CHAPTER III
MANAGERIAL EFFECTIVENESS-A RESEARCH REVIEW

Campbell, Dunnette, Lawler and Weick (1970) view the effective manager as an optimizer in utilizing both internal and external resources (human, material and financial) in order to ensure the continuance of the organizational unit for which the manager is responsible. Emphasis in the definition is primarily on managerial actions and behaviour implicit, relevant and essential for appropriate utilization of resources. Rational evaluations therefore require the specification of the wide spectrum of managerial behaviors that are considered to be important for good performance. Beginning with the determination of the domain of a manager's responsibilities, the process continues with simultaneous identification of critical behaviour vital for the appropriate use of existing and potential resources. Suffice it to say that the aim is to derive meaningful expositions of effective managerial performance. Such revelations enable a more comprehensive understanding of the complex inter-relationships existing between various job behaviours and organizational performances or outcomes measured.

The aim of this chapter is to examine the knowledge about managerial effectiveness that has
accumulated over the years through research. The review of literature is confined to large scale empirical research falling within the ambit of global measures, predictive studies, objective personality and interest inventories, leadership ability tests and rater perceptions.

Global Measures

Numerous studies of managerial success have utilized objective, global or administrative criteria (e.g. Laurent 1961, 1962; Mahoney, Jerdey and Nash 1960). Critical examinations have indicated that global measures such as supervisory rankings of composite managerial effectiveness, salary or organizational hierarchical level statistically corrected for age or length of service have several advantages. In cases where supervisory rankings are in operation one superior normally ranks a limited number of subordinate managers. Hence the test-retest and interrater reliabilities tend to be high. Further, such rankings quite often include a large sample of behaviours conditioned over a period of time. Consequently, unintentionally, many a times the manager becomes the subject of judgement rather than the organizational factors. Peer performance is used as standards for comparison of the manager's performance. Such comparison of the manager with his peers is appropriate
since all of them are usually collectively responsible for optimizing the use of similar quantities of resources (Campbell et al., 1970).

Overall measures of effectiveness or ratings of success are factorially complex. Their factorial composition undergoes constant changes as a result of complex interactions among the various variables (Dunnette 1963; Weitz 1961). Quite naturally such measures obscure rather than reveal managerial behaviour for success. The role of intervening variables such as providence, education, training etc. is relatively unknown. Global or administrative criteria explain the position of the manager on the success continuum, but is silent on how it was attained (Campbell et al., 1970).

Behaviours relevant to managerial success changes over time (Korman, 1968). It is imperative that effectiveness dimensions be reviewed to keep pace with organizational changes. Identical is the case of objective measures of managerial effectiveness. The Employee Relations Index (ERI) developed by Merrihue and Katzell (1955) at the General Electric Company is a composite of eight objective indicators of managerial effectiveness, viz., absence rate, separation rate, dispensary visits, number of suggestions submitted, disciplinary suspensions, grievance
rates, work stoppages and extent of employee participation in the company sponsored insurance plan. In early studies, significant relationships were disclosed between the magnitude of the ERI and subjective estimates of work effectiveness. However, they were found to be insignificant in subsequent studies. Consequently, the use of the ERI was discontinued.

Dunnette and Kirchner (1958) administered the Miller Analogies Test (MAT), the Wechsler Adult Intelligence Scale (WAIS), the Strong Vocational Interest Blank (SVIB), the California Psychological Inventory (CPI) and the Edwards Personal Preference Schedule (EPPS) to twenty six sales managers working with the 3M Co. Simultaneously the general manager and vice president using forced choice method selected the better manager out of each combination pair of managers. The global effectiveness score was thus the proportion of times the manager was selected as the more effective of the various pairs. They hypothesized that the effective manager would be more intelligent, have numerous interests and have stronger personalities than other managers. Intelligence was tested by comparing the scores of managers above and below the median on the MAT and those above and below an IQ of 125 on the WAIS. Effectiveness scores of managers with moderate to strong interest in
at least two occupational groups on the Strong Vocational Interest Blank were compared with those of managers showing interest on none or only one occupational group, in order to test hypothesis relevant to interest. To test for strong personality, the manager's profiles on the CPI and the EPPS were examined and divided into high and weak personalities. Effectiveness scores were then compared. Managers having "favourable" reports on two or all three of the variables were compared with those having "unfavourable" reports on at least two variables. Results showed that combining data from separate measures yield better relationship with rated effectiveness than if they are considered separately.

Campbell et al. (1970) observed that objective criteria suffers from both deficiency and contamination. Also these measured only a small portion of the variance due to individual managerial behaviour. Variations in these measures were observed to be dependent on many, but often irrelevant, factors. As these factors are not under the direct control of the manager, it is difficult to monitor them.

In order to overcome the shortcomings of objective criteria, subjective criteria may be resorted to (Cascio, 1978). Such assessment must be based on observations of persons at their jobs. But this practice is often
confronted with problems. Inadequate sampling of the job behaviour domain, lack of knowledge, lack of cooperation by the ratees, differing expectations and perceptions of ratees, changes in the job or job environment are some cases of error and contamination. However, it is now possible to overcome these by scale development methods and training. For a more accurate picture of performance global criteria should preferably be supplemented with systematic observations and recordings of behaviour for better understanding of managerial success (Cascio, 1978).

