1986

A Comparison of Test Scores Obtained from Equivalent Forms of a Multiple Choice In-Basket Exercise and a Free Response In-Basket Exercise

German B. Romero
University of Central Florida

A COMPARISON OF TEST SCORES OBTAINED FROM EQUIVALENT FORMS OF A MULTIPLE CHOICE IN-BASKET EXERCISE AND A FREE RESPONSE IN-BASKET EXERCISE

BY

GERMAN B. ROMERO
B.A., University of Central Florida, 1978

THESIS

Submitted in partial fulfillment of the requirements
for the Master of Science degree in Industrial/Organizational Psychology
in the Graduate Studies Program of the College of Arts and Sciences
University of Central Florida
Orlando, Florida

Fall Term
1986

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Management has always sought better and more efficient methods of selecting supervisors and managers. The need for selection tools appears more important than ever. Management personnel decisions have been made more critical by the increasing complexity of the business environment and by the increased intervention of litigation. The increasing cost of time and human resources further demands that managerial selection decisions be made quickly and as accurately as possible.

Although probably still in use today by some organizations, the original method of selecting managers by relationship or recommendation from a relative or friend has faded. Business connections were invaluable in getting ahead. As the business environment became more complex this method became impractical. Psychologists sought to provide an alternative by developing evaluation techniques (DuBrin, 1972; Strauss & Sayles, 1972). Personality tests, profiles, and inventories were soon in wide use throughout business organizations. But these too proved imperfect; not only were the courts deciding that they violated the rights of women and minorities to equal employment opportunities but also, the users found these tests inadequate in selecting the right candidate for the job. Although still in limited

use today, personality tests are taking on less importance as other methods or techniques have evolved (Baritz, 1960; DuBrin, 1972; Megginson, 1967; Strauss & Sayles, 1972).

One such technique that has evolved is the Assessment Center. The Assessment Center originated with the concept that the best predictor of future managerial potential is past managerial performance in a similar or related position. This concept has been refined to mean that the best prediction of managerial potential is made by observing and evaluating behavior exhibited during the performance of relevant managerial tasks. The typical assessment center is made up of various exercises designed to evaluate various appropriate managerial dimensions.

One of the most commonly used exercises in an Assessment Center is the In-Basket Exercise. It was developed by Dr. Normal Frederiksen and the Educational Testing Service in an attempt to devise a sensitive measure of managerial potential which could also be objectively and reliably scored (Frederiksen, Saunders, & Wand, 1957). Sparked by the complexity of modern business and government enterprises and a need for accurate and realistic techniques to identify, select, and develop people with the ability to hold key positions, the In-Basket has evolved from a business game into an integral part of the personnel selection system (Lopez, 1966). The In-Basket has been shown to be reliable and valid (when used appropriately and in combination with other tools) in selecting managers and supervisors for all levels of management.

One of its strengths and at the same time its major drawback is its complex scoring system. A typical In-Basket may take from two to four hours to be accurately scored. This can prove to be quite an obstacle when it is considered that a key position may require the screening of 100 or more candidates.

This paper will attempt to study the problem of the In-Basket scoring system. Is it possible to develop an easily scorable In-Basket while retaining its validity and reliability? More specifically, will an In-Basket Exercise designed on a multiple choice answering format be as effective as the standard free-response format In-Basket?

CHAPTER II

REVIEW OF LITERATURE

To address the potential of a multiple choice answer In-Basket in the selection of managers, the job of the manager must be analyzed. In a study by Meyer (1970) a 38 item In-Basket Exercise was designed and given to 81 unit managers. Ratings of observed job performance of these 81 managers were collected from higher level managers. Factor analysis of the scores on the items and on the ratings revealed two major dimensions of the manager's job. The first was a supervision or human relations factor and the second an administrative or planning factor. When Meyer performed correlations between the In-Basket scores and the supervisory ratings, the results showed that scores were better predictors of the administrative/planning factor than of the supervisory factor. When results were cross validated with an additional 45 unit managers using weights developed by comparing the specific courses of action taken by managers who received above average scores on the ratings of each of the two factors with those who received below average ratings, only the administrative/planning factor was found to have any predictive validity. The correlations were found significant at the .05 level of confidence. Additional work by Mintzberg (1973), Shapira and Dunbar (1980) sought to provide a taxonomy of managerial work. Based on a longitudinal study of five chief executives, Mintzberg was able to divide managerial work into

three categories: interpersonal, informational, and decisional. In a series of studies, Shapira and Dunbar attempted to test Mintzberg's theory. Using a total of 166 subjects including students and managers of various levels, Shapira and Dunbar administered an In-Basket containing 16 items. Results indicated that the manager's work could be regrouped into an information generation and processing role and a decision-making role. The interpersonal role was relegated to a supporting role of the other two roles.

The results of these studies indicate that the In-Basket Exercise is not a reliable measuring instrument of a manager's interpersonal skills. This inability is grounded on the very nature of the exercise. As Shapira and Dunbar point out, a candidate's behavior during evaluation by the In-Basket method does not fully correspond to the actual on-the-job behavior because a manager is not typically limited to written communication in the performance of his/her duties. Behavior during the exercise, however, is limited to written communication. The candidate is instructed to write what he/she would do in the given situations. The situations may require personal contact with others which is not possible during the exercise. The written response may or may not correspond precisely to what the candidate would do under actual "real" circumstances. The results of these studies indicate that the In-Basket should not be used to select for the possession of interpersonal skills.

