.21), and the High group had a mean rank of 3.58 (M=12.33, S.D.=3.33). This difference is significant (Z=2.74, p=.006) and in the expected direction. For the semantic differential items, the Moderate group had a mean rank of 5.67 (M=4.5, S.D.=1.64) and the High group had a mean rank of 7.33 (M=5.5, S.D.=.84). This difference is not significant (Z=.74, p=.46).

Since the second manipulation check included three groups, a Kruskal-Wallis test was used. For the Likert-scaled items, the Low group had a mean rank of 14.08 (M=15.0, S.D.=2.19), the Moderate group had a mean rank of 9.92 (M=14.82, S.D.=3.37), and the High group had a mean rank of 4.50 (M=14.66, S.D.=4.27). This difference is significant (X =9.77, p=.007) and in the expected direction. For the semantic differential items, the Low group had a mean rank of 9.50 (M=4.33, S.D.=.82), the Moderate group had a mean rank of 7.67 (M=4.00, S.D.=.89), and the High group
had a mean rank of 11.33 (M=4.66, S.D.=1.03). This difference is not significant (X^2=1.60, p=.45).

Each manipulation check questionnaire also contained seven semantic differential items which asked about feelings or perceptions which were not associated with level of involvement in self-assessment. It was hypothesized that responses to these questions would not differ between groups.

For the first questionnaire, the mean rank for the Moderate group on these items was 6.25 (M=17.33, S.D.=1.75), and for the High group the mean rank was 6.75 (M=17.50, S.D.=4.46). This difference is not significant (Z=.16, p=.87). On the second questionnaire, the mean rank for the Low group was 6.75 (M=14.50, S.D.=1.75), for the Moderate group it was 9.80 (M=16.83, S.D.=2.71), and for the High group it was 10.58 (M=17.33, S.D.=2.42). This difference also is not significant (X^2=1.93, p=.38).

Self-Assessment - 16 Dimensions

There were 16 dimensions used in the assessment center. The assesses rated their perceived abilities on each dimension on a five-point scale (1=poor, 5=excellent). They rated themselves once before the assessment center and again immediately after the assessment center. The vast majority of the ratings were either "average", "very good", or
"excellent" (3, 4, or 5). For the pretest ratings, only 5% of the ratings were below "average" (see Table 1), and for the posttest, only 2% of the ratings were below "average" (see Table 2).

Table 1

Pretest Self-Ratings - All 16 Dimensions

<table>
<thead>
<tr>
<th>Rating</th>
<th>N</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poor)</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2 (fair)</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>3 (average)</td>
<td>82</td>
<td>29</td>
</tr>
<tr>
<td>4 (very good)</td>
<td>136</td>
<td>47</td>
</tr>
<tr>
<td>5 (excellent)</td>
<td>55</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>288</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2

Posttest Self-Ratings - All 16 Dimensions

<table>
<thead>
<tr>
<th>Rating</th>
<th>N</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poor)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 (fair)</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>3 (average)</td>
<td>71</td>
<td>25</td>
</tr>
<tr>
<td>4 (very good)</td>
<td>139</td>
<td>48</td>
</tr>
<tr>
<td>5 (excellent)</td>
<td>71</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>288</td>
<td>100%</td>
</tr>
</tbody>
</table>

There was significant variation in assesses' evaluation of themselves across the 16 dimensions. On the pretest, the average ratings ranged from a low of 3.11 for Evaluation to a high of 4.39 for Initiative. A t-test of the difference between these two means revealed that the
difference is significant \((t=3.86, \ p<.002)\). For the posttest, average ratings ranged from 3.50 for Written Communication to 4.33 for Sensitivity. This difference is also significant \((t=3.59, \ p<.003)\). The average ratings for each dimension on the pretest and posttest are shown in Table 3.
### Table 3

Mean Ratings by Dimension

#### Pretest Self-Ratings (N=18)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>3.83</td>
<td>.71</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Written Communication</td>
<td>3.94</td>
<td>.80</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.83</td>
<td>.71</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Initiative</td>
<td>4.39</td>
<td>.61</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Planning</td>
<td>4.28</td>
<td>.70</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>3.72</td>
<td>.70</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Development</td>
<td>3.33</td>
<td>.97</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Flexibility</td>
<td>3.89</td>
<td>.68</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>3.61</td>
<td>.78</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.61</td>
<td>.70</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Objectivity</td>
<td>3.67</td>
<td>.59</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Perception</td>
<td>4.06</td>
<td>.73</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.89</td>
<td>.83</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.17</td>
<td>.79</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>4.28</td>
<td>.83</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.11</td>
<td>1.23</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

#### Posttest Self-Ratings (N=18)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>3.83</td>
<td>.79</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Written Communication</td>
<td>3.50</td>
<td>.79</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.72</td>
<td>.75</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.83</td>
<td>.79</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Planning</td>
<td>4.28</td>
<td>.70</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>4.17</td>
<td>.62</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Development</td>
<td>3.78</td>
<td>.81</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4.17</td>
<td>.51</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>4.22</td>
<td>.88</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.83</td>
<td>.99</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Objectivity</td>
<td>4.00</td>
<td>.69</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Perception</td>
<td>4.17</td>
<td>.71</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>4.33</td>
<td>.49</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.72</td>
<td>.89</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>4.11</td>
<td>.83</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.56</td>
<td>.51</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Self-Assessment (Pretest)

If the three groups were equivalent prior to the manipulation there should have been no difference between the groups in the way they rated themselves on the pretest. A total self-rating score for each assessee was formed by summing the ratings on each of the 16 dimensions. Each total score was then ranked. As it turned out, the Low group rated themselves higher than the other two groups. For the Low group, the mean rank was 13.75, for the Moderate group the mean rank was 8.75, and for the High group the mean rank was 6.0. A Kruskal-Wallis test, with average scores used for ties, was significant ($\chi^2=6.52$, $p=.04$). In addition to looking at the overall ratings, each of the 16 dimensions was also tested for significance, using the Kruskal-Wallis test (see Table 4). For 11 of the 16 dimensions, the Low group rated themselves higher than the Moderate and High group, and for three of these dimensions (Written communication, Planning, and Management control), the difference was significant.
Table 4
Mean Ratings by Group
Pretest Self-Ratings

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>S.D.</th>
<th>MOD</th>
<th>S.D.</th>
<th>HIGH</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>4.00</td>
<td>.63</td>
<td>3.83</td>
<td>.98</td>
<td>3.67</td>
<td>.52</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>4.50</td>
<td>.55</td>
<td>4.17</td>
<td>.75</td>
<td>3.17</td>
<td>.41</td>
</tr>
<tr>
<td>Leadership</td>
<td>4.00</td>
<td>.63</td>
<td>4.17</td>
<td>.41</td>
<td>3.33</td>
<td>.82</td>
</tr>
<tr>
<td>Initiative</td>
<td>4.50</td>
<td>.55</td>
<td>4.17</td>
<td>.75</td>
<td>4.50</td>
<td>.55</td>
</tr>
<tr>
<td>Planning</td>
<td>4.83</td>
<td>.41</td>
<td>3.83</td>
<td>.75</td>
<td>4.17</td>
<td>.41</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>4.00</td>
<td>.63</td>
<td>3.83</td>
<td>.75</td>
<td>3.33</td>
<td>.52</td>
</tr>
<tr>
<td>Development</td>
<td>3.67</td>
<td>.52</td>
<td>3.00</td>
<td>1.41</td>
<td>3.33</td>
<td>.82</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4.00</td>
<td>.63</td>
<td>4.17</td>
<td>.75</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>4.00</td>
<td>.63</td>
<td>3.33</td>
<td>1.03</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>4.00</td>
<td>.63</td>
<td>3.67</td>
<td>.82</td>
<td>3.17</td>
<td>.41</td>
</tr>
<tr>
<td>Objectivity</td>
<td>3.67</td>
<td>.52</td>
<td>3.83</td>
<td>.75</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Perception</td>
<td>4.16</td>
<td>.41</td>
<td>4.17</td>
<td>.75</td>
<td>3.83</td>
<td>.98</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>4.16</td>
<td>.75</td>
<td>4.00</td>
<td>.89</td>
<td>3.50</td>
<td>.84</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.83</td>
<td>.75</td>
<td>3.00</td>
<td>.63</td>
<td>2.67</td>
<td>.52</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>4.67</td>
<td>.52</td>
<td>4.33</td>
<td>.52</td>
<td>3.83</td>
<td>1.17</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.83</td>
<td>.75</td>
<td>3.00</td>
<td>1.41</td>
<td>2.50</td>
<td>1.22</td>
</tr>
</tbody>
</table>
Table 4, cont.