**Predictive Studies**

The taxonomy, developed by Meehl (1954), makes distinction possible between the various types of instruments used and the types of prediction used in the study of managerial success. Statistical or mechanical predictions assess the individual on some instrument. The scores obtained are subsequently correlated with a criterion measure. Objective personality inventories, biographical data and some ability tests fall into this category. Judgemental or clinical prediction, requires the subjective combination of these scores to forecast criterion status. Observations of behaviour at work and assessment interviews are examples of clinical predictions (Cascio 1978).
Chiselli (1963, 1966), in his review of studies on managerial success, reported that tests of intellectual ability and perceptual accuracy and personality inventories ('r's ranging between 0.25 and 0.30) are best forecasts of managerial success. These results are attributed to the use of criteria differing in reliability, validity and degree of contamination. Concurrent and predictive studies were combined and the validities were averaged. The studies differed on many dimensions such as the relative size of sampling errors, restricted ranges, relative reliabilities of measures used etc. In view of these, the resultant "summarized" validities tended to be underestimates of the actual degree of relationship.

Korman (1968) reviewed only predictive validity studies, solely because of less ambiguity in the interpretation of results. Verbal ability tests, used to measure intelligence, were found to be good predictors of first line supervisory performance, but not performance at the higher levels, because of restriction imposed by the range problem.

Griswely and Jarrett (1973) used a matched group, concurrent-validity design to determine the extent to which mental ability test scores and self description inventory scores obtained during assessment prior to employment
distinguish between top and middle level managers. The matched group design was used to control two moderator variables (age and education) which were posited to be related to test performances and managerial achievements. Out of 437 applicants 100 subjects were selected. Each of the 50 top managers was paired with one of the 50 middle managers considering age and college education. Success criterion was based on the level of managerial responsibility attained prior to assessment. An advantage of the design is that the data was gathered under actual employment conditions. On the contrary, the disadvantages are the large wastage in subjects used and the difficulty in interpretation of results because "matching the variables controlled inevitably makes one of the matched samples non representative of the population from which it was drawn" (Cascio, 1978, p.237). Inspite of this problem, Grimsely and Jarrett chose to compromise representativeness for increased control of these two moderators which affect the relationship between test scores and managerial achievement.

AT & T's management progress study was begun in 1956 to investigate the career development of persons employed by the organization. A number of significant reports of results (Berlew, 1965; Berlew & Hall, 1964, 1966; Bray, 1962, 1964, 1966; Bray and Grant Berlew, Rychlak and Katkovsky, 1965) has been made.
Bray (1962, 1964) undertook the study to trace the careers of managers from the time of their entry into Bell System. The study intended to obtain an appropriate psychological assessment of a number of new young employees. It also attempted to study them periodically throughout their career. Paper and pencil tests, work samples, projective tests, clinical interviews and participation in group problems and leader's group discussions were used as assessment measures. Studies to assess subject's life, work environment, training, salary, traditions and goals of his organization, and the morale of the employees' work group were intended. Follow up of men who had terminated their employment with AT & T together with employees continuing at the Bell System was also undertaken.

Four hundred and twenty two men employed in 6 companies were assessed over several years. Most of these studies related to the nature of judgments made during assessment, and its relationship to later job performances. The initial step in such assessment procedures consisted of identifying characteristics to be assessed. Out of the numerous items identified, a list of twenty five qualities was selected. These variables were studied by techniques like in-basket method, interview, manufacturing problem, group discussion, projective tests, paper and pencil tests,
and questionnaires. Information was also obtained through a personal history questionnaire, an autobiographical essay and seventy Q-sort self descriptive items.

Interviews were conducted by individual staff members. Tests, questionnaires and other exercises were administered in groups. The group consisted of six participants at a time. Two staff members recorded their assessment of each participant and evaluated them separately. The in-basket technique was reviewed by one staff member. One observer for the group exercises evaluated performance of the individuals in these exercises. Ratings of peers and observers was also considered. This was reviewed and an individual report was prepared by a trained psychologist. The paper and pencil tests were also scored. For Q-sort items the employee identified those that were "most" and "least" like him. The five point scale was used to rate each individual on twenty five characteristics viz., organization and planning, decision making, creativity, human relations skills, behaviour flexibility, personal impact, tolerance of uncertainty, resistance to stress, scholastic aptitude, range of interests, inner work standards, primacy of work, oral communication skills, perception of social cues, self-objectivity, energy, realism of expectations, Bell System value orientation, social
objectivity, need for advancement, ability to delay gratification, need for superior approval, need for peer approval, goal flexibility and need for security.

After completing the ratings on all variables the individual's potential as a management person in the Bell System was evaluated by the staff. Projections of an employee's continuance at the Bell System and the likelihood that he would achieve middle management position and conclusions of whether he should be promoted thus was made. These ratings were discussed and reviewed by staff members wherever it was considered necessary.

Results of these studies provided evidence in four areas viz., factors underlying the 25 assessment variables, the relative contribution of these assessment methods and tests to ratings on different assessment factors and their overall predic ion capability, relationships between the staff's overall predictions and later promotions and salarly levels, relationships between early job challenges and performance and later career success.

The rating variables used in the first study differed from those in subsequent assessments. For purposes of factor analysis, the employees were divided into those with college degrees at the time of entry into service
(N=207) and those without degrees (N=148). Correlation matrices were drawn up for each of the groups. Wherry’s (1959) hierarchical method was used for factor analysis. Similar factors were obtained for both samples. General effectiveness, administrative skills, interpersonal skills, control of feeling, intellectual ability, work oriented motivation, passivity and dependency were the eight common factors.

Correlations between various tests and assessment methods and scores on each of the variables identified by the factor analysis revealed that some methods were more helpful to the staff for making judgements of individuals concerned.