Because the In-Basket scores are not traditionally used as the sole criterion in selection, the inability of the In-Basket to predict

potential success in the interpersonal phase of managerial work is overcome by evaluating interpersonal skills through two or more different Assessment Center Exercises. Thus the evaluation of skills needed to effectively work one-on-one with another person becomes secondary when evaluating the effectiveness of a multiple-choice format In-Basket Exercise. Effectiveness must be measured by how well the exercise is able to predict success in the administrative, planning, and decision-making duties of managerial work.

Lopez (1966) provided some insight into the usefulness of a multiple choice In-Basket in the evaluation of administrative planning and decision-making skills. Lopez proposed that a manager's job is a set of responses to a set of inputs. The range of responses varies from highly effective to totally ineffective. Lopez further stated that the inputs are not random and that each input requires the subject to choose from an array of alternatives available to him/her. The pattern of responses typically chosen describes the manager's style of managing and determines his/her effectiveness. Overall effectiveness can be gauged by the importance of problems concentrated on, the amount of work produced, appropriateness of the decision (financial and social consequences), the ability to judge a situation correctly, the ability to make decisions under time pressure, and the ability to take steps to secure the cooperation of peers, superiors and subordinates. The In-Basket Exercise can be designed to parallel these activities.

Because the inputs encountered by the manager are not random, it is possible to develop highly effective In-Basket Exercises based on representative samples of managerial work. The range of responses available to handle a typical input is also not random nor usually unique. The effectiveness of the response to each problem is determined by how accurately the manager assesses the input information and selects the most effective choice of possible actions. The information provided during the exercise is very specific and forms a set of interlocking situations that forces the participant to make decisions. Procrastination or avoiding the critical issues is itself a decision. Thus what the candidate does or does not do provides important clues as to the effectiveness of the candidate as an administrator or decision-maker. A well constructed In-Basket can be prepared listing all possible actions available to the candidate. The actions chosen can then be easily evaluated along the desired dimensions.

Studies by both Lopez (1966) and Frederiksen (1962) indicate that the courses of actions taken by candidates is limited to a fairly small number of responses, thus making it easier to construct a proper multiple choice In-Basket Exercise. Using an exercise called the Bureau of Business Test, similar to an In-Basket Exercise, Frederiksen was able to develop a list of up to only 10 courses of action for each of 31 problems and five phone calls to evaluate the content of the responses given by 335 subjects. Although the range of responses given by the candidates to each problem was fairly narrow, the typical

exercise does allow for the evaluation of unusual or imaginative responses. However, Lopez has found that the analysis of unusual responses yields not a measure of creativity of the candidate but rather a measure of uncertainty or nonconformity. Therefore, there is evidence to suggest that it is possible to develop multiple choice In-Baskets. Items can be carefully developed based on the job position to be filled, as in the regular In-Basket Exercise. Through consultation with job experts, all possible courses of action can be determined and weights assigned based on the job experts' evaluations of degrees of effectiveness or ineffectiveness. Courses of action for each item along various dimensions can be plotted. After the candidates have selected their responses, the In-Baskets can be scored manually or by computer scanners along the predetermined dimensions. Thus fast, reliable, highly objective scores can be obtained from all candidates.

To more accurately determine the usefulness of multiple choice In-Baskets the scoring dimensions used in evaluating managerial potential must be investigated. The scoring dimensions or categories are or should be determined by the specific job in question. After all, the task is to evaluate the candidate's potential for the particular job. As a first step in defining these dimensions and designing the In-Basket to accurately measure them, a Job Analysis is conducted (Cohen, 1980). The objective of the Job Analysis is to uncover the tasks and situations relevant to the demands of the job. This phase leads to the identification of knowledge, skills, abilities

and other personal characteristics (KSAPs) necessary to perform the critical tasks of the job. Identification of these critical KSAPs dictates the construction of the In-Basket Exercise. The items in the In-Basket are designed to allow the candidate to demonstrate the degree of possession or lack of possession of the critical KSAPs through the responses given. Trained scorers evaluate the responses along the dimensions previously identified and assign ratings or points on the basis of the predetermined scales (Brass & Oldham, 1976).

Much of the success of the In-Basket Exercise can be attributed to the design of its scoring system. However, its major drawback also lies in its time consuming scoring format which acts to restrain its use in the selection process. The first systematic effort to analyze the results numerically in directly making selection decisions, by the Port of New York Authority in 1960 for a police lieutenant's job, proved too costly (Lopez, 1966). The cost resulted from the very complicated scoring system required. The Port Authority later used a less complicated and less costly scoring procedure for selecting clerical and secretarial employees, but the success of the In-Basket must be measured in its effectiveness to select managers and supervisors. Eighty-five candidates were evaluated on 47 problems on operating and administrative aspects of the police command over a four hour period which included a one hour discussion by each candidate to prioritize problems and explain the reasons for the actions taken. Scoring was accomplished by comparing an individual's actions with a

list of possible actions prepared by a panel of scorees in 11 major categories. The candidate's response was given the numerical grade assigned to the action most resembling the response. A great deal of effort was given to training scorers and developing detailed ground rules for handling the many difficult scoring decisions that arose out of the variations in participant responses. In the beginning a scorer required seven hours to evaluate a single candidate; with practice the scorers were able to reduce scoring time to three hours. Interscorer reliability ranged from -.20 to +.97 with a median reliability of about +.60. With 85 candidates to evaluate, the 255 plus man hours required for the selection decision made future use of the exercise prohibitive without further reduction in the time required for scoring.