Mean Rankings by Group

Pretest Self-Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Low</th>
<th>Mod</th>
<th>High</th>
<th>X</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>10.75</td>
<td>9.25</td>
<td>8.50</td>
<td>.66</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>13.00</td>
<td>10.92</td>
<td>4.58</td>
<td>9.14</td>
<td>.01</td>
</tr>
<tr>
<td>Initiative</td>
<td>10.25</td>
<td>8.00</td>
<td>10.25</td>
<td>.90</td>
<td>n.s.</td>
</tr>
<tr>
<td>Planning</td>
<td>13.67</td>
<td>6.50</td>
<td>8.33</td>
<td>7.14</td>
<td>.03</td>
</tr>
<tr>
<td>Development</td>
<td>11.17</td>
<td>8.00</td>
<td>9.33</td>
<td>1.21</td>
<td>n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>10.33</td>
<td>11.42</td>
<td>6.75</td>
<td>3.12</td>
<td>n.s.</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>12.08</td>
<td>7.67</td>
<td>8.75</td>
<td>2.61</td>
<td>n.s.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>12.42</td>
<td>9.75</td>
<td>6.33</td>
<td>4.79</td>
<td>n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>9.67</td>
<td>10.58</td>
<td>8.25</td>
<td>.75</td>
<td>n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>9.83</td>
<td>10.08</td>
<td>8.58</td>
<td>.39</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>11.08</td>
<td>10.00</td>
<td>7.42</td>
<td>1.75</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>13.58</td>
<td>8.58</td>
<td>6.33</td>
<td>7.11</td>
<td>.03</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>11.83</td>
<td>9.17</td>
<td>7.50</td>
<td>2.43</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>12.58</td>
<td>8.92</td>
<td>7.00</td>
<td>3.70</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Self-Assessment (Posttest)

The findings for the self-ratings done after the assessment center were essentially the same as for the ratings done before the assessment center. As with the pretest, the Low group rated themselves higher than the Moderate and High groups. The mean rank for the Low group was 12.92, for the Moderate group the mean rank was 10.42, and the High group the mean rank was 5.17 (X =6.60, p=.04). The mean rank for each of the 16 dimensions is shown in Table 5. For 12 of the dimensions, the Low group rated themselves higher than the other two groups. For two of the
dimensions (Management control and Evaluation) the difference was significant. For Perception, the Moderate group gave themselves the highest ratings, and this difference was significant.

Table 5
Mean Ratings by Group
Posttest Self-Ratings

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>S.D.</th>
<th>MOD</th>
<th>S.D.</th>
<th>HIGH</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>4.00</td>
<td>.63</td>
<td>4.00</td>
<td>.89</td>
<td>3.50</td>
<td>.84</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>3.83</td>
<td>.75</td>
<td>3.67</td>
<td>.82</td>
<td>3.00</td>
<td>.63</td>
</tr>
<tr>
<td>Leadership</td>
<td>4.00</td>
<td>.63</td>
<td>3.67</td>
<td>.82</td>
<td>3.50</td>
<td>.84</td>
</tr>
<tr>
<td>Initiative</td>
<td>4.33</td>
<td>.82</td>
<td>3.67</td>
<td>.82</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Planning</td>
<td>4.50</td>
<td>.55</td>
<td>4.33</td>
<td>.82</td>
<td>4.00</td>
<td>.63</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>4.33</td>
<td>.52</td>
<td>4.33</td>
<td>.52</td>
<td>3.83</td>
<td>.75</td>
</tr>
<tr>
<td>Development</td>
<td>4.17</td>
<td>.75</td>
<td>3.83</td>
<td>.75</td>
<td>3.33</td>
<td>.82</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4.33</td>
<td>.52</td>
<td>4.17</td>
<td>.41</td>
<td>4.00</td>
<td>.63</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>4.50</td>
<td>.84</td>
<td>4.17</td>
<td>.98</td>
<td>4.00</td>
<td>.89</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>4.17</td>
<td>.75</td>
<td>4.17</td>
<td>.75</td>
<td>3.17</td>
<td>1.17</td>
</tr>
<tr>
<td>Objectivity</td>
<td>4.33</td>
<td>.52</td>
<td>4.17</td>
<td>.75</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Perception</td>
<td>4.33</td>
<td>.52</td>
<td>4.67</td>
<td>.52</td>
<td>3.50</td>
<td>.55</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>4.33</td>
<td>.52</td>
<td>4.33</td>
<td>.52</td>
<td>4.33</td>
<td>.52</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>4.50</td>
<td>.55</td>
<td>3.67</td>
<td>.52</td>
<td>3.00</td>
<td>.89</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>4.50</td>
<td>.55</td>
<td>4.33</td>
<td>.52</td>
<td>3.50</td>
<td>1.05</td>
</tr>
<tr>
<td>Evaluation</td>
<td>4.50</td>
<td>0.0</td>
<td>3.33</td>
<td>.52</td>
<td>3.33</td>
<td>.52</td>
</tr>
</tbody>
</table>
Table 5, cont.

Mean Rankings by Group

Posttest Self-Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Low</th>
<th>Mod</th>
<th>High</th>
<th>X^2</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>10.75</td>
<td>10.50</td>
<td>7.25</td>
<td>1.84</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>11.67</td>
<td>10.42</td>
<td>6.42</td>
<td>3.77</td>
<td>n.s.</td>
</tr>
<tr>
<td>Leadership</td>
<td>11.58</td>
<td>9.08</td>
<td>7.83</td>
<td>1.80</td>
<td>n.s.</td>
</tr>
<tr>
<td>Initiative</td>
<td>12.58</td>
<td>8.42</td>
<td>7.50</td>
<td>3.54</td>
<td>n.s.</td>
</tr>
<tr>
<td>Planning</td>
<td>11.00</td>
<td>10.08</td>
<td>7.42</td>
<td>1.79</td>
<td>n.s.</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>10.67</td>
<td>10.67</td>
<td>7.17</td>
<td>2.29</td>
<td>n.s.</td>
</tr>
<tr>
<td>Development</td>
<td>11.83</td>
<td>9.67</td>
<td>7.00</td>
<td>2.90</td>
<td>n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>10.83</td>
<td>9.42</td>
<td>8.25</td>
<td>1.15</td>
<td>n.s.</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>11.08</td>
<td>9.25</td>
<td>8.17</td>
<td>1.08</td>
<td>n.s.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>11.08</td>
<td>11.08</td>
<td>6.33</td>
<td>3.48</td>
<td>n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>11.83</td>
<td>10.67</td>
<td>6.00</td>
<td>4.96</td>
<td>n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>10.50</td>
<td>13.00</td>
<td>5.00</td>
<td>8.44</td>
<td>.01</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>9.50</td>
<td>9.50</td>
<td>9.50</td>
<td>.00</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>14.00</td>
<td>8.83</td>
<td>5.67</td>
<td>8.66</td>
<td>.01</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>11.75</td>
<td>10.50</td>
<td>6.25</td>
<td>4.17</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>13.50</td>
<td>7.50</td>
<td>7.50</td>
<td>6.80</td>
<td>.03</td>
</tr>
</tbody>
</table>

Change in Self-Ratings

Since each assessee rated themselves both before and after the assessment center, it was possible to compare pretest and posttest ratings to determine whether assesses' perceptions of their own abilities changed as a result of participating in the assessment center. A t-test was used to evaluate whether the difference between pretest and posttest was significantly different from zero. The overall difference was not significant.
In addition to looking at the overall ratings, differences between pretest and posttest ratings on each of the 16 dimensions were also examined. For six of the dimensions, the difference was significantly different from zero (see Table 6). For Written communication and Initiative, pretest scores were significantly higher than posttest scores. For Decision-making, Organizational sensitivity, Sensitivity, and Management control, the posttest ratings were significantly higher than the pretest.

Although there was not a significant change in self-ratings from pretest to posttest for the assessees as a whole, this overall finding might have masked changes that occurred for each group. To determine if this was the case, pretest and posttest rating were compared for each of the three groups. These findings were very similar to the results for all of the groups combined, indicating that the group the assessees was in did not affect whether the assessees' self-ratings changed from pretest to posttest.
Table 6
Differences Between Pretest and Posttest Ratings (N=18)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>0.0</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Communication</td>
<td>2.68</td>
<td>.02</td>
</tr>
<tr>
<td>Leadership</td>
<td>.52</td>
<td>n.s.</td>
</tr>
<tr>
<td>Initiative</td>
<td>2.75</td>
<td>.01</td>
</tr>
<tr>
<td>Planning</td>
<td>0.0</td>
<td>n.s.</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>-3.69</td>
<td>.002</td>
</tr>
<tr>
<td>Development</td>
<td>-2.05</td>
<td>n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>-1.57</td>
<td>n.s.</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>-3.05</td>
<td>.007</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>-1.07</td>
<td>n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>-2.06</td>
<td>n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>-.52</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>-2.41</td>
<td>.03</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>-3.01</td>
<td>.008</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>.82</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>-1.57</td>
<td>n.s.</td>
</tr>
<tr>
<td>Overall</td>
<td>-1.71</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Assessor Ratings - 16 Dimensions

The assesses rated themselves on each of the 16 dimensions on a 5-point scale. The assessors used the same scale to make their ratings. As with the self-ratings, there were very few "poor" ratings. However, the assessors gave many more "average" ratings than did the assesses, and far fewer "excellent" ratings (see Table 7). Overall, the assessors' ratings were much closer to the middle of the scale than were the assesses' self-ratings.
Table 7

Assessor Ratings - All 16 Dimensions

<table>
<thead>
<tr>
<th>Rating</th>
<th>N</th>
<th>% of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (poor)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2 (fair)</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>3 (average)</td>
<td>120</td>
<td>42</td>
</tr>
<tr>
<td>4 (very good)</td>
<td>127</td>
<td>44</td>
</tr>
<tr>
<td>5 (excellent)</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>288</td>
<td>100%</td>
</tr>
</tbody>
</table>