Bray and Grant (1966) determined that the behaviour on the in-basket, performance in manufacturing problem and group exercises accounted for fifty percent of the variance in the overall staff predictions for the college sample and thirty per cent of the variance in the non college sample. Mental ability measures accounted for 6 per cent and 12 per cent of variance in the two samples. In July 1965, five companies provided the progress made by the subjects in the study. The management and salary levels when compared with predictions made at the time of assessment showed that 82 per cent and 75 per cent of the
college and non-college educated subjects had been correctly identified as likely to achieve middle management level. 94 per cent were correctly identified as those who would not progress beyond first level of management. Situational techniques were found to predict success to a considerable extent. Cognitive and personality measures had limited scope.

The Management Progress study data was used to study the effects of early job challenge on later managerial effectiveness (Berlew, 1965, Berlew and Hall, 1964, 1966). Forty four college men continuing on their jobs with the company were studied. An estimate of the individuals degree of job challenge during his first year with the company was prepared based on information elicited through interviews with company officials. From this information, Berlew and Hall estimated company expectations for each individual in eighteen behaviour areas. They are technical competence, learning capacity, imagination, persuasiveness, group membership skills, communication skills, supervisory skills, decision making, organizing ability, time-energy commitment, sacrifice of autonomy, sociability, acceptance of company norms, self development, maintenance of public image, loyalty, productivity and initiative.

Expectation for each individual was rated from one
(low) to three (high). Job challenge was measured as the total of all ratings held by the company for each person. If expectations were high it was assumed that challenge was high. Data were collected to determine relationship between success and job challenge. In 1962 a global appraisal of each man's overall performance and potential was combined with an index based on present and starting salaries to constitute a success index. A number of measures of job performance was also included to form a performance index. Analyses revealed that persons judged to have more challenging first jobs were found to accomplish more than those with less challenging first jobs after four or five years in the company. The foregoing results indicate that first year job challenge is reasonably associated with later judgements of success and job performance.

The group exercises and the performance on the in-basket test were found to be the most important determinants of the staff member's assessment on most variables. The predictive validities of the global predictions were moderately high. With the establishment of the personnel assessment program on an experimental basis an increase in proportion of persons with better performance and high potential into first line management positions was observed.

The promotional patterns of one hundred and
eighteen college men hired during 1956 and 1957 and continuing to work at the Prudential Insurance Company was studied by Selover (1962). At the time of hiring, information about the schools attended, course of study, activities undertaken at college, work and military experience, college grades, scores on tests of verbal and arithmetic reasoning ability were obtained. Of these, scores on arithmetic reasoning and vocabulary tests, scholastic achievement, part-time work and extra-curricular participation, leadership abilities and number of jobs held before joining Prudential Insurance Company were found to have a strong impact on advancement. These five measures were combined into a single advancement potential score.

The general applicability of the scores for predicting promotion was experimented on four additional groups. The first group consisted of thirty employees hired between 1948 and 1951, the second eighty five employees hired between 1952 to 1955, the third thirty six employees hired between 1958 and 1959. These employees were continuing with the company in 1961. In addition, the advancement potential score was computed for 48 men who had been asked to resign and 25 men who had been promoted to high level jobs. The phi co-efficients of 0.36, 0.40, 0.35 and 0.65 indicated that the scoring system was moderately related to promotions in all four cases.
The results support the view that scholastic and non-scholastic factors facilitate prediction of career success for managers. Selover (1962) opines that college graduates who were promoted rapidly did better on general ability tests, achieved high academic grades, were active in college, displayed greater leadership abilities in college or military service and joined the company soon after graduation.

Husband (1957) reported results of his study of career success of 368 graduates of Dartmouth College. Large median differences in current income between scholastic and nonscholastic men were observed. Men who were academically good, participated in extracurricular activities, and demonstrated leadership potential were found to be more successful in their careers as measured by promotion and salary levels.

One hundred and ninety six former MBA students of the Stanford University Graduate School of Business were the subjects of the study conducted by Williams and Harrel (1964). They found that the current salary after adjusting for length of years out of school was moderately related to grades received at the school ('r's of 0.13, 0.14 and 0.22). It was also related to overall school accomplishment (r=0.18) and to leadership in campus (r=0.24).
Harrell (1967) observed that the most suitable criteria of job success were Hemphill's Position Concern Questions, present compensation, present job success as perceived by the MBA graduates and the number of hours in the work week. Since inventories such as the Public Opinion Questionnaire, initiating structure and decision making on Ghiselli's Self-Description Inventory, Undergraduate Grade Point Average, Individual Background Survey, Ascendence on the G-Z Temperament Survey, initiative on the Ghiselli Self-Description Inventory and Manic Scale on the MMPI yielded positive results, its use was recommended for admission to Graduate Schools of Business.

Meyer (1963, 1965a, 1965b) studied the utility of test measures when other factors related to promotion are held relatively constant. His findings are based on a series of studies conducted at the Jewel Tea Company in 1959. The purpose of the study was to identify managers for promotion. Each manager was assigned a promotability rating on a six-point rating scale. A follow up study was made in 1965 of one hundred and seventy eight managers assessed during 1959 and 1960. Assessment ratings, biographical factors and test scores were related to promotions during the intervening years. The assessments showed a strong relationship to later promotions. Of the eighty five
persons identified for promotion, sixty seven were actually promoted. Of the ninety three persons not considered suitable for promotion by the assessment only twelve were promoted. Since Meyer anticipated that the assessment ratings would have had an impact on the actual promotion decision, he investigated the role of other factors. Such factors as age, education, and initial management level were found to be moderately related to promotional experience. Each of these factors and the assessment rating were then weighted equally to yield a "promotability probability index", which was found to be strongly related to promotion.

The study highlighted the fact that promotion policies in the Jewel Tea Co. from 1958 to 1965 gave importance to managers who were relatively younger, more educated, had begun their career at lower levels and were working in the buying, merchandising or administrative departments of the company.