Other users of the In-Basket have encountered similar problems. One study by Jon Benz (Lopez, 1966) used a 37 item exercise. Scoring was found to be "time consuming and tedious" requiring from four to eight hours per In-Basket. A 34 page manual was written to score the nine dimensions developed. The In-Basket had been designed to evaluate candidates' ability to successfully hold the job of store manager for the Sears Company. Fifty-three subjects were tested along categories similar to other research studies of the In-Basket. As with the Sears study and the Port Authority studies, a study by General Electric's Research Division observed that while the In-Basket was a useful and valuable management selection aid, "its major weakness lay in its complex, tedious scoring process" (Lopez, 1966, p. 89).

A closer examination of how the dimensions are scored and the training received by the scorers adds credence to the use of a multiple choice In-Basket over the free response In-Basket. After the In-Basket items are constructed and before they are administered to candidates, scorers receive extensive training and practice in the evaluation of the testing material (Slevin, 1972). Training may require two to four days of instruction and practice. It may require the learning of a scoring guide or manual of a few pages or as many as 165 pages (Frederiksen, 1962). The training sessions concentrate on developing and/or understanding all possible ways of responding to the Exercise items. The scorers discuss and evaluate which courses of actions are the most effective for the circumstances given. Weights or points may be assigned to the various courses of actions possible based on the scorers' evaluations of degrees of effectiveness. The In-Basket is then administered to the candidates. After the candidates complete the exercise, the scorers evaluate the answers by comparing candidate responses with the previously studied courses of action. An interview with the candidate after completion of the exercise is usually held to get further insight into how the candidate intended to handle each item. The scorer matches the candidate's response with the course of action it most nearly resembles in the list of possible actions and awards the points assigned to that course of action or awards a positive mark for effective actions and a negative mark for ineffective actions. In the event of an unusual or unique answer the scorer may consult with other scorers to determine

appropriateness of the response. The score for each dimension is usually determined by adding the points awarded to each dimension or comparing the number of positive and negative marks awarded to the dimension across items.

Although scoring dimensions may be different in In-Basket Exercises, a set of dimensions is typically used: delegation, decisiveness, organizing and planning, perception or analysis and judgment or decision-making. Delegation (assigning work and responsibility to subordinates) is scored by comparing the number of times a candidate delegates work assignments and the completeness of the instructions given with a predetermined standard. A scale of appropriateness of delegation from too much delegation to too little delegation can be established. Completeness of instructions can be scaled from adequate to inadequate. Decisiveness (readiness to make decisions, render judgments, take action or commit oneself, and not needlessly seek or wait for further information) is measured by totalling the number of decisions made during the exercise. Organizing and planning (establishing a course of action to accomplish specific goals; making proper assignments to personnel and appropriate allocation of resources; establishing priorities) is measured by the number and order of handling problems, the number of meetings scheduled, the number of assignments made, the number of deadlines established, the lack of conflicting meetings, deadlines, etc., and the use of the calendar. Perception or analysis (identifying existing or potential problems or opportunities, obtaining relevant information, relating data from

different sources and identifying possible causes of problems) is scored by the number of problems correctly identified, the number of conflicting dates or assignments identified, the number of times additional information is correctly sought. Judgment or decision-making (making realistic decisions based on logical assumptions and facts) is measured by the total number of points or weights assigned to the various decisions made.

The scoring procedures are objective and fairly mechanical. The task becomes complex because many of the items are related and the scorer must consider how each decision affects the other items. This may involve sifting through pages of candidate responses to determine how other items were handled. Unlike other Assessment Center Exercises where scorers are required to learn observation and listening skills, the scorer's task is of a clerical nature. Subjective judgments are limited to correctly matching candidate responses to the most similar course of action studied by the scorers. A carefully constructed multiple choice In-Basket that includes all possible courses of actions can be used to parallel the scorer's job. With the advanced computer technology, a multiple choice In-Basket can be quickly scored by adding point values accorded to the various actions selected by the candidate. Printouts can quickly reproduce a list of all effective actions and ineffective actions under each scoring dimension for each comparison between candidates.

The multiple choice In-Basket can have two advantages over the free response In-Basket in addition to reducing scoring time. The

first is that by forcing the candidate to choose from the alternatives, a clear understanding of the candidate's actions is possible. In the free response In-Basket, it is sometimes necessary to infer and possibly misinterpret a candidate's intended course of action. A second advantage is that a multiple choice In-Basket eliminates the subjectivity and potentially biased scorer's judgment. Taft (1959) contended that there are differences in the assessors' abilities to evaluate personnel. Abilities to evaluate can be influenced by the familiarity with the criterion measures and the problem situations. Although In-Basket scorers have generally yielded satisfactory reliability coefficients in the .20 to .95 range, any procedure which can increase interscorer reliability should be investigated (Brass & Oldham, 1976; Frederiksen et al., 1957). The extensive training received by scorers is one method to increase interscorer reliability. Another method is to reduce a scorer's judgment to a group decision on the value to place on each course of action, as is possible with an objective multiple choice In-Basket Exercise.