Across the 16 dimensions, there was significant variation in the average ratings given by the assessors. The ratings ranged from a low of 3.11 for Development of coworkers to a high of 3.89 for Planning. This difference is significant ($t=2.61$, $p<.02$). The average ratings for each dimension are shown in Table 8.
### Table 8
Mean Ratings by Dimension

**Assessor Ratings (N=18)**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>3.61</td>
<td>.69</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Written Communication</td>
<td>3.44</td>
<td>.86</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.28</td>
<td>.67</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.28</td>
<td>.67</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Planning</td>
<td>3.89</td>
<td>.58</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>3.39</td>
<td>.61</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Development</td>
<td>3.11</td>
<td>1.02</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Flexibility</td>
<td>3.17</td>
<td>.71</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>3.67</td>
<td>.77</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.39</td>
<td>.78</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Objectivity</td>
<td>3.50</td>
<td>.62</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Perception</td>
<td>3.39</td>
<td>.61</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.50</td>
<td>.86</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.33</td>
<td>.69</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>3.67</td>
<td>.77</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.22</td>
<td>.81</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

**Overall Rating**

In addition to rating the assesses on each dimension, the assessors also gave each assessee an overall assessment center rating using the following scale:

1=should exceed normal mgmt. expectations
2=should meet normal mgmt. expectations
3=could meet normal mgmt. expectations with development
4=meeting normal mgmt. expectations is questionable

Out of eighteen assesses, two of them received an overall rating of "1", eight received an overall rating of "2", and eight were given an overall rating of "3". None of the assesses received a "4" rating.
The distribution by group is as follows: For the Low group, one assessee received an overall rating of "1", three received a "2", and the other three were given a "3". For the Moderate group, three assessees received a "2", and the other three received a "3". For the High group, one assessee was given a "1", two assessees received a "2", and three assessees received a "3".

Assessor Ratings - By Group

Although the assessor ratings varied from dimension to dimension, they did not vary from group to group. A total score for each assessee was calculated by summing the ratings given by the assessors on each dimension. Each total score was then ranked. For the Low group, the average rank was 10.08, for the Moderate group the average rank was 9.50, and for the High group the average rank was 8.92. A Kruskal-Wallis test showed these differences to be non-significant \( \chi^2 = .14, p = .93 \).

Since the assessors gave each assessee an overall assessment center rating, it also was possible to compare these ratings by group. For the Low group, the average overall rank was 8.33, for The Moderate group the average overall rank was 10.50, and for the High group it was 9.67. These differences were also non-significant \( \chi^2 = .61, p = .74 \).
When broken down by dimension, there were still no differences in the way the assessors rated the assesseees in different groups. As shown in Table 9, the only significant difference was for Planning, with the Low group receiving the highest ratings and the High group receiving the lowest rating. Given the large number of dimensions, this significant difference may be the result of chance variation.

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>MOD</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>3.67</td>
<td>.52</td>
<td>4.00</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>3.83</td>
<td>.41</td>
<td>3.33</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.33</td>
<td>.82</td>
<td>3.17</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.33</td>
<td>.52</td>
<td>3.50</td>
</tr>
<tr>
<td>Planning</td>
<td>4.33</td>
<td>.52</td>
<td>3.83</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>3.33</td>
<td>.82</td>
<td>3.33</td>
</tr>
<tr>
<td>Development</td>
<td>3.00</td>
<td>.63</td>
<td>3.33</td>
</tr>
<tr>
<td>Flexibility</td>
<td>2.83</td>
<td>.75</td>
<td>3.33</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>3.83</td>
<td>.98</td>
<td>3.17</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.50</td>
<td>.55</td>
<td>3.67</td>
</tr>
<tr>
<td>Objectivity</td>
<td>3.83</td>
<td>.41</td>
<td>3.33</td>
</tr>
<tr>
<td>Perception</td>
<td>3.50</td>
<td>.55</td>
<td>3.50</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.17</td>
<td>.75</td>
<td>3.83</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.33</td>
<td>.52</td>
<td>3.00</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>3.50</td>
<td>.55</td>
<td>3.67</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.50</td>
<td>.55</td>
<td>3.00</td>
</tr>
</tbody>
</table>
Table 9, cont.

Mean Rankings by Group

Assessor Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Low</th>
<th>Mod</th>
<th>High</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>9.33</td>
<td>12.00</td>
<td>7.17</td>
<td>3.99</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Commun.</td>
<td>11.67</td>
<td>9.67</td>
<td>7.17</td>
<td>2.84</td>
<td>n.s.</td>
</tr>
<tr>
<td>Leadership</td>
<td>9.25</td>
<td>8.83</td>
<td>10.42</td>
<td>.41</td>
<td>n.s.</td>
</tr>
<tr>
<td>Initiative</td>
<td>9.67</td>
<td>11.00</td>
<td>7.83</td>
<td>1.30</td>
<td>n.s.</td>
</tr>
<tr>
<td>Planning</td>
<td>12.83</td>
<td>9.17</td>
<td>6.50</td>
<td>6.14</td>
<td>.05</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>9.42</td>
<td>8.83</td>
<td>10.25</td>
<td>.27</td>
<td>n.s.</td>
</tr>
<tr>
<td>Development</td>
<td>9.00</td>
<td>10.50</td>
<td>9.00</td>
<td>.36</td>
<td>n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>7.25</td>
<td>10.75</td>
<td>10.50</td>
<td>1.92</td>
<td>n.s.</td>
</tr>
<tr>
<td>Org. Sensitivity</td>
<td>10.25</td>
<td>6.25</td>
<td>12.00</td>
<td>4.38</td>
<td>n.s.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>9.75</td>
<td>11.00</td>
<td>7.75</td>
<td>1.41</td>
<td>n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>12.08</td>
<td>8.58</td>
<td>7.83</td>
<td>2.81</td>
<td>n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>10.25</td>
<td>10.25</td>
<td>8.00</td>
<td>.90</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>7.58</td>
<td>11.42</td>
<td>9.50</td>
<td>1.75</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>9.67</td>
<td>7.73</td>
<td>11.50</td>
<td>2.44</td>
<td>n.s.</td>
</tr>
<tr>
<td>Collaborative.</td>
<td>8.25</td>
<td>9.83</td>
<td>10.42</td>
<td>.63</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>11.00</td>
<td>8.17</td>
<td>9.33</td>
<td>.98</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Comparison of Assessor and Self-Ratings

Since past research has shown that self-ratings and the ratings of others are often not congruent, assessor ratings were compared with both pretest and posttest self-ratings. T-tests were used to determine if the difference between ratings was significant. For the pretest ratings, there were significant differences between assessor and assessee ratings on seven of the sixteen dimensions. The overall rating, which was calculated by taking the sum of the ratings on the sixteen dimensions, also showed a significant
The differences between assessor ratings and posttest ratings are shown in Table 11. For ten of the sixteen dimensions, the posttest self-ratings were significantly higher than the assessor ratings. The overall difference was highly significant. Comparing the pretest and posttest ratings, it appears that the assessees were less accurate making their posttest ratings than their pretest ratings.
Table 11
Differences Between Assessor and Posttest Ratings (N=18)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Posttest Mean</th>
<th>Assessor Mean</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>3.83</td>
<td>3.61</td>
<td>1.00</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Communication</td>
<td>3.50</td>
<td>3.44</td>
<td>.29</td>
<td>n.s.</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.72</td>
<td>3.28</td>
<td>2.05</td>
<td>.05</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.83</td>
<td>3.28</td>
<td>3.01</td>
<td>.008</td>
</tr>
<tr>
<td>Planning</td>
<td>4.28</td>
<td>3.89</td>
<td>2.36</td>
<td>.03</td>
</tr>
<tr>
<td>Decision-Making</td>
<td>4.17</td>
<td>3.39</td>
<td>3.76</td>
<td>.002</td>
</tr>
<tr>
<td>Development</td>
<td>3.78</td>
<td>3.11</td>
<td>2.61</td>
<td>.02</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4.17</td>
<td>3.17</td>
<td>5.53</td>
<td>.0001</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.83</td>
<td>3.39</td>
<td>2.41</td>
<td>.03</td>
</tr>
<tr>
<td>Objectivity</td>
<td>4.00</td>
<td>3.50</td>
<td>3.43</td>
<td>.003</td>
</tr>
<tr>
<td>Perception</td>
<td>4.17</td>
<td>3.39</td>
<td>3.76</td>
<td>.002</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>4.33</td>
<td>3.50</td>
<td>3.59</td>
<td>.002</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.72</td>
<td>3.33</td>
<td>1.51</td>
<td>n.s.</td>
</tr>
<tr>
<td>Collaborativeness</td>
<td>4.11</td>
<td>3.67</td>
<td>1.72</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.56</td>
<td>3.22</td>
<td>1.68</td>
<td>n.s.</td>
</tr>
<tr>
<td>Total</td>
<td>63:22</td>
<td>54.84</td>
<td>3.99</td>
<td>.0009</td>
</tr>
</tbody>
</table>

Comparison of Assessor and Self-Ratings - By Group

The first hypothesis deals with accuracy of self-ratings. It was hypothesized that assessees in the High involvement condition would report self-ratings that were the closest to the assessor ratings while assessees in the Low involvement condition would report self-ratings that were the farthest from the assessor ratings.