The degree of relationship between the various cognitive and personality tests used in the assessment program and promotion was determined. The other influential factors were held relatively constant. The test information for promoted and non-promoted persons with the same or similar scores on the "promotability probability index" was
perused. The promoted managers tended to have higher scores on the general intelligence test than the non-promoted managers. Meyers observed that those who were promoted were better delegators, more flexible, less manipulative of people, not constantly dependent on superior's approval and were more forthright and emotionally well disposed.

Test information was related to various effectiveness measures by Albright (1966). One hundred and eighty three field sales managers working with the American Oil Company were taken as sample for the study. Predictive relationships were based on tests administered to nearly one third of the men in 1955, who were then working as salesmen in the firm. Concurrent relationships, on the other hand, were based on tests administered to managers shortly before effectiveness ratings were obtained. Global performance and promotability ratings made by superiors, and salary growth rate were the measures for effectiveness. Correlations between all test scores taken together and each of the criteria were computed. Computations for the total group and for subgroups designed according to age above and below forty, and according to functional areas were made.

In some instances low or moderate concurrent relationships (ranging between 0.15 and 0.40) were obtained between global performance rating for vocabulary knowledge,
measures of dominance, self sufficiency, decisiveness and SVIB scales. Moderate predictive relationships (ranging between 0.25 and 0.45) was observed for global performance rating for SVIB scales concerned with positive interest in sales, advertising, verbal and business occupation and negative interest in skilled trades and outdoor occupations. Global promotability ratings for intelligence and vocabulary measures, political and business oriented values, personality measures and SVIB pertaining to business verbal and administrative jobs showed low to moderate concurrent and predictive relationships (0.15 to 0.35). Concurrent relationships were greater for men over forty than for those under forty.

Studies were carried out at the North American Aviation (Tenopyr, 1960, 1961a, 1961b, 1961c), to assess the validity of tests such as the Otis Mental Ability Test, supervisory knowledge tests, the Thurstone Test of Mental Alertness etc. for predicting success as measured by salary advancement, promotion, performance rating etc. Correlations obtained were uniformly near zero. The studies showed that none of the tests was capable of assessing supervisory effectiveness.

Later studies conducted at the same organization showed more positive results (Tenopyr 1962, 1963, 1965;
Tenopyr and Ruch, 1965). Tests showed that successful managers were brighter \((r=0.66\) with Otis IQ) and more proficient in mathematics. Verbal comprehension and Leadership Evaluation and Development scale yielded moderate correlations of 0.29 and 0.36 respectively. Possibility of curvilinear relationships were checked out by computing etas between tests and salary criterion. Relationship with verbal comprehension showed no significant change, but the relationship between salary and human relations was curvilinear \((r=0.36; \eta=0.46)\). Utility of examining curvilinear relationships between predictors and criterion measures was thus established.

Albrecht, Glaser and Marks (1964) studied thirty one district marketing managers who soon after promotion were given a sentence completion test and a human relations test, intelligence and critical thinking tests besides participating in an interview. They were also asked to complete a questionnaire on their personal history. Three psychologists ranked the subjects according to predicted effectiveness in four areas viz., forecasting and budgeting effectiveness, sales performance, interpersonal relationships and overall performance. After completion of a year on the job, these managers were ranked on the above factors by two superiors and peers. These three sets of rankings
were then combined to form a composite index for each individual in each area. The psychologists' predictions were then correlated with the composite index scores. Scores on the tests were also correlated with the composite rankings.

The combined information was found to yield better predictions of later effectiveness of these subjects than any test taken individually. This difference was least in case of forecasting and budgeting effectiveness and most for predicting sales performance.

Flanagan and Krug (1964) administered the SCORES battery consisting of 38 variables distinguished into 6 broad areas viz., supervision (S), creativity (C), organization (O), research (R), engineering (E) and salesmanship (S) to sixty engineers working with Lockheed Aircraft Corporation. They were divided into four homogeneous subgroups as equally matched as possible on age, education and company experience. They were differentiated into supervisory and non supervisory categories and rapid or slow promotability. High and low promotion categories scored identically on creativity and engineering. High promotion managers scored higher on supervision, organization, research and salesmanship. Engineers in management jobs differed from those in technical jobs.
Flanagan and Krug were able to assign participants into the four groups based on the individual's SCORES record. The study supports the view that a battery of tests aimed at determining functional area aptitudes and traits will be capable of identifying "general" effectiveness in management and for making placement decisions.

Objective Personality and Interest Inventories

A number of investigators (Guion and Gottier, 1966; Hedlund, 1965; Korman, 1968; Nash, 1965) have obtained negative results on using personality and interest measures in forecasting managerial effectiveness. However Chiselli's review (1966) has been positive. The matched-group design used by Grimsely and Jarrett (1973, 1975), yielded much higher coefficients from several standardized personality measures used. A questionnaire comprising the general activity, sociability and emotional stability scales based on the Guilford-Zimmerman Temperament Survey yielded a multiple R of 0.52. The General Activity scale when combined with the verbal reasoning and numerical ability scales of the Employee Aptitude Survey yielded a multiple R of 0.61. These coefficients denoted concurrent validity. Results were ascribed to the severe restrictions imposed by the matched group experimental design and the fact that the
data were gathered under employment rather than research conditions.

The Early Identification of Management Potential (EIMP) study, is a pioneer in the validation of tests and biographical inventories against measures of managerial effectiveness. The management of the Standard Oil Company of New Jersey (SONJ) supported this research in order to shed light on the measurement of success and identification of persons with potential for success. Results of the study are reported by Laurent (1962).