There have been some attempts to develop multiple choice In-Baskets. In 1963, as part of the AMA Company Management Course, a 111 item multiple choice questionnaire was developed to cover 20 problems in an In-Basket Exercise (Lopez, 1966, p. 109). The machine scorable questionnaire was designed from the analysis of the responses noted in the "Reasons for Action Form." The "Reasons for Action Form" was originally given to participants after they completed the exercise to gain a better understanding of why, how and what the candidate

intended to do. Careful analysis of the responses revealed that the candidate was faced with an initial decision after reading each item, e.g., respond to a letter, send it to someone, put it aside for later, etc. After the candidate chose a path of action, subsequent sets of actions from which he/she must select confront him/her, e.g., who to sent it to, what to say, etc. The authors of the AMA Company Exercise noted that the structure of sequential decision-making suggested a basic response format by which an In-Basket participant could record his/her actions systematically, a machine scorable multiple choice form. By selecting the alternatives that best describe his/her actions, the participant was actually scoring his/her own responses in a manner not unlike that of a trained scorer. The major advantage, aside from time economy, was that the participant was aware of his/her own intentions, whereas a third party could only infer them from what he read or observed. The participants were also given a form to record "unusual actions" for which there was no suitable response in the multiple choice form, but many of the participants did not utilize it. This first attempt at a multiple choice was described by the authors as "quite promising although not overwhelmingly successful."

The question of the reliability and validity of a multiple choice versus a free response open-ended test has often been debated (Guion, 1965; Ward, Frederiksen, & Carlson, 1980). Proponents of the free response argue that the open-ended test requires the subject to produce an answer rather than to recognize it. Presumably this calls for a somewhat greater depth of knowledge, although Guion states that no

empirical evidence has been found to support it. Proponents also argue that open-ended questions provide the subject with more opportunity to show how well he can analyze a problem, think it through, organize ideas, and follow these ideas through a logical conclusion. Supporters for the multiple choice testing format claim that any ability that is clearly specified can be assessed by procedures in which the correct answer need only be chosen from the alternatives presented. Ward et al. (1980) proposed a study to resolve the question of the validity of a multiple choice test. The authors used an instrument called "Formulating Hypotheses" which consisted of a set of problems not clearly defined, the information needed to solve the problem was not immediately available nor initially apparent, there was no clear criterion for testing a proposed solution, no clearly defined process for applying a criterion and the set of problems had no right or wrong answers but many possible answers with different degrees of quality. The subjects were required to read a brief description of an experiment or field study, to study a graph or table showing the results and to write or choose a hypothesis or possible explanation that could account for the major findings of the study. This "Formulating Hypotheses" test is not very unlike the In-Basket Exercise which is a set of problems not clearly defined, the information to solve the problem must be gathered from various sources and there is no right or wrong answer or a method to test a possible proposed solution. One hundred seventy-four paid volunteers completed a free response test followed by a multiple choice version in which

nine options were presented. The options represented ideas that had been proposed by subjects completing the free response form. The nine options varied as broadly and evenly as was possible along a quality of response scale. The results of the study showed that the highest scoring subjects obtained quality scores very close to the maximum possible on both forms but the worst performances obtained were very much worse on the free response version. A possible explanation given by the authors was that a subject may be able to discriminate a good from a poor alternative even when he was unable to generate a good response. Although differences were found between the forms, the authors concluded that for quality scores the free response quality scores would add little, if anything, to what is measured by the multiple choice test. Given the subjective liabilities and the complicated and time-consuming task of scoring, it may be advantageous to create multiple choice test versions.

This study attempted to test the hypothesis that the scores obtained from subjects taking a free response and a parallel multiple choice version In-Basket Exercise will not be significantly different due to test version. Research results indicate that it is possible to carefully construct a multiple choice In-Basket to evaluate the sequential decision-making role of the manager. A carefully constructed In-Basket which includes all possible options available to a manager when he/she is confronted with a problem can effectively replace the clerical, tedious, complex scorer's task of evaluating administrative, planning, decision-making skills of candidates.

CHAPTER III

METHODOLOGY

Apparatus: In-Basket Exercises

Two In-Basket Exercises were designed, Exercises A and B. Each consisted of an introduction, calendar, organizational chart and 26 items. The items used in each of the exercises included an approximately equal number of items which related or impacted other items within the exercise as well as items which were independent or did not contain information which related to the other items. Each In-Basket Exercise was comprised of two main issues or problems, 15 issues of various degrees of importance/priority and three items of little importance/priority.

The two In-Basket Exercises were constructed to be as parallel as possible without using identical items. The Exercises were constructed to simulate the role of a first level claims manager of a large national insurance company. The Exercises were constructed with the assistance of a panel of five representatives from three insurance companies. The five representatives had a combined 13 years experience at the managerial level with a mean of $2\frac{1}{2}$ years and a range of no managerial experience to 6 years experience and an average of approximately 6 years experience in the insurance field. The Exercises were developed by having the panel cite a list of typical problems/situations which were encountered by a first level claims manager. The

panel was also asked to rate each item on the importance of handling the problem/situation effectively and quickly using a five-point scale. The items were then assigned to two sets with each set receiving an approximate equal number of items from each point along the importance scale. The author and the most senior member of the panel developed two imaginary insurance companies with a corresponding organizational chart and employees. Using the organizational chart and the list of situations/problems, the author and the senior panel member developed letters, memos, notes, etc. The author and the expert were careful to be consistent in the development of each of the In-Basket Exercises. For example, each In-Basket Exercise had an equal number of conflicts, meetings scheduled at the same time, equal number of requests, report deadlines, etc.