To answer this question, the data was broken down by groups. As shown in Table 12, the groups were not equally accurate prior to the involvement in self-assessment manipulation. The Low group was much less accurate than the
Moderate and High groups. For the Low group, on six of the sixteen dimensions the assessees gave themselves significantly higher ratings than they received from the assessors. For the Moderate group, the assessees rated themselves significantly higher than did the assessors on one dimension (Leadership). For the High group, the assessees gave themselves higher ratings than they actually received from the assessors on two dimensions (Initiative and Planning).

Table 12
Differences Between Assessor Ratings and Pretest Ratings
By Group

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Low</th>
<th>Mod</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Comm.</td>
<td>1.00 n.s.</td>
<td>-.42 n.s.</td>
<td>1.00 n.s.</td>
</tr>
<tr>
<td>Written Comm.</td>
<td>2.00 n.s.</td>
<td>1.54 n.s.</td>
<td>0.0 n.s.</td>
</tr>
<tr>
<td>Leadership</td>
<td>3.16 .03</td>
<td>3.87 .01</td>
<td>0.0 n.s.</td>
</tr>
<tr>
<td>Initiative</td>
<td>7.00 .0009</td>
<td>2.00 n.s.</td>
<td>4.39 .007</td>
</tr>
<tr>
<td>Planning</td>
<td>2.24 n.s.</td>
<td>0.0 n.s.</td>
<td>3.16 .03</td>
</tr>
<tr>
<td>Decis.-Making</td>
<td>1.20 n.s.</td>
<td>1.46 n.s.</td>
<td>-.54 n.s.</td>
</tr>
<tr>
<td>Development</td>
<td>2.00 n.s.</td>
<td>-.36 n.s.</td>
<td>.67 n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>7.00 .0009</td>
<td>1.75 n.s.</td>
<td>.54 n.s.</td>
</tr>
<tr>
<td>Org. Sensit.</td>
<td>.54 n.s.</td>
<td>.35 n.s.</td>
<td>-1.46 n.s.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>2.24 n.s.</td>
<td>0.0 n.s.</td>
<td>.42 n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>-1.00 n.s.</td>
<td>1.00 n.s.</td>
<td>.54 n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>3.16 .03</td>
<td>1.35 n.s.</td>
<td>1.20 n.s.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.87 .01</td>
<td>.28 n.s.</td>
<td>0.0 n.s.</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>1.17 n.s.</td>
<td>0.0 n.s.</td>
<td>-1.94 n.s.</td>
</tr>
<tr>
<td>Collaborative.</td>
<td>3.80 .01</td>
<td>1.20 n.s.</td>
<td>0.0 n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>1.00 n.s.</td>
<td>0.0 n.s.</td>
<td>-.79 n.s.</td>
</tr>
<tr>
<td>Overall</td>
<td>4.56 .006</td>
<td>1.06 n.s.</td>
<td>.38 n.s.</td>
</tr>
</tbody>
</table>
The pattern changed somewhat for the posttest ratings (see Table 13). The Low group still gave themselves significantly higher ratings than the assessors on six of the dimensions. The difference in overall evaluation (sum of the 16 dimensions) was still significant, but not as significant as the pretest ratings. For the Moderate group, the assessees rated themselves significantly higher than the assessors on five of the dimensions. The overall difference, which was not significant on the pretest, was significant on the posttest. The High group rated themselves significantly higher then did the assessors on only one dimension (Sensitivity).
Table 13

Differences Between Assessor Ratings and Posttest Ratings

By Group

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Low</th>
<th>Low</th>
<th>Mod</th>
<th>Mod</th>
<th>High</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
<td>( p )</td>
<td>( t )</td>
<td>( p )</td>
<td>( t )</td>
<td>( p )</td>
</tr>
<tr>
<td>Oral Comm.</td>
<td>1.00</td>
<td>n.s.</td>
<td>0.0</td>
<td>n.s.</td>
<td>.67</td>
<td>n.s.</td>
</tr>
<tr>
<td>Written Comm.</td>
<td>0.0</td>
<td>n.s.</td>
<td>.67</td>
<td>n.s.</td>
<td>-1.00</td>
<td>n.s.</td>
</tr>
<tr>
<td>Leadership</td>
<td>1.58</td>
<td>n.s.</td>
<td>1.17</td>
<td>n.s.</td>
<td>.54</td>
<td>n.s.</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.87</td>
<td>.01</td>
<td>1.00</td>
<td>n.s.</td>
<td>1.17</td>
<td>n.s.</td>
</tr>
<tr>
<td>Planning</td>
<td>1.00</td>
<td>n.s.</td>
<td>1.46</td>
<td>n.s.</td>
<td>1.46</td>
<td>n.s.</td>
</tr>
<tr>
<td>Decis.-Making</td>
<td>2.24</td>
<td>n.s.</td>
<td>3.87</td>
<td>.01</td>
<td>1.00</td>
<td>n.s.</td>
</tr>
<tr>
<td>Development</td>
<td>3.80</td>
<td>n.s.</td>
<td>.89</td>
<td>n.s.</td>
<td>.79</td>
<td>n.s.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>6.71</td>
<td>.001</td>
<td>2.71</td>
<td>.04</td>
<td>2.00</td>
<td>n.s.</td>
</tr>
<tr>
<td>Org. Sensit.</td>
<td>2.00</td>
<td>n.s.</td>
<td>1.94</td>
<td>n.s.</td>
<td>0.0</td>
<td>n.s.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>3.16</td>
<td>.03</td>
<td>2.24</td>
<td>n.s.</td>
<td>.35</td>
<td>n.s.</td>
</tr>
<tr>
<td>Objectivity</td>
<td>2.24</td>
<td>n.s.</td>
<td>5.00</td>
<td>.004</td>
<td>.54</td>
<td>n.s.</td>
</tr>
<tr>
<td>Perception</td>
<td>2.08</td>
<td>n.s.</td>
<td>7.00</td>
<td>.0009</td>
<td>.79</td>
<td>n.s.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.80</td>
<td>.01</td>
<td>.89</td>
<td>n.s.</td>
<td>2.71</td>
<td>.04</td>
</tr>
<tr>
<td>Mgmt. Control</td>
<td>3.80</td>
<td>.01</td>
<td>3.16</td>
<td>.03</td>
<td>-1.58</td>
<td>n.s.</td>
</tr>
<tr>
<td>Collaborative.</td>
<td>2.74</td>
<td>.04</td>
<td>1.58</td>
<td>n.s.</td>
<td>-.79</td>
<td>n.s.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>2.24</td>
<td>n.s.</td>
<td>1.00</td>
<td>n.s.</td>
<td>.35</td>
<td>n.s.</td>
</tr>
<tr>
<td>Overall</td>
<td>3.96</td>
<td>.01</td>
<td>2.57</td>
<td>.05</td>
<td>.92</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Correlation of Assessor and Self-Ratings

Although assessees tended to give themselves higher ratings than they received from the assessors, the self-ratings and assessor ratings might still be highly correlated. However, this was not the case. For the pretest ratings, there were no significant correlations between self-ratings on the 16 dimensions and the corresponding assessor ratings. For the posttest ratings, three of the dimensions showed significant correlations between self-ratings and assessor ratings--Written
communication (r=.52, p=.025), Assertiveness (r=.63, p=.005), and Objectivity (r=.55, p=.017).

Correlations between overall self-ratings (sum of the 16 dimensions) and overall assessor ratings were also non-significant. Although the correlation between pretest self-ratings and posttest self-ratings was highly significant (r=.65, p=.0035), the correlation between pretest and assessor ratings (r=.01, p=.96), and between posttest and assessor ratings (r=.27, p=.28) were not significant.

Correlation of Assessor Ratings

The intercorrelation of the assessor ratings on all of the 16 dimensions are shown in Table 14. Of the 256 correlations, sixteen are significant at the .05 level, and nineteen are significant at the .01 level. The problem of halo error with assessor ratings and its impact on the feedback interview was addressed earlier.
### Table 14