Four hundred and forty three managers working for SONJ and five of its affiliate companies were the subjects of the study. These subjects were drawn from all the different functional areas. Three measures of managerial effectiveness were chosen - position level, salary history and effectiveness ranking. These measures were then combined to form an overall success index. Factor analytic methods ensured its independence from both age and length of experience in the firm. This measure of relative managerial effectiveness, did not specifically identify or quantify directly, actual managerial behaviour. However, it reasonably differentiated between managers in their utilisation of organizational resources in their careers. Each manager completed a questionnaire about family,
education, vocational planning, finance, hobbies and leisure time activities, health history and social aspects. Also included were measures of verbal ability, inductive reasoning, management judgement, managerial attitudes and personality measures in the Guilford-Zimmerman Temperament Survey. These test scores were correlated with the overall success index. The degree of relationship between the overall success index and items in the biographical survey, management judgement and Guilford-Zimmerman inventories was examined. Double cross validation was resorted to assure stability of relationships discovered. The total sample was divided into two groups of 222 and 221 managers each. Scoring keys developed on each group was cross validated on the other group. Next the test and responses yielding highest and most uniform correlations with the overall success index was combined to yield a single composite test score. The correlation between the overall success and composite scores was 0.70. Four measures emerged as useful for the evaluation of success. They were the Guilford-Zimmerman Survey, the Management Judgement Test and the two measures of intelligence. Evidence from the data was supportive of the qualities identified in the study as those related to general potential for management, regardless of functional, company or job differences within the organization. It was further contended that the special keys
and weighting methods developed in the study could be applied across different companies and even different countries with some but insignificant loss of predictiveness. To that extent the study is regarded as successful in identifying personal or trait correlates of managerial effectiveness.

Another significant large scale study of the personal correlates of managerial effectiveness was done by the Industrial Relations Centre of the University of Minnesota (Mahoney, Jerdee and Carroll, 1963; Mahoney, Jerdee and Nash, 1960; Mahoney, Sorenson, Jerdee and Nash, 1963). Four hundred and fifty two managers from thirteen diverse firms in Minnesota constituted the sample for the study. Each of the participants was asked to complete a battery of tests consisting of the Wonderlic Personnel Test, Empathy Test, Strong Vocational Interest Blank (SVIB), California Psychological Inventory (CPI) and a Biographical questionnaire. In total ninety eight predictors were tried. Managerial effectiveness was measured by a panel of top company officials who ranked the participants on the basis of management competence. The results showed that the ranking scores were independent of manager's age but related to organizational level in some of the companies. The sample was split randomly into two. Scores of the various
tests were studied on the basis of their capacity to distinguish between the seventy five top and bottom ranked managers in one of the groups. The measures that were statistically significant were used to develop a scoring system which was cross validated on the managers in the other sample. Results obtained in the cross-validation groups from applying patterns of predictors and cutting scores showed that 29 per cent and 62 per cent were judged less and more effective, and received scores above the cutting score on predictor combination. 71 per cent and 38 per cent scored below the cutting score and were judged less and more effective respectively. The Wonderlic Personnel Test and dominance measured by California Psychological Inventory were significant. The Strong Vocational Interest Blank showed that effective managers tend to have similar interests as others in business. Their interests were found to differ from those in agriculture and skilled trades. The Biographical questionnaire showed that effective managers were more educated, had better educated spouses who resigned their jobs shortly after marriage, participated in sports, hobbies and were relatively young as compared to the less effective managers. Effective managers were also found to be more intelligent and dominant than the less effective ones.
Kirkpatrick (1961) has described some of the results of the study conducted to determine the biographical correlates of job effectiveness for the executives of the American Chamber of Commerce. He developed a biographical questionnaire based on interviews with eighteen Chamber of Commerce executives. It consisted of all background information such as early history, family background, education, employment history, military service, physical conditions and social information. Six hundred executives of the Chamber of Commerce responded to the questionnaire. Flanagan's (1951) listing of the critical requirements of Air Force Officer's jobs was used to form a list of effective executive behaviours. This was used to group each of the respondents as above average, average or below average in meeting the critical requirements identified by Flanagan. Adopting the "forced" choice technique fifty percent of the subjects were grouped "average" and twenty five percent each were termed "above average" and "below average". Two groups consisting of fifty members each designated as above and below average, and who were matched on age and length of job experience and rated high on effectiveness were chosen. Comparison between the responses of these two groups yielded thirty five items. These were combined and scored to enable prediction of above average performance. The scoring system was again cross-validated.
on two groups of thirty below average and thirty nine above average executives. The point biserial correlation obtained was 0.56.

Kirkpatric (1961) concluded that the successful Chamber of Commerce executive has a middle class socio economic background, was well educated, engaged in many extra curricular activities, emphasised communication skills and entered the Chamber of Commerce relatively early in life. However some of the biographical variables tend to describe success rather than predict it.

In the Sears Roebuck investigations an executive battery of standardized tests was used to predict criteria of managerial success such as promotion rates, nomination of managers with high success potential and high employee attitude scale results. Bentz (1963, 1967) has reported results pertaining to the study. The Sears organizations used four tests namely American Council on Education Test, Guilford-Martin Inventories, Allport-Vernon Survey of Values and Kuder Preference Record to predict executive effectiveness. The tests were correlated against employee morale scores. Analyses were made using discriminant function and multiple regression.

Multiple correlations approaching 0.75 and falling below 0.40 were reported. The results indicate that
objective personality and interest inventories can predict managerial success to a considerable extent. Moreover, "a cluster of psychological characteristics contributes to general executive competence that transcends the boundaries of specialized or non-specialized assignments" (Bentz, p. 118).

Morse and Wagner (1978) developed an instrument to measure and evaluate managerial performance. The process of managerial performance was the focus of the study. A paper and pencil instrument was developed to determine similarities and differences among managerial jobs.