The remaining panel members were then asked to review the items, i.e., memos, notes, etc. The panel was instructed to judge the "realness" of the letters, memos and notes and to provide a list of possible actions that a manager could take to respond to each item. Five items were deleted or revised as a result of the panel's input. The panel was then asked to rate each response on a five-point scale (very effective, effective, not measured, ineffective, very ineffective) in each of five dimensions: decisiveness, decision-making, organizing and planning, delegation and perception. The author and the panel members then developed a scoring guide for the free response In-Basket Exercise by using these ratings. Key points were identified for each item of the exercises in each of the scoring

dimensions (dependent variables). For each Key point positive actions/decisions and negative actions/decisions were identified. A list of these Key points and actions/decisions was prepared for each item.

Members of the author's family and friends were asked to take the In-Basket Exercises A and B. These subjects did not have experience as insurance adjusters or claims representatives. Some did have managerial/supervisory experience in other fields. The purpose of this step was to generate more responses to the items. The responses by these subjects were in the majority of cases equal or very similar to the responses which had been generated by the panel of experts. The panel of experts rated those responses which were different along the same five dimensions.

Using the list of responses and Key points, the panel of experts and the author also developed a multiple choice answer sheet for each In-Basket Exercise. The answer sheet was developed with five choices for each item. The five choices were designed to incorporate the full range of actions that were available to the subjects taking the exercises.

Each In-Basket Exercise was prepared with two sets of instructions. One set instructed the subject to read the items and demonstrate in writing how the subject would respond to each item. The subject was directed to be specific and provide details on what actions he/she would take. The second set of instructions directed the subject to read the items, determine how he/she would respond to

each item and then select the one choice in the multiple choice answer sheet which most nearly matched the actions which he/she had decided to take to respond to the item.

The Study Design

The Study Design utilized was a two by two repeated measures design with five dependent variables. The two independent variables were Test Form (Exercises A and B) and Test Format (Multiple Choice and Free Response). Four test groups (I, II, III, and IV) were used with each group receiving two tests as shown in Figure 1 below. The design provided for each group to receive both the Free Response and the Multiple Choice In-Basket Exercise.

		Test 1		Test 2	
		Test Format		Test Format	
		Multiple Choice	Free Response	Multiple Choice	Free Response
Test Form	A	I	II	IV	III
	B	III	IV	II	I

FIGURE 1. STUDY DESIGN--2 X 2 REPEATED MEASURES

The Dependent Variatles were defined as follows:

1. Decisiveness--readiness to make decisions, render judgments, take action or make commitments; not needlessly seeking or waiting for more information.

2. Decision-Making--Making realistic decisions based on logical assumptions which reflect the facts and take situational resources into consideration.

3. Organizing and Planning--Establishing a course of action for self and others to accomplish a specific goal. Making proper assignments to personnel and appropriate allocation of resources. Establishing priorities and time sequences of activities.

4. Delegation--Making work assignments for subordinates. Assigning data gathering and research support to subordinate staff as well as assigning responsibility and authority to others. Providing clear instructions and scheduling follow-up.

5. Perception--Identifying existing or potential problems or opportunities, obtaining relevant information, relating data from different sources and identifying possible causes of problems.

6. Total--Average of the scores of the four dependent variables.

Hypothesis Test

Ho = The independent variable, Test Format (Multiple Choice or Free Response) has no effect on the scores of the subjects.

H1 = The independent variable, Test Format, has an effect on the scores of the subjects.

Subjects

Twelve subjects were used for the study. All 12 subjects were employed as claims adjusters for a large insurance company. Three subjects had an Associates college degree or equivalent, eight had earned Bachelors degrees and one had one year of postbaccalaureate college credit. Groups I, III, and IV had an average of 3.3 years of college credit and Group II averaged 4.3 years.

The subjects' experience in the insurance industry ranged from 5 months to 14 years with an average of 3.4 years. The average per

group was 2.6 years, 2.5 years, 3.2 years and 5.5 years for groups I, II, III, and IV respectively. Group IV average was affected by the one subject who had 14 years of experience.

Procedures

The 12 subjects were randomly assigned to the four study groups, three subjects per group. Each subject was given one In-Basket Exercise, the form and format corresponding to the group to which the subject was assigned.

After completing the In-Basket Exercise, the exercises were scored by three members of the panel of experts who were appointed as assessors. The same three members scored all the exercises. The assessors, as a team reviewed the response of each item of the exercise and identified which Key points from the assessor's guide was included and/or excluded in the subject's response. This generated a list of positive marks (+), negative marks (-), and zero marks (0) for each response in each of the scoring dimensions. Each scoring dimension was then individually scored by reviewing the number of +'s and -'s and 0's each subject had received in the 26 items. The assessors scored the dimension by awarding a rating of 0 to 10 based on the number of +'s, -'s, and 0's received by the subject. The Total Score was determined by the numbered average of the other five scoring dimensions. The multiple choice In-Basket Exercise was also scored by the same assessors. Each choice in the exercise had previously been scored with +'s, -'s and 0's for each scoring dimension. The assessors reviewed the choice the subject had selected for each item

and awarded the +'s, -'s and 0's based on the pre-established scoring key. This resulted in a list of marks for each item in each scoring dimension for each item and the total for each scoring dimension. The assessors awarded a score of 0 to 10 on each scoring dimension based on the same scale/standard used for the free response exercise.