Correlation of Assessor Ratings

<table>
<thead>
<tr>
<th>Writ</th>
<th>Lead</th>
<th>Init</th>
<th>Plan</th>
<th>Decis</th>
<th>Devel</th>
<th>Flex</th>
<th>Org</th>
<th>Sens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>.01</td>
<td>.49*</td>
<td>.62#</td>
<td>.18</td>
<td>.23</td>
<td>.47*</td>
<td>.26</td>
<td>.18</td>
</tr>
<tr>
<td>Writ</td>
<td>.08</td>
<td>.18</td>
<td>.46</td>
<td>.10</td>
<td>-.26</td>
<td>-.23</td>
<td>-.12</td>
<td></td>
</tr>
<tr>
<td>Lead</td>
<td>.61#</td>
<td>.08</td>
<td>.44</td>
<td>.55*</td>
<td>.39</td>
<td>.53*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Init</td>
<td>.54*</td>
<td>.44</td>
<td>.21</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan</td>
<td>.12</td>
<td>-.18</td>
<td>-.38</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decis</td>
<td>.21</td>
<td>.52*</td>
<td>.55#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>.70#</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assert</th>
<th>Object</th>
<th>Perc</th>
<th>Sens</th>
<th>Mgmt</th>
<th>Collab</th>
<th>Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>.73#</td>
<td>-.06</td>
<td>.52*</td>
<td>.25</td>
<td>.29</td>
<td>.29</td>
</tr>
<tr>
<td>Writ</td>
<td>.08</td>
<td>0.0</td>
<td>.16</td>
<td>-.16</td>
<td>-.12</td>
<td>-.12</td>
</tr>
<tr>
<td>Lead</td>
<td>.46</td>
<td>.21</td>
<td>.44</td>
<td>.36</td>
<td>.68#</td>
<td>.53*</td>
</tr>
<tr>
<td>Init</td>
<td>.68#</td>
<td>.07</td>
<td>.44</td>
<td>.15</td>
<td>.42</td>
<td>.31</td>
</tr>
<tr>
<td>Plan</td>
<td>.36</td>
<td>.33</td>
<td>.13</td>
<td>-.35</td>
<td>.09</td>
<td>.04</td>
</tr>
<tr>
<td>Decis</td>
<td>.16</td>
<td>.23</td>
<td>.36</td>
<td>.39</td>
<td>.51*</td>
<td>.55*</td>
</tr>
<tr>
<td>Devel</td>
<td>.09</td>
<td>.28</td>
<td>.59*</td>
<td>.74#</td>
<td>.44</td>
<td>.73#</td>
</tr>
<tr>
<td>Flex</td>
<td>-.12</td>
<td>.20</td>
<td>.39</td>
<td>.73</td>
<td>.36</td>
<td>.65#</td>
</tr>
<tr>
<td>Org Sens</td>
<td>.13</td>
<td>.25</td>
<td>.29</td>
<td>.18</td>
<td>.67#</td>
<td>.40</td>
</tr>
<tr>
<td>Assert</td>
<td>-.06</td>
<td>.40</td>
<td>-.04</td>
<td>.29</td>
<td>.13</td>
<td>.50*</td>
</tr>
<tr>
<td>Object</td>
<td>.54*</td>
<td>.27</td>
<td>.42</td>
<td>.49*</td>
<td>.59*</td>
<td></td>
</tr>
<tr>
<td>Percep</td>
<td>.62#</td>
<td>.52*</td>
<td>.67#</td>
<td>.65#</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensit</td>
<td></td>
<td>.20</td>
<td>.72#</td>
<td>.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mgmt</td>
<td></td>
<td>.67#</td>
<td>.71#</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.69#</td>
</tr>
</tbody>
</table>

* p < .05  
# p < .01

### Satisfaction

The second hypothesis dealt with the assesses' satisfaction with the assessment center process.

Immediately after the assessment center, the Moderate and High groups were given an evaluation questionnaire which asked about their satisfaction with the assessment center. Due to the small sample size, it was not possible to test
the items for internal reliability. A Wilcoxon two-sample test showed no difference between the Moderate and High groups on this variable (\(Z=.41, p=.69\)).

Several weeks after the assessment center, all three groups were once again asked about their satisfaction with the assessment center. A Kruskal-Wallis test showed the difference between these three groups to be non-significant \(\left(\chi^2=1.56, p=.46\right)\).

Lower scores indicate greater satisfaction. With a possible range of 4-20, the average for this study was 12.94 (s.d.=2.2).

Learning

The third hypothesis stated that assessees in the High involvement group would report that they learned more about themselves than the Moderate involvement group, who in turn would report greater learning than the Low involvement group.

On the first questionnaire, there was not a significant difference between the Moderate and High groups (\(Z=.25, p=.81\)). However, a significant difference emerged for the second questionnaire \(\left(\chi^2=6.10, p=.05\right)\). The mean rank for the Low group was 5.17, for the Moderate group the mean rank was 11.83, and for the High group the mean rank was 11.50.
Since these results were the opposite of what was hypothesized, the data were broken down further in an attempt to discover why these results occurred. There were three questions on the second questionnaire which asked about learning:

1. I feel I learned quite a bit about myself as a result of participating in the assessment center.

2. My perceptions of my abilities changed as a result of participating in the assessment center.

3. The assessment center gave me added information about what is expected of me on the job.

Each question was evaluated separately using the Kruskal-Wallis test. For the first question, the mean rank for the Low group was 8.25, for the Moderate group the mean rank was 9.92, and for the High group the mean rank was 10.33. This difference was not significant ($X^2=.53$, $p=.74$). For the second question, the Low group had a mean rank of 4.50, the Moderate group had a mean rank of 11.85, and the High group had a mean rank of 12.42. This difference was significant ($X^2=9.14$, $p=.01$). For the third question, the mean rank for the Low group was 6.42, for the Moderate group
the mean rank was 12.08, and for the High group the mean rank was 10.00 ($X^2=3.76$, $p=.15$).

With lower scores indicating greater learning, the possible range was 3-15. The average rating was 11.0 (s.d.=2.68).

**Perceived Validity**

The fourth hypothesis stated that assessees in the High involvement group would be more likely to feel that the assessment center was a valid way to assess abilities than assessees in the Low involvement group. Responses of assessees in the Moderate involvement group were expected to fall in between the other two groups.

There was not a significant difference between the Moderate and High groups in their perceptions of the validity of the assessment center process on the first questionnaire ($Z=.25$, $p=.81$). On the second questionnaire, all the items were repeated; in addition, there were several questions which asked about the feedback interview. However, since three of the assessees had not yet had their feedback interview, these items were deleted from the analysis. The difference between the three groups on the second questionnaire was not significant ($X^2=1.22$, $p=.54$).
The average rating was 14.17 (s.d.=2.57) with a possible range of 4-20. Lower scores indicate greater perceived validity.

**Intenstions to Engage in Developmental Activities**

It was hypothesized that assessees in the High involvement group would report more intentions to engage in developmental activities than assessees in the Moderate involvement group, who in turn would report more intentions to engage in developmental activities than assessees in the Low involvement group.

On the questionnaire sent out several weeks after the assessment center, assessees were asked to indicate the number of developmental activities they planned to engage in during the next 12 months. There was not a significant difference between the three groups ($X^2=2.43$, $p=.30$). It should be noted, however, that some of the assessees refused to respond to this question. One assessees said that he did not wish to commit himself to a developmental plan at this stage; another said she was waiting to receive suggestions on developmental activities from the assessment center staff. These assessees were given a zero for number of developmental activities.
Self-Assessment

Assessees' ratings of themselves in this study paralleled the results of previous studies (DeNisi & Shaw, 1977; Meyer, 1980). Two-thirds of the self-ratings were either "very good" or "excellent", and one-quarter were "average." There were very few "fair" ratings, and practically no "poor" ratings. The assessees saw their abilities as at least average, and usually above average.

Despite the fact that the ratings were all quite high, there was variability across the 16 dimensions. For the pretest ratings, the self-ratings ranged from a low of 3.11 (1=poor, 5=excellent) for Evaluation to a high of 4.39 for Initiative. Evaluation also had a large standard deviation (1.23). The definition of evaluation is not as clear as that of the other dimensions, and it may have been that the assessees were uncertain how to rate themselves. The high ratings for Initiative, Planning (M=4.28), and Collaborativeness (M=4.28) may indicate that the assessees
consider these to be critical dimensions, ones in which they are reluctant to give themselves low ratings.

The overall change in self-ratings did not change from pretest to posttest. However, this overall lack of significance masked the fact that the assesses rated themselves higher on the pretest than the posttest for some dimensions, and lower on the pretest than the posttest for other dimensions. For Written communication and Initiative, pretest scores were significantly higher than posttest scores. Some of the assesses expressed dissatisfaction with the written portions of the exercises, and claimed that their jobs did not involve much writing. Therefore, the assesses may have felt pressured at being asked to do much writing within limited time periods, and may have felt they did not perform as well as they could have. The assesses may have felt the assessment center did not allow them to display their initiative. Since the assesses were pretty much told what to do in each exercise, they may not have seen many opportunities to begin actions without stimulation from others.

For Decision-making, Organizational sensitivity, Sensitivity, and Management control, the posttest ratings were higher than the pretest ratings, which was surprising considering that the pretest ratings were already skewed toward the high end of the scale. While the assesses may
not have felt they had opportunities to show initiative, they may have seen many opportunities to display their decision-making ability, organizational sensitivity, and management control. And, since they had overall favorable perceptions of themselves, may have felt they had performed well in the exercises designed to test these abilities. The change in self-assessment for Sensitivity is difficult to explain. This would seem to be a dimension with which people were very familiar, and not one that would be raised through an evaluation of performance in the assessment center exercises.

Self-Assessment vs. Assessor Ratings

When comparing the self-ratings to the assessor ratings, it was found that the assessees rated themselves higher than they were rated by the assessors. The assessor ratings were not as skewed toward the high end of the scale as were the self-ratings. This finding is consistent with the majority of studies on self-assessment which have found that individuals rate themselves higher than they are rated by others (Thornton, 1980).

Given the fact that self-ratings are generally higher than the ratings given by others, many studies have used correlations to determine if there is any agreement between self-assessments and assessments given by others. Brief,
Aldag, and Van Sell (1977) reviewed nine studies and found modest correlations ranging from a low of .05 (Klimoski & London, 1974) to a high of .60 (Williams & Seiler, 1973). In the present study, the correlations between pretest and posttest self-ratings and assessor ratings were very low (.01 and .27), which indicates that the assessees and assessors did not agree on which assessees were the most and least capable.