In the first stage of the study, specific behaviour and activities characteristic of managerial work was identified. Based on literature reviews and interviews with six top corporate executives in six different companies, one hundred and six item statements were developed. They were then clustered around nine roles viz., strategic problem solving, resource handling, conflict handling, organizing, information handling, motivating providing for growth and development, coordinating and managing the organization's environment. This instrument was administered to one hundred and fifteen managers. The instrument was modified based on suggestions made by them. This enabled the reduction of items to ninety six, clustered around the nine roles. Four hundred and six managers
completed the modified questionnaire on another manager. The Crowne and Marlowe (1960) Social Desirability Scale was also completed by the participating managers on themselves.

This modified instrument was factor analyzed and a principal component solution was extracted. Six factors with eigen values exceeding 1.00 were obtained. Based on results obtained the instrument was reduced to contain fifty one items. Correlations between item score and total score exceeded 0.40 and correlation of overall score with Crowne and Marlowe scale was 0.11 only.

The final fifty one item instrument was factor analyzed using correlation matrix obtained from the 406 managers. Six factors emerged with eigen values exceeding 1.00. The factors in descending order are - managing the organization's environment and its resources, organizing and coordinating, information handling, providing for growth and development, motivating and conflict handling, and strategic problem solving. Orthogonal rotations of the six factors to the varimax criterion resulted in 56 per cent of the variance being accounted for. Analytic oblique rotations of the six factors was resorted to to improve the structure and test for interrelatedness. Moderate relationships were obtained. Results suggest that inspite of the distinctive character of roles, they interrelate to describe similarities and differences in managers' jobs.
Both content and concurrent validity of the scale was established by the data. Internal reliability was found to be satisfactory. Replication of the study yielded similar results. The 51-item instrument was tried on a new sample of 420 managers. These managers represented a number of functions, levels and experience and were drawn from two organizations. The replication again identified six factors with eigen values exceeding 1.00. Oblique rotations resulted in a factor pattern matrix similar to the original study. Factors were found to be moderately interrelated (high of 0.44 to a low of 0.29).

The replication differed from the original in two aspects. First, the variance accounted for was 53 per cent (i.e. 3 per cent less than the first case). While factor one accounted for 17 per cent, factor six accounted for only 5 per cent of the variance. However in the replication study, factor three accounted for 12 per cent and factor one only 6 per cent of the variance accounted for. Second, item statements loading on a factor remained uniform but the actual factor loadings differed. The two aspects denote that the variance associated with each factor cannot be generalized and will vary according to the situation. Importance of the item in defining a factor also cannot be generalized and depends on the specific situation.
Multiple regression was used on the data from 2 different organizations to test the influence of the situation on effectiveness. In the first case the significant factors associated with managerial effectiveness were information handling, and strategic problem solving, while in the second it was managing the organization's environment and its resources, organizing and coordinating, and motivating and conflict handling. The discrepancy in results indicates that the importance of each item statement in defining the factor loadings varies according to the specific managerial situation. Based on personality, predispositions, managerial style, skills and abilities of managers some particular roles are emphasized while yet others are ignored.

Reviews concurred on the point that there is a great need to focus attention on the various situational and individual difference variables which are likely to affect the predictability of managerial effectiveness. Job behaviours which contribute to managerial success or failure differ across hierarchical levels and organizations. According to evidence obtained by Guion and Gottier (1966), custom made scoring keys validated for specific organizational settings are more useful than standardised personality measures with standard scoring systems. Scales
developed to predict occupational choice (e.g. the Strong or the Kuder) or Psychological Alienation Tests (e.g. the MMPI) are seldom capable of predicting managerial behaviours such as effective planning or delegation (Campbell et al., 1970).

**Leadership Ability Tests**

Measures of leadership ability have proved to be good predictors of managerial success since they are directly relevant to managerial job requirements. Scales designed to measure two major constructs underlying managerial behaviour - consideration and initiating structure have been developed and used in many situations (Fleishman, 1973). In one review of 7 studies which used the Leader Opinion Questionnaire (LOQ), Korman (1966) reported a median value of -0.02 for validity coefficients. However, only 2 of these studies used industrial samples and were predictive in nature. Another review by Korman (1966) of studies using the Leader Behaviour Description Questionnaire (LBDQ), reported higher validities. All these studies were concurrent, and hence it was impossible to determine whether initiating structure and consideration were causes or effects of managerial effectiveness. Description of managerial behaviour by relevant others are more accurate as predictors of managerial success than are
self descriptions. Kerr et al. (1974) have observed that the use of LOQ, LBDQ and SBDQ (Subordinate Behaviour Description Questionnaire) yields significant concurrent relationships between leader behaviour predictors and various criteria. Such enhanced predictability, could be attributed to researchers' efforts to conceptualize and measure situational variables relating to leadership behaviour and infrequent use of averaged data. In such cases analysis is done by the work group rather than by the individual. Frequent use of moderator variable research strategies was observed. A requisite of this is the independent analysis of the individual. On the other hand the causal implications of consideration and initiating structure are unknown. The predictive validity of the scales remains to be proved as yet (Kerr and Schreinsheim, 1974).