The same 12 subjects received a second In-Basket Exercise after a waiting period. Eight subjects received the second test 14 days later, two subjects 15 days late, one subject 16 days later and the last subject 18 days later. The form and format of the second test was determined by the group to which the subject had been randomly assigned. The second exercises were scored by the same three assessors using the same scoring procedures.

The scores were subsequently analyzed using six analyses of variance for factorial designs with repeated measures.

CHAPTER IV

RESULTS

The study was designed to observe if significant differences occurred in scores obtained by the administration of two formats of equivalent In-Basket Exercises, a free response answer format and a multiple choice answer format. To test the null hypothesis that there was no significant difference in scores due to test format, the analysis of variance with repeated measures was utilized.

Six analysis of variance tests were performed, one for each of the six dependent variables resulting in a total of 42 F values. The results are summarized in Tables 1 through 6. In addition to the test of significance for the main independent variable, test format, F values were available for the second independent variable, test form A and B and for the repeated measures, test 1 and test 2 as well as for the interaction effect of test form and test format, test form and repeated measures, test format and repeated measures and test form, test format and repeated measures. The correlation between the multiple choice and the free response scores for each subject was also performed by applying the Product Moment Correlation Coefficient to the repeated scores. The results are summarized in the far right column of Table 7.

Table 1 summarizes the analysis of variance for the dependent variable, Decisiveness Score. None of the F values were found to be

significant, thereby supporting the null hypothesis that the test format did not have an effect on the scores obtained from the subjects. Additionally, none of the variability of the data was attributed to the test format, as shown by the 0 for the sum of squares for test format in Table 1. The correlation between the multiple choice scores and the free response scores for each subject using the Product Moment correlation coefficient was +.25. Generally, for the Decisiveness Score there was little variability within the data, with the greatest proportion of variance (52%) attributable to the within-cells sum of squares.

TABLE 1

ANALYSIS OF VARIANCE FOR DECISIVENESS SCORE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	1.50	1	1.50	2.77
Format	0.00	1	0.00	0.00
Form x Format	.67	1	.67	1.23
Error Between	4.33	8	.54	
Repeated Measures	.17	1	.17	.11
Form x Repeated	1.50	1	1.50	1.03
Format x Repeated	0.00	1	0.00	0.00
Form x Format x Repeated	2.67	1	2.67	1.83
Error Within	11.67	8	1.45	
Total	22.50	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

The Analysis of Variance for the Decision-Making Score is summarized in Table 2. As with the Decisiveness Score, none of the F values were found to be significant at either the .05 or .01 level of confidence. Approximately 1% of the sum of squares was attributable to Test Format. The Product Moment Correlation Coefficient was +.97. The null hypothesis that test format does not have a significant effect on In-Basket scores was supported.

TABLE 2

ANALYSIS OF VARIANCE FOR DECISION MAKING SCORE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	1.04	1	1.04	1.56
Format	.37	1	.37	.56
Form x Format	3.37	1	3.37	5.06
Error Between	5.33	8	.67	
Repeated Measures	2.04	1	2.04	1.29
Form x Repeated	7.04	1	7.04	4.45
Format x Repeated	.37	1	.37	.24
Form x Format x Repeated	.37	1	.37	.24
Error Within	12.67	8	1.58	
Total	32.62	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

The third dependent variable studied was the Organizing and Planning Score which is summarized in Table 3. One F value was found to be significant at the .05 level but not at the .01 level of

confidence. The significant F value was the interaction effect of Test Form and Test Format. The F value for Test Format was not significant and the Sum of Squares for Test Format accounted for less than 1% of the total Sum of Squares. The correlation bettween the multiple choice and the free response scores was +.46.

TABLE 3

ANALYSIS OF VARIANCE FOR ORGANIZING AND PLANNING SCORE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	3.37	1	3.37	3.11
Format	.04	1	.04	.04
Form x Format	7.04	1	7.04	6.50*
Error Between	8.67	8	1.08	
Repeated Measures	1.04	1	1.04	.50
Form x Repeated	3.37	1	3.37	1.62
Format x Repeated	.04	1	.04	.02
Form x Format x Repeated	.37	1	.37	.18
Error Within	16.67	8	2.08	
Total	40.62	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

*Significant at the .05 level

Table 4 summarizes the Analysis of Variance for the Delegation Score. One of the seven F values was found significant at the .01 level of confidence. Significance due to Test Form was observed. No significant difference was found due to Test Format and the Sum of

Squares for Test Format indicated that only 3% of the total variance was due to Test Format, with 40% of the variance accounted by the sum of squares within cells. None of the interaction effects were found to be significant. The correlation for the repeated measures, multiple choice, and free response was found to be +.06.

TABLE 4

ANALYSIS OF VARIANCE FOR DELEGATION SCORE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	9.37	1	9.37	28.12*
Format	1.04	1	1.04	3.12
Form x Format	.37	1	.37	1.12
Error Between	2.67	8	.33	
Repeated Measures	2.04	1	2.04	1.17
Form x Repeated	5.04	1	5.04	2.88
Format x Repeated	.37	1	.37	.21
Form x Format x Repeated	.04	1	.04	.02
Error Within	13.99	8	1.75	
Total	34.96	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

*Significant at the .01 level

The Analysis of Variance for the Perception Score is summarized in Table 5. None of the seven F values were found to be significant. The null hypothesis that Test Format does not affect scores was retained. Sum of Squares of Test Format includes less than 1% of the

total variability of the data and the correlation was found to be +.29.