There are several factors which have been suggested as affecting the accuracy of self-ratings. The first factor is anonymity. Research has shown that publicly announced self-appraisals tend to be more modest than self-appraisals given in confidence (Parker, et al., 1959). According to Bradley (1978), if your self-evaluation is to be made public, and if others will be evaluating you on the same dimensions you have evaluated yourself on, it would be embarrassing to give yourself ratings that are higher than you deserved. In the present study, the self-ratings were partially anonymous. Assessees were asked to put their names on their self-ratings, but they were also told that the ratings would be kept confidential and would not be shown to the assessors.

A second factor is whether the self-evaluation measure is an ability or performance assessment. Self-evaluations of ability involve an interpretational process in which the
individual must decide to what extent ability was a factor in determining performance. Self-evaluations of performance need not be interpreted and therefore tend to be accurate (Mabe & West, 1982). In this context, the asseesees were making inferences about their abilities based on their past performance. This was a more difficult task than rating their past performance.

Third, a self-evaluation phrased in relative terms is expected to correlate higher with criterion measures than a self-evaluation phrased in absolute terms. It is easier to rate oneself as average, below average, or above average, compared to one's peers, than it is to rate oneself as fair, poor, or excellent. The scale that asseesees used to make their self-ratings was essentially an absolute one; they were asked to rate themselves as poor, fair, average, very good, or excellent.

Four, Festinger (1954) proposed that there are essentially two motivations operating in evaluation of ability: the desire to obtain accurate evaluation information (self-evaluation), and the desire to enhance the perception of one's competence (self-enhancement). These two motivations would be expected to effect self-evaluations quite differently. When people are motivated to obtain accurate evaluation information, their self-evaluations would be more realistic and more valid, while people
motivated by self-enhancement desires would make inflated self-evaluations that were less valid. For the assessees, both motivations were probably operating. They desired accurate evaluation information since this would help them to be more effective on their jobs and could increase the likelihood that they would be promoted. On the other hand, since performance in the assessment center would also be one factor used in considering whom to promote, they would also be motivated by self-enhancement.

Although the assessees were not accurate overall in making their self-ratings, they were more accurate on some dimensions than on others. For the pretest ratings, there were nine dimensions on which there were not significant differences between assessee and assessor ratings. These dimensions are Oral communication, Decision-making, Development of coworkers, Organizational sensitivity, Assertiveness, Objectivity, Sensitivity, Management control, and Evaluation. For the posttest ratings, the assessees were most accurate on Oral communication, Written communication, Leadership, Management control, Collaborativeness, and Evaluation. Mihal & Graumenz (1984) found that self-raters and assessors agreed most on dimensions where the behaviors were most concrete (i.e., writing, speaking), and least where behaviors were more abstract (i.e., sensitivity, work standards). The findings
here were not so clearcut. Although assessees and assessors agreed on ratings for a concrete dimension such as Written communication, they also agreed on Management control, which is quite abstract and open to individual interpretations.

The three groups were not equivalent in the way they rated themselves. The Low group gave themselves higher ratings than the other two groups on both the pretest and the posttest. This was understandable considering that the Low group was comprised mostly of Supervisors-in-Training, who had presumably been promoted to that position because of their supervisory abilities. However, if they were in fact more capable than the agents in the Moderate and High groups, they should have received higher ratings from the assessors. This was not the case. There was not a significant difference in the assessors' ratings of the assessees in the three groups. The Low group's high self-ratings, therefore, were not justified. The finding that the assessors did not rate the three groups differently showed that the assessors' ratings were not swayed by the assessees' current position. It also shows that, if they were aware that the High group had received self-assessment training, which they probably were not, it did not affect the way they rated the assessees in that group.

One of the most surprising findings in this study was that the assessees were less accurate in making their
posttest self-ratings than they were in making their pretest self-ratings. The opposite results were expected. More accurate evaluations are obtained when there is agreement between the self-evaluation measure and the criterion measure of interest (Mabe & West, 1982). The assessors based their ratings on the assessee's performance in the assessment center exercises. For the posttest self-ratings, assessee were told to evaluate their abilities focusing on their performance in the assessment center exercises. In contrast, with the pretest ratings, the assessee were using all of their past experiences to make their self-evaluations. Since the assessor ratings and the posttest ratings were evaluations of the same behaviors in the same contexts, they should have been more similar than the assessor and pretest ratings.

One suggestion made by Shrauger and Osberg (1981) for enhancing the accuracy of self-assessments is to facilitate the recall of relevant past experiences. This was done with the posttest ratings by telling the assessee which dimensions were being assessed in what exercises, and by asking them to focus on their performance in those exercises in making their self-ratings. For the pretest ratings, the assessee were not directed to recall past experiences.

Another reason why the posttest ratings should have been more accurate is because of experience.
Self-evaluations generally become more accurate as people gain experience in evaluating their abilities (Kooker, 1974; Levine, et al., 1977). The posttest ratings were the second time the assesses rated themselves on these dimensions. Although having one previous experience cannot be considered extensive experience in self-evaluation, it still should have helped make the posttest ratings more accurate, not less accurate.

It was hypothesized that the assesses in the High group would be the most accurate in making their self-assessments, and the Low group would be the least accurate. Since the High group were trained in self-assessment, their ratings should have been the most accurate. In addition, both the Moderate and High groups were encouraged to assess themselves during the assessment center, so their attention was focused on themselves. Several studies have found that when self-focused attention is heightened, self-reports are more accurate (Gibbons, et al., 1979; Pryor, et al., 1977). A comparison of assessor and posttest ratings would seem to support this hypothesis. The High group's self-ratings were not significantly different from the assessor ratings, the Moderate group was significantly different at the .05 level, and the Low group was significantly different at the .01 level. However, the comparison of assessor and pretest ratings indicate that the
groups were not equivalent prior to the assessment center. The Low group's self-ratings were significantly higher than the assessor ratings at the .006 level, while the Moderate and High groups were not significantly different. The Low group became somewhat more accurate as they went from pretest to posttest, the Moderate group became less accurate, and the High group was slightly less accurate. Therefore, Hypothesis 1 was not supported.

Satisfaction, Perceived Validity

It was hypothesized that assessees in the High involvement group would respond more favorably to the assessment center than assessees in the Low involvement group. According to adult learning theory, the more involved individuals are in the learning process, the more satisfied they will be with the learning activity (Long, 1983), and the more likely they are to accept the process as valid (Knowles, 1978).

The findings of this study did not support this theory. There was not a significant difference between the three groups in their reported satisfaction with the assessment center and their perceptions of the validity of the assessment center Thornton and Byham (1982) stated that the assessees generally like the assessment center process and perceive it to be a valid way to assess a person's
abilities. When responses are skewed toward the high end of the scale, it is more difficult to discriminate between groups. However, the responses given in this study for satisfaction and perceived validity were fairly well spread out; they were not skewed toward one end of the scale or the other. The lack of findings were probably due to the fact that the manipulation was unsuccessful. Since the groups did not differ on involvement, it cannot be expected that they would differ on the dependent variables.

Intentions to Engage in Developmental Activities

A repeated finding in the adult learning literature is that a self-diagnosed need for learning produces much greater motivation to learn than an externally diagnosed need (Knowles, 1980). Since the High involvement group was involved in self-assessment, and the Low involvement group was not involved in self-assessment, it was hypothesized that assesses in the High involvement condition would report more intentions to engage in developmental activities than assesses in the Low involvement condition. The results did not support this hypothesis. There was no significant difference between the three groups in their stated intentions to participate in developmental activities over the next 12 months.
The lack of significance may have been due to some problems with the question which was intended to measure this variable. The assessees were supposed to indicate how many and what types of developmental they planned to participate in during the next 12 months. However, some of the assessees seemed to feel that if they wrote anything down on the questionnaire, they would be committing themselves to participating in those activities. So, rather than answer the question, they wrote notes saying they wanted to wait awhile before choosing developmental activities. They were then given a zero for this question. This may indicate a lack of trust on the part of the assessees, perhaps because they were aware that the assessment center results might impact on their career in the organization. If all assessees had answered this question as it was intended, a significant difference might have been found. One positive finding this does bring out, though, is that the assessees took the questionnaire seriously and were careful in how they responded to the questions.

Learning

The more active the learners are in a learning situation, the more they will learn (Knowles, 1950). Therefore, it was hypothesized that the assessees in the
High involvement condition would report that they learned more about themselves than assessees in the Low involvement group. The results concerning this hypothesis were significant but the opposite of what was expected. The Low involvement group reported the most learning, and the High involvement group reported the least learning. Since this finding was so surprising, the three questions which make up this variable were examined individually.

One question appeared to be responsible for most of the variation between groups, "My perceptions of my abilities changed as a result of participating in the assessment center." The Low group gave themselves self-ratings that were much higher than the self-ratings of the other two groups, yet did not receive higher ratings from the assessors. Therefore, there was more of a discrepancy between the Low group's self-ratings and the assessor ratings than between the Moderate and High group's self-ratings and assessor ratings. When these differences came out during the feedback process, the Low group would have had to change their perceptions of their abilities more than the assessees in the Moderate or High groups.

Another question which was not significant by itself but which contributed to the overall significance of the variable was "The assessment center gave me added information about what is expected of me on the job." The
Low group were almost all Supervisors-in-Training, while the Moderate and High groups were County agents. Since the assessment center was designed for managerial jobs, the Supervisors-in-Training may have found the assessment center exercises and dimensions to be closer to their jobs than did the County agents. Therefore, they were able to learn more about what would be considered good performance on their jobs.