The relation between ratings of leadership effectiveness and conflict resolution methods was investigated in a large midwestern manufacturing firm (Bernardin and Alvares, 1976). Hypotheses based upon the managerial grid developed by Blake and Mouton (1964) were tested. The team, the middle-of-the-road and the task oriented supervisor did not prefer the confrontation, compromise and forcing strategies respectively. No relation between managerial orientation and leadership effectiveness
ratings of first line supervisors was found. Supervisors did not seem to differ in their preferences for conflict resolution strategy as a function of managerial orientation. Perceptions of resolution strategies were observed to differ as a function of organizational level and conflict type. Thus the results indicated that the grid placement was a poor predictor of either perceived managerial effectiveness or conflict resolution methods employed. In a rejoinder to these observations Blake and Mouton (1976) had refuted the use of self assessments obtained prior to a seminar to test hypotheses about managerial effectiveness or conflict resolving methods, as the data are unreliable over estimated judgements by the participants.

Rater Perceptions

Since research has repeatedly proved the necessity to consider ratings of different constituencies, the use of appraisal techniques by each of these groups of raters is examined herein. A prime criterion for evaluation of success is that appraisal scales should be used only by people who are conversant with the aims and objectives of the job; people who usually observe the ratee performing on the job; and people who are capable of making reliable and valid estimates (Latham and Wexley, 1981; Thorndike, 1949).
Comparison of peer nomination, ranking, and ratings by Love (1981) yielded reliable and valid results. No bias on account of friendship was observed in these cases.

Based on a study of the accuracy of peer ratings of British managers, Imada (1982) concluded that behaviour-based information often reduces halo error as the rater is able to perceive the multi-dimensionality of performance criteria. Kane and Lawler (1980) have also observed that peer ratings are good measures of performance.

Musford (1983) observed that multiple raters counterbalance observational errors and increase reliability. Peers have realistic contacts amongst each other. Hence, in spite of lack of objective measures for most tasks, peers arrive at valid ratings since they observe and compare the task-relevant abilities of their colleagues. Moreover, these data are not significantly influenced by factors such as friendship, race or sex discriminations. Attention is based on performance alone.

Information about performance in a leadership or managerial role can be useful for peer raters belonging to a sample constituting different behavioural dispositions. The idea has received considerable support from subsequent researchers (Korsan, 1968; Kane and Lawler, 1978). In the
industrial setting too, they have proved to be effective predictors (Mayfield et al., 1980; Roadman 1964; Weitz 1958).

Mayfield et al. (1980) requested three thousand agents from three different Life Insurance Companies to rate each other on work oriented and socially oriented items. Subsequently, 154 agents were promoted as assistant managers, without considering peer ratings. Supervisors rated them six and eighteen months after promotion. The test retest reliability after a year's interval was 0.77. Assistant managers were expressed as letter grades ranging from A to E. The chi-squared value representing the peer rating scores and the percentage of assistant managers who were rated successful at each letter grade was 12.50 (p<0.05). Further, peer ratings made in one group was not valid if the individuals promoted were subsequently transferred to other groups. Peer ratings were made in established groups where new agents were found to be normally included only on the termination of established agents. Where agents are subject to varying degrees of exposure, it is possible that new agents might receive low ratings as their abilities have not been demonstrated as desired. Based on length of service data Mayfield et al. (1980) showed that peer ratings are unlikely to be effective
predictors for individuals with less than six months experience.

Siegel (1982) obtained paired comparison evaluations by prospective peers (N=20) and superiors (N=4) for a group of twenty savings and loan association branch managers. Only sixteen of the peers however provided their evaluations. The superiors were informed that their ratings together with peer ratings would be used for the four promotional decisions they were to make in the near future. The branch managers knew that appointments to higher posts were to be made shortly. However they did not know that they were actually participating in the selection process. Each individual evaluator's comparison yielded a score for every branch manager who was being evaluated.

A month prior to the paired comparison, nineteen of the participants were administered a battery consisting of the Thurstone Test of Mental Alertness (TMA), planning, judgement, comprehension and expression tests belonging to the Flanagan Industrial Tests (FIT), the Strong-Campbell Interest Inventory (SCII), the Survey of Interpersonal Values (SIV), the Supervisory Index (SI) and the Guilford-Zimmerman Temperament Survey (GZ).

Z-score conversions of assessments made by peers and superiors facilitated selection of three incumbents as
the two sets of evaluators seemed to agree alike. In the fourth instance however ratings by peers far exceeded those of the superiors. Since the candidate was found more favourable to the peers who would subsequently become subordinates, the superiors accepted the candidate for promotion.

Correlation between the mean effectiveness scores by peers and supervisors was 0.84. Interrater correlations ranged between 0.31 and 0.86 in case of peer comparison matrix. Interjudge agreement was found to be high, extending to other independent psychological characteristics of the subjects evaluated. The peer evaluations enabled acceptable promotional decisions in this study. Discussion by supervisors, of the differences in the evaluations made by them individually revealed important differences in perceptions about the determinants of managerial effectiveness. Two important views emerged from the study. One attached greater importance to human relations skill, while the other tended to emphasize knowledge about financial matters as a more important component of branch manager effectiveness. These two aspects were considered to be complementary to each other by the group of superiors.

The purpose of the study conducted by Mount (1984) was to compare the psychometric properties of subordinate
ratings of managerial performance to those of supervisor and self ratings. Eighty middle level managers, their immediate superiors (N=80) and their subordinates (N=365), from a high-technology multinational corporation participated in the study.

A questionnaire was administered to the subjects to assess managerial effectiveness. The first section consisted of items pertaining to background information of the respondents. The second contained eight behavioural items (know-how, administration, training, work direction, feedback, motivation, innovation and consideration).

The multitrait-multimethod analysis was used to study convergent validity, discriminant validity, leniency error and halo error. Computation of Duncan's Multiple Range Test revealed that self ratings were the most lenient (M=6.47), followed by supervisor ratings (M=6.27) and subordinate ratings (M=6.02). Differences in mean levels of ratings was significant across rating sources in case of know-how, work direction, feedback, motivation and consideration.