TABLE 5

ANALYSIS OF VARIANCE FOR PERCEPTION SCORE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	7.04	1	7.04	3.93
Format	.04	1	.04	.02
Form x Format	7.04	1	7.04	3.93
Error Between	14.33	8	1.79	
Repeated Measures	1.04	1	1.04	1.92
Form x Repeated	2.04	1	2.04	3.77
Format x Repeated	.04	1	.04	.08
Form x Format x Repeated	.04	1	.04	.08
Error Within	4.33	8	.54	
Total	26.77	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

The final dependent variable, the Total score, was the mathematical average of the other five dependent variables. The analyses of variance for Total Score are summarized in Table 6. One F value for Test Form was found to be significant at the .05 level. It was not significant at the .01 level of confidence. The Product Moment Correlation Coefficient obtained was +.67 for the correlation between the multiple choice and the free response scores.

TABLE 6

ANALYSIS OF VARIANCE FOR TOTAL SCORES

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Form	3.84	1	3.84	6.64*
Format	6.67	1	6.67	.01
Form x Format	2.94	1	2.94	5.08
Error Between	4.63	8	.58	
Repeated Measures	1.13	1	1.13	.88
Form x Repeated	3.53	1	3.53	2.77
Format x Repeated	.11	1	.11	.08
Form x Format x Repeated	.43	1	.43	.33
Error Within	10.17	8	1.27	
Total	35.96	23		

Note. Significant F values: 5.32 at .05 level, 11.26 at .01 level

*Significant at the .05 level

Table 7 summarizes the means and standard deviations of the scores for both the free response and the multiple choice exercise for each of the dependent variables. The correlation between the repeated scores using the Product Moment Correlation Coefficient is also summarized in Table 7. The study design provided for each subject to receive a different format of the exercise at the second administration, therefore the correlation of the repeated scores is the correlation of the multiple choice and the free response score for each subject. The test of significance for the correlation

coefficients revealed that only the Total Score and the Decision-Making Score were statistically significant.

TABLE 7

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR TEST FORMAT

	Free Response Exercise		Multiple Choice Exercise		
Variable	Mean	Standard Deviation	Mean	Standard Deviation	Correlation
Decisiveness	7.25	1.14	7.25	0.87	+.25
Decision-Making	6.25	1.42	6.00	0.95	+.97*
Organizing & Planning	5.91	1.62	5.83	1.02	+.46
Delegation	7.08	1.44	7.50	1.00	+.06
Perception	5.75	1.48	5.83	1.03	+.29
Total Score	6.45	1.32	6.48	.84	+.67**

Note. Significant values for the correlation coefficients are .576 at the .05 level and .708 at the .01 level.

*Significant at the .01 level.

**Significant at the .05 level.

CHAPTER V

DISCUSSION

The purpose of the study was to determine if a multiple choice In-Basket Exercise would yield equivalent scores to those obtained through the use of the standard free choice In-Basket Exercise. The free choice In-Basket Exercise has been an integral part of the Assessment Center personnel selection system since 1957 (Frederiksen, Saunders, & Wand, 1957; Lopez, 1966). The In-Basket Exercise has withstood the test of time and has been shown to be reliable and valid in selecting managers and supervisors for all levels of management.

The In-Basket Exercise consists of the presentation of a series of real-life situations in the form of memos, letters and notes to a group of candidates. Each candidate is asked to write how he/she would handle each situation. A panel of assessors then evaluates each response and rates the response along several criteria, i.e., decision-making, organizing and planning, delegation. The ratings when used in connection with other exercises are used to make judgments about the candidate's ability to handle managerial responsibilities.

The process for making these judgments is very complex and requires a tremendous amount of time, as much as four hours per candidate. This can be a difficult obstacle when there are many candidates and limited time. The purpose of this study has been to determine if

an easily scorable In-Basket Exercise could be developed which generates ratings/scores that are not significantly different from those obtained by the traditional free response In-Basket Exercise. If no significant difference is observed, support is provided for the proposition that the multiple-choice In-Basket Exercise can be used in place of the free response form and thereby reduce the time required to evaluate a candidate's performance.

Forty-two F values were calculated from the data generated by the study. This included test of significance for two independent variables, test format and test form, repeated measures and the interaction effect. The significance of the test format was the focus of the study with the repeated measures and the test form needed to add support for the main hypothesis.

The six F values for the independent variable, test format, were not found significant for the six dependent variables, thereby the null hypothesis was retained in each case suggesting that the multiple choice and the free response In-Basket Exercise could produce equivalent scores. The results are not surprising if one examines the process of scoring In-Basket Exercises. In the traditional free response exercise the assessors develop a list of possible actions that can be taken and assign weights or ratings to each of the actions along various dimensions, i.e., decisiveness, delegation, etc. The assessor's task then becomes one of matching the subject's response with the action from the list of alternatives which it most closely resembles. The assessor then assigns that rating or points to the

subject's response (Brass & Oldham, 1976). In the multiple choice exercise, the list of alternative actions is provided to the subject; when the subject selects a response, the points or ratings assigned to that response are given to the subject for that item. The difference between the two processes is that for the multiple choice exercise the assessors do not duplicate the subject's task of selecting a course of action.