Limitations of the Current Study

Every study has limitations; this section will discuss the limitations of the current study, and how these limitations may have shaped the results.

The sample size for this study was very small. Given the time, expense, and resources involved in sending a person to an assessment center, organizations do not assess large numbers of people at once. Therefore, a small sample size is a reality of doing this type of assessment center research. The problem with small sample sizes is that the groups are less likely to be equivalent prior to the manipulation, and differences between groups will be less likely. In the present study, the differences between groups were probably due to the fact that people were not randomly assigned to groups, which will be discussed next. With a large sample size, it is possible to detect very
small differences between groups on the dependent variables. However, these differences, while they may be statistically significant, may be too small to be of any practical value. Although a larger sample size may have generated more significant results which would have been of interest to researchers, they might have been too small to be of use to anyone working with assessment centers in an applied setting.

The sample size did limit the analyses which could be done. It was not possible to test the scales for internal reliability, or to factor analyze all the items to determine how many different variables actually existed.

The assesses were not randomly assigned to the Low group. The assignment to groups for the Moderate and High groups was essentially random, since both groups were County agents, and no attempt was made to assign certain people to one group or the other. But the Low group was different from the other two groups. Five or the six were Supervisors-in-Training, not County agents, and they gave themselves higher self-ratings than did the other two groups. The fact that the Low group had more favorable perceptions of their own abilities, or other potential differences between the groups, may have affected their responses to some of the other dependent variables, such as satisfaction or intentions to engage in developmental
activities. Different results may have been obtained if the groups had been more similar.

The design of the study was not completely balanced. The Low group did not receive some of the questionnaires which were given to the other two groups. For example, they did not receive the first manipulation check. The Low group also did not receive the first questionnaire which asked about satisfaction, learning, and perceived validity. Since significant results were not obtained for satisfaction and perceived validity, and the results for learning were puzzling, the study would have been stronger if these variables had been assessed at two points in time for all three groups.

It was intended that all assesseses would have had their feedback interview before receiving the second questionnaire. It was felt that negative reactions to the feedback they received might alter the assesseses' perceptions of the assessment center process. But because of scheduling problems between the assesseses and assessors, three of the assesseses had not yet had their feedback interview when the second questionnaires were sent out. Due to the small sample size, it was not possible to compare assesseses who had or had not had their feedback interviews to determine if their responses to questionnaire differed. However, the assesseses who had had their feedback interview reported it
to be a positive and useful experience, so it does not seem likely that their reactions to the assessment center were adversely effected.

There was only one question which asked about intentions to engage in developmental activities and it turned out to have problems. It was designed to be specific so that assesses would have to think about their personal development. Unfortunately, some assesses refused to respond so specifically. A more general question, where assesses just had to state the number of developmental activities they planned to engage in, and didn't have to write in the specific types of activities, might have yielded a higher response rate. Another solution would be to simply wait a year and then follow-up with each of the assesses, asking them how many developmental activities they have actually participated in.

The most serious limitation to this study was the fact that the manipulation check was only partially successful. Some of the questions showed differences between the three groups, but other questions did not. For the comparisons of self-ratings and assessor ratings, this is not a problem, since these findings were not dependent on there being differences between groups. But it does impact on the conclusions we can draw for the dependent variables satisfaction, learning, perceived validity, and intentions
to engage in developmental activities. The fact that no differences emerged in this study does not mean that the theory was incorrect. Since it can't be proved that the groups differed on involvement, it can't be expected that they would differ on the dependent variables.

**Directions for Future Research**

Research on assessment centers has always used archival data; this is the first study in which an attempt was made to manipulate anything in the assessment center. In addition, this was the first study which looked specifically at developmental assessment centers. Obviously, much more work is needed in this area.

Assessment centers are a natural environment in which to study self-assessment, since the assessors are specifically trained to evaluate behavior, and since the assesseees and assessors can evaluate the same behavioral episodes. In this study, it was found that there was more agreement between the assesseees and assessors on some dimensions than on others although there was no clear pattern to the results. Future research could examine under what types of circumstances or for what types of dimensions do self-ratings tend to agree with the ratings made by others.
Another interesting aspect to assessment centers is that assessees are given feedback from the assessors, so they have a chance to compare their self-evaluations to the evaluations given by the assessors. It would be interesting to find out how assessees deal with evaluations from the assessors that differ from the assessees' self-ratings, whether they change their self-ratings to be more congruent with the assessors' ratings, and whether this change is a lasting one.

For the present study, involvement was manipulated by varying the information assessees were given on self-assessment and whether or not they were encouraged to self-assess. There are other ways to encourage involvement, although they would be very difficult to implement. Assessees could be involved in developing the assessment center, deciding what dimensions should be assessed, etc. Or they could be given more decision-making power within the assessment center, such as having choices of what exercises they would participate in, or how the feedback interview would be given. These activities would all be consistent with the conceptual definition of involvement provided by Webster (1961).

In developmental assessment centers, a primary concern is whether assessees will take the information about themselves which they have learned in the assessment center
and use it to develop areas where their skills need improvement. If they do not attempt to change, then the assessment center has served only an evaluative function, not a developmental one. Research is needed on how assessment centers do or do not motivate people to participate in developmental activities. Research should also focus on what types of developmental activities people choose. Do they focus on their weakest areas, where the most work is needed, or do they choose to focus on stronger areas, where they are assured of succeeding? Do they participate in activities which are congruent with their self-assessed needs, or their needs as evaluated by the assessors? These are just some of the questions that need answering.

**Summary and Conclusions**

The purpose of this study was to examine how assesses in assessment centers rate themselves, how their self-ratings might be made more accurate through training, and how this training would affect the assesses’ perceptions of the assessment center. Adult learning theory states that adults need to be actively involved in the learning process. In this study, involvement was manipulated by varying the amount of training assesses were
given on self-assessment and the extent to which they were encouraged to assess themselves.

The results showed that the assessees rated their abilities higher than did the assessors, and this difference was more pronounced after the assessment center than before the assessment center. Although it is unclear why the post-assessment center ratings were less accurate than the pre-assessment center ratings, this does show that even when individuals and observers are evaluating the exact same instances of behavior, their interpretations of the individual's abilities are different.

Another finding was that assessees in the Low involvement group reported greater learning than did assessees in the other two groups. The assessees in the Low group were mostly supervisors-in-training and the assessment center exercises were more similar to their jobs than they were to the jobs of the county agents in the other two groups. Therefore, the Low group learned more about what was expected on the job by participating in the assessment center. In addition, since the Low group gave themselves self-ratings that were much higher than they deserved, they learned more about the true extent of their abilities in the feedback interview than did the assessees who had a more accurate picture of their abilities.


APPENDIX A

FIRST ASSESSMENT CENTER EVALUATION
ASSESSMENT CENTER EVALUATION

Name____________________________

The staff at Cooperative Extension Service is interested in learning what you think about the assessment center you just participated in. Your feedback will help use to revise and improve the assessment center. We are asking that you put your name on the questionnaire, so that we know who has responded. However, all responses will be kept strictly confidential and will not be shown to any of the assessors.

Use the following scale:

1=strongly agree
2=agree
3=neither agree nor disagree
4=disagree
5=strongly disagree

___1) I found participating in the assessment center to be an enjoyable experience.

___2) The two days of the assessment center were difficult one for me.

___3) The assessment center did not meet my expectations.

___4) I enjoyed all of the exercises.

___5) I was satisfied with my performance in all of the exercises.

___6) I feel I learned quite a bit about myself as a result of participating in the assessment center.

___7) I feel I'll gain added information about my abilities from the feedback interview.

___8) My perception of my abilities changed as a result of participating in the assessment center.

___9) The assessment center gave me added information about what is expected of me on the job.
10) I felt my performance in the assessment center exercises was an adequate reflection of my abilities.

11) During some of the exercises I acted in ways that are not typical for me.

12) I feel that an assessment center is an appropriate way to evaluate a person's skills and abilities.

13) I didn't see the purpose behind some of the exercises.

14) I was told what dimensions I would be evaluated on prior to the start of the assessment center.

15) My evaluation of my own abilities was treated as an important part of the assessment center process.

16) I was given no information about what would happen in the assessment center prior to the first day of the assessment center.

17) I was not told anything about why the assessment center is an appropriate way to evaluate a person's abilities.
APPENDIX B

SECOND ASSESSMENT CENTER EVALUATION
The staff at Cooperative Extension Service is interested in learning what you think about the assessment center you participated in on February 25-26. Your feedback will help us to revise and improve the assessment center. We have enclosed a stamped, self-addressed envelope for your convenience. We are asking you to put your name on the questionnaire, so that we know who has responded. However, all responses will be kept strictly confidential and will not be shown to any of the assessors.

For the first 16 questions, use the following scale:

1=strongly agree
2=agree
3=neither agree nor disagree
4=disagree
5=strongly disagree

___1) I found participating in the assessment center to be an enjoyable experience.

___2) The two days of the assessment center were difficult ones for me.

___3) The assessment center did not meet my expectations.

___4) I enjoyed all of the exercises.

___5) I was satisfied with my performance in all of the exercises.

___6) I feel I learned quite a bit about myself as a result of participating in the assessment center.

___7) My perception of my abilities changed as a result of participating in the assessment center.

___8) The assessment center gave me added information about what is expected of me on the job.
9) I felt my performance in the assessment center exercises was an adequate reflection of my abilities.

10) During some of the exercises I acted in ways that are not typical for me.

11) I feel that an assessment center is an appropriate way to evaluate a person's skills and abilities.

12) I didn't see the purpose behind some of the exercises.

13) I was told what dimensions I would be evaluated on prior to the start of the assessment center.

14) My evaluation of my own abilities was treated as an important part of the assessment center process.

15) I was given no information about what would happen in the assessment center prior to the first day of the assessment center.

16) I was not told anything about why the assessment center is an appropriate way to evaluate a person's abilities.

17. Have you had your feedback interview with the assessor?
   Yes (go to Q.19) No (go to Q.18)

18. Answer this question only if you have not yet had your feedback interview with the assessor.

18a) I feel I'll gain added information about my abilities from the feedback interview.

1. strongly agree 4. disagree
2. agree 5. strongly disagree
3. neither agree nor disagree

19. Answer these questions only if you have already had your feedback interview with the assessor.

Use the following scale:

1=strongly agree 4=disagree
2=agree 5=strongly disagree
3=neither agree nor disagree
19a) I felt I learned more about my abilities from the feedback interview than I learned during the actual assessment center.

19b) I felt the assessors accurately rated my abilities on the assessment center dimensions.

19c) I felt the assessors gave me some ratings that were lower than I actually deserved.

19d) I felt the assessors gave me some ratings that were higher than I actually deserved.

Comments about the feedback interview: ____________________
You have had some time to think about what types of developmental activities you would like to engage in based on your evaluation of your skill levels on each of the 16 assessment center dimensions. In the spaces below, indicate which activities you will participate in during the next 12 months. To the left of each activity, indicate the number you will participate in, then list each topic area below. For example, if you plan on taking two college courses:

____ College courses
  (topics):

_____________________________
_____________________________
_____________________________
_____________________________

____ College Courses
  (topics):

_____________________________
_____________________________
_____________________________
_____________________________

____ Workshops
  (topics):

_____________________________
_____________________________
_____________________________
_____________________________

____ Readings
  (topics):

_____________________________
_____________________________
_____________________________
_____________________________

____ Other
  (describe):

_____________________________
_____________________________
_____________________________
_____________________________
Your cooperation in completing this questionnaire is greatly appreciated. Please return completed questionnaire to:

Keith Smith
Leader, Personnel Development
Cooperative Extension Service
2120 Fyffe Road
Columbus, Ohio 43210
APPENDIX C

SECOND MANIPULATION CHECK
PRE-ASSESSMENT CENTER QUESTIONNAIRE

Name______________________________

The staff at Cooperative Extension Service is interested in finding out how much information people feel they have about the assessment center prior to when the assessment center starts. The following questions ask about things you might or might not have been told about the assessment center. This information could have come from anyone, including assessment center staff, and might have been conveyed either orally or in writing. Try and remember what you had been told about the assessment center prior to the start of the first assessment center exercise. Although we ask that you put your name on the questionnaire, your responses will be kept strictly confidential.

Use the following scale:

1=strongly agree
2=agree
3=neither agree nor disagree
4=disagree
5=strongly disagree

___1) My ratings of my own abilities were considered an important part of the assessment center process.

___2) I was told what factors people use to make their assessments of their own abilities.

___3) I was told how self-assessments differ from the ratings of others.

___4) I was told why assessment centers are an appropriate way to assess a person's abilities.

___5) I was given information about self-assessment which I felt would help me evaluate myself more accurately.
The purpose of the following items is to determine how you felt about the assessment center while you were participating in the assessment center exercises. Each item has a 1 to 5 scale with each end-point being anchored by an adjective. For each item, circle the number which most closely describes, to the best of your recollection, your feelings or your perceptions about the assessment center while you were participating in it.

1------------2------------3------------4------------5
Involved
Uninvolved

1------------2------------3------------4------------5
Committed
Uncommitted

1------------2------------3------------4------------5
Tense
Calm

1------------2------------3------------4------------5
Confident
Unsure

1------------2------------3------------4------------5
Self-directed
Directed by Others

1------------2------------3------------4------------5
Happy
Unhappy

1------------2------------3------------4------------5
Important
Unimportant

1------------2------------3------------4------------5
Interesting
Boring

1------------2------------3------------4------------5
Difficult
Easy
PRE-ASSESSMENT CENTER QUESTIONNAIRE

Name ____________________________

The staff at Cooperative Extension Service is interested in finding out how much information people feel they have about the assessment center prior to when the assessment center starts. The following questions ask about things you might or might not have been told about the assessment center. This information could have come from anyone, including assessment center staff, and might have been conveyed either orally or in writing. Respond to the questions using any information you have been given up to this moment. Although we ask that you put your name on the questionnaire, your responses will be kept strictly confidential and will not be shown to any of the assessors.

Use the following scale:

1=strongly agree
2=agree
3=neither agree nor disagree
4=disagree
5=strongly disagree

____1) My ratings of my own abilities are considered an important part of the assessment center process.

____2) I was told what factors people use to make their assessments of their own abilities.

____3) I was told how self-assessments differ from the ratings of others.

____4) I was told why assessment centers are an appropriate way to assess a person's abilities.

____5) I was given information about self-assessment which I feel will help me rate myself more accurately.

(continued on next page)
The purpose of the following items is to determine how you are currently feeling or your current perceptions about the assessment center. Each item has a 1 to 5 scale with each end-point being anchored by an adjective. For each item, circle the number which most closely describes your feelings or your perceptions.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involved</td>
<td>Uninvolved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Committed</td>
<td>Uncommitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tense</td>
<td>Calm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confident</td>
<td>Unsure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-directed</td>
<td>Directed by Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>Unhappy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>Unimportant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td>Boring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

SELF-ASSESSMENT TRAINING
On May 13th, you will be attending a two-day assessment center. The purpose of the assessment center is to give you a better understanding of your own abilities, particularly as they relate to the job of County Chair. During the assessment center, the assessors will be observing your behavior during each of the exercises and using your performance to make inferences about your abilities. Their evaluations will be communicated to you at a feedback interview which will be held shortly after the assessment center.

However, since you're the one who has to take responsibility for your development, we don't want you to rely solely on the judgments of others. We'd also like you to be assessing your own abilities. You have already been sent the list of assessment center dimensions and have rated yourself on each dimension. In order to assist you in your self-evaluations during the assessment center, we will be holding a short training session immediately prior to the assessment center. You will be asked to rate yourself again on each dimension after the assessment center.

The agenda for the training session is as follows:

I. Reason for using assessment centers to assess a person's abilities
   1. History of assessment centers
   2. Why assessment centers are an improvement on performance appraisals by supervisors

II. How people develop their self-assessments
   1. Use our direct experiences and the effects produced by our actions
   2. Observe the actions of others and compare them to our own actions
   3. Use the judgments that others express about our abilities
   4. Use our powers of reasoning to decide how effectively we could deal with new situations

III. Self-assessments vs. the ratings of others
1. Our self-ratings are usually higher than the ratings we get from others
2. Individuals may need to distort their self-ratings in order to maintain their self-esteem
3. Individuals and others may be using different criteria to rate the individual's performance

IV. Discussion of dimensions

The assessors were trained to focus on specific behaviors. In order to make your self-ratings, you also need to focus on behavior and make some decisions about what behaviors indicate ability on each dimension.

1. Oral Communication
2. Written Communication
3. Sensitivity
4. Initiative
5. Leadership/Persuasiveness
6. Objectivity
7. Assertiveness
8. Perception
10. Planning/Organizing
11. Behavioral Flexibility/Adaptability
12. Development of Coworkers
13. Collaborativeness
14. Organizational Sensitivity
15. Management Control
16. Evaluation

The training session will start at 9:00 on May 13th and the assessment center will begin at 10:00. We appreciate your cooperation and hope that this training session will make the assessment center a useful learning experience for you.
APPENDIX F

PERFORMANCE SCORES AND SELF-REPORT ITEMS
CORRELATION OF PERFORMANCE SCORES TO SELF-REPORT ITEMS

There are two ways to determine the assesses's overall performance. One way is to add together the consensus ratings of the assessors on each of the 16 dimensions to get a total score; the other way would be to use the single overall rating which was also developed by the assessors. With the total rating, the higher the score, the better the performance. The reverse is true for the overall rating.

The assesses's performance on the assessment center exercises could effect the way he/she felt about the assessment center. The correlations between the two performance measures and the assesses's reported satisfaction, perceived learning, perceived validity, and intentions to engage in developmental activities are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Total Score</th>
<th>Overall Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>-.48 (.04)</td>
<td>.44 (.07)</td>
</tr>
<tr>
<td>Learning</td>
<td>-.48 (.04)</td>
<td>.54 (.02)</td>
</tr>
<tr>
<td>Validity</td>
<td>.40 (.10)</td>
<td>-.43 (.07)</td>
</tr>
<tr>
<td>Intentions</td>
<td>.47 (.05)</td>
<td>-.60 (.009)</td>
</tr>
<tr>
<td>Total Score</td>
<td>-.87</td>
<td>(.0001)</td>
</tr>
</tbody>
</table>