In their studies Lopez (1966) and Frederiksen (1962) proposed that a manager's job involves the selection of appropriate responses when presented with an input, problem or task. The responses typically available to the manager are limited and not unique for each problem, hence a multiple choice type process. The manager must analyze the inputs, i.e., a request, a conflict in meetings, deadline, priority, etc., and choose a course of action from the limited set of options available. As with Lopez (1966), the present study demonstrated that when given the free response exercise, the subjects provided actions which were fairly typical. The assessors found only three examples of actions which were not provided as alternatives in the multiple choice format, therefore the options available through the multiple choice exercise were shown to be usually sufficient for the subjects to demonstrate effective managerial abilities as well as weak or ineffective managerial abilities.

Other F values which were reviewed included the repeated measures. None of these values were found to be significant. The repeated measures obtained from each subject included a score from

each of the two test formats. The subjects were used as their own control to minimize error from differing degrees of managerial ability between subjects. To control for carry-over effect, two test forms were used. The two test forms presented different problems, items, requests, etc. However, both test forms were designed to evaluate the same knowledge, skills, and abilities. The finding of non-significance provided further support to the hypothesis that test format did not significantly affect scores.

Three of the 42 F values were found to be significant, the Delegation Score was significant for test form (.01 level), the Total Score was also significant for test form (.05 level) and the Organizing and Planning Score was significant for interaction effect between test form and test format (.05 level). One plausible explanation for the significance is random error. At the .05 level of confidence, it is expected that 2 or 3 of the 42 F values can be found significantly different due to chance not attributable to the test instruments. A review by the assessors of the test instruments for the delegation score did not reveal a source of possible contamination.

A second test of significance which was used to try to establish the equivalency of the two test formats was the statistical significance of the correlation coefficients. The correlation for the repeated measures was determined which by the study design was the correlation between the Free Response and the Multiple Choice scores for each subject. It would be expected that there would be a high

positive statistically significant correlation for each of the dependent variables. The results of the study, however, produced only two statistically significant correlations, the Total Score and the Decision-Making Score.

The lack of statistical significance in the correlations of four of the dependent variables required further evaluation to establish the cause. One of the possible causes is the low number of subjects used in the study and the relatively low variability within the data as shown by the data in Table 7. A second explanation is that although both exercise formats are evaluating the subject's possession or lack of possession of skills and abilities as defined by the dependent variables, each format may be emphasizing a different construct of the dependent variables. The group means are not significantly different but each subject's standing within the group changes as a function of the construct which each exercise taps. The construction of the multiple choice exercise was very time-consuming and complicated and required a delicate balance between providing alternatives and not providing clues to the most correct response which would not be available in the free response exercise. It is possible that this process is sufficient to alter, however slightly, the exercise's measurement of some of the dependent variables. Managerial ability, which is what the In-Basket Exercise is attempting to measure, is a broad concept even when divided into a number of traits or skills.

In summary, although the most important of the dependent variables, the Total Score was found to have a statistically

significant correlation between test formats and to have group means not significantly different due to test format, the evidence was not conclusive that the multiple choice exercise can be used in place of the free response exercise to reduce scoring time. Although differences were observed, there was enough equivalency to proceed with additional studies.

A study with larger subject size is needed as well as a more basic study of the constructs which define the managerial traits, decision-making ability, decisiveness, perception, etc., which are being evaluated. The multiple choice exercise can reduce the scoring time from over four hours per participant to less than one-half hour; however, assurance is required that it is duplicating the free response exercise nearly exactly. A study which examines/compares the steps by which the subject arrives at the decisions/actions is needed before any definitive conclusions can be made. Additionally, a predictive validity study is also needed to determine if both the free response and the multiple choice exercise continue to be successful predictors of managerial performance given any changes required to make the formats equivalent.

REFERENCES

Baritz, L. (1960). The servants of power, a history of the use of social science in American industry. Middletown, CT: Wesleyan University Press.

Brass, D., & Oldham, G. (1976). Validating an In-Basket Test using an alternative set of leadership scoring dimensions. Journal of Applied Psychology, 61, 652-657.

Cohen, S. (1980). Pre-package vs. tailor made: The assessment center debate. Personnel Journal, 59, 989-991.

DuBrin, A. J. (1972). The practice of managerial psychology. New York: Pergamon Press.

Frederiksen, N. (1962). Factors in In-Basket performance. Psychological Monographs, 76(22).

Frederiksen, N., Saunders, D. R., & Wand, B. (1957). Psychological Monographs, 71(9, Whole No. 438).

Guion, R. (1965). Personnel testing. New York: McGraw-Hill.

Lopez, F. (1966). Evaluating executive decision making--The In-Basket technique. New York: American Management Association.

Megginson, L. C. (1967). Personnel: A behavioral approach to administration. Homewood, IL: Richard Irwin.

Meyer, H. (1970). The validity of the In-Basket Test as a measure of performance. Personnel Psychology, 23, 297-307.

Mintzberg, H. (1973). The nature of managerial work. New York: Harper & Row.

Shapira, Z., & Dunbar, R. (1980). Testing Mintzberg's managerial roles classification using an In-Basket simulation. Journal of Applied Psychology, 65(1), 87-95.

Slevin, D. (1972). The assessment center: Breakthrough in management appraisal and development. Personnel Journal, 51, 255-261.

Strauss, G., & Sayles, L. (1972). Personnel: The human problems of management (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Taft, R. (1959). Multiple methods of personality assessment. Psychological Bulletin, 56(5), 333-352.

Ward, W., Frederiksen, N., & Carlson, S. (1980). Construct validity of free-response and machine-scorable forms of a test. Journal of Educational Measurement, 17(1), 11-29.