Data showed evidence of convergent validity and halo effect. Discrimination among traits was observed to be large. Results of ANOVA confirmed that differences among
managers are attributable to the rating sources. Discriminant validity was higher than convergent validity but equal to halo effect. The highest convergent validity was between superiors and subordinates (0.24) and the lowest (0.16) between superiors and self ratings.

Interrater agreement was assessed through interclass correlations. Agreement was greater in case of (Training and Motivation) than on (Delegation and Consideration).

Comparison of results obtained in this study with those of others revealed that this study yielded smaller convergent validities and halo effect and greater discriminant validity. But the same could not be said about self ratings. The results also indicate that interrater agreement is greater on some dimensions like training and motivation than on others such as delegation and consideration. The study is perhaps one of the first to analyze psychometric properties of subordinate ratings. Since subordinate ratings are potential sources of performance information its use has been recommended. Use of at least three subordinates' ratings for a single manager has been suggested.

The findings imply that the use of multiple rating sources would ensure inclusion of more useful information in
appraisals. Inspite of possibility of low interrater agreement its use is emphasized because of greater predictability. Since some dimensions are rated more reliably by some rating sources, it is inevitable that they be included in the appraisal process.

Meyers (1980) observed that self ratings were unrealistically high. Thornton (1980) dissuaded its use for administrative decision making. Bauman and Dent (1982) on the other hand observed that the perception that self assessments would be compared with other objective measures increased the accuracy of self ratings. The view was supported by Mabe and West (1982).

The use of self assessments for performance evaluation was recommended by Levine (1980). Locke and Latham (1984) demonstrated its effectiveness in their goal setting theory.

Caldwell and O'Reilly (1982) studied individuals in a large franchise organization. The study demonstrated the implications of self-monitoring. The individuals were found to perform better if they were sensitive to social cues and were able to modify their behaviour to situational requirements.

Tsui and Ohlott, (1988) studied three sources of disagreement in the effectiveness ratings viz., rating
errors, selective perceptions and variations in criteria weight. All three causes provide reasons of varying degrees for the low agreement coefficient. Three hundred and forty four managers of a Fortune 100 company participated in the study. The ratee managers were given six questionnaires and requested to complete a questionnaire and distribute one each to a superior, one peer and one subordinate with whom the ratee worked least well, and one subordinate and peer with whom the ratee interacted frequently. Confidentiality was assured.

Three hypotheses on criterion type and criterion weight found consensus in the effectiveness ratings of superiors, subordinates and peers. Agreement was strongest between superiors and the peers. Agreement among the raters was high on role behaviours and personality traits of the managers as criteria for effectiveness. Role behaviours accounted for largest variance in all three models (R^2 = 0.07, 0.11 and 0.12) for superiors, subordinates and peers respectively. F-values indicated that the criteria for effectiveness did not significantly differ among the three sets of raters. F-statistic was not significant on either the personal or the role behaviour variables. Raters at the same level tended to assign similar criterion weights.

The results suggest that rating error tendencies,
informational differences and criteria differences are possible causes for low interrater agreement. The study emphasizes the need to consider viewpoints of different rating sources to arrive at accurate explanations of managerial effectiveness.

In yet another predictive study, (Kraut, 1975) showed that peer ratings taken under a "real" set could be used among middle and higher level managers to predict later success. Peer ratings were obtained from 156 middle level managers and 83 higher level executives. Two years later it was found that it did not predict performance of the middle manager sample, however 7 of the 13 rating dimensions predicted promotions significantly (median r=0.35). The highest predictive validity coefficient was shown by mobility (r=0.44). On the other hand for the executive grade ten of the peer ratings significantly predicted promotions. The highest validity coefficient in this group was general impressions (r = 0.44). According to Kraut (1975), the use of peer ratings for evaluation of managers may merely formalize a practice in which managers are already involved.

Results of a study on "feedback-seeking" behaviour of three hundred and eighty seven managers, as observed by their superiors, peers, and subordinates are
reported by Ashford and Tsui (1991). It showed that managers' inclination to seek negative feedback, increased their understanding of the evaluation techniques adopted by these feedback sources. While seeking negative feedback tended to positively influence the opinions of all the three categories of raters. As regards overall effectiveness, seeking positive feedback decreased the same.

The authors opine that the willingness of effective managers to seek negative feedback may be a result of their confidence. Similarly strong egos might have prompted them to use direct enquiry while seeking feedback.

Velsor et al. (1993) examined the relationship among self-perception accuracy, self-awareness, gender, and leader effectiveness. Conclusions were based on data provided by three groups of managers (N=648; N=168 and N=79). Results show that discrepancy is a result of differences between self ratings and ratings by others. The finding was consistent across all three samples, and no gender differences were found. Underraters constantly rated themselves lower than the accurate or overraters. They were rated highest by their direct reports. Overraters are rated lowest by their subordinates. However they tended to constantly rate themselves high. The data also suggested that underraters are the most effective managers. They were
also found to be rated highest in self-awareness based on direct reports. Overraters were perceived as the lowest of the three groups in self awareness and effectiveness.

Nilsen and Campbell (1993) studied the differences between the perceptions of subjects and their raters. Discrepancies in perception were observed on the 360-degree feedback instrument. The discrepancy on skill based multiple rating instruments related to discrepancy on personality based multirater instruments. Self perception is contended to be a stable individual difference. Self-other discrepancy was also found to be stable over time.

A salient feature of studies reported here is their emphasis on self, peer, and subordinate ratings. A noteworthy point in peer assessments is that researchers have shifted their attention from stressing reliability and validity to the study of reactions of actual users besides attempting a theoretical understanding of the reasons behind high reliability and validity (Latham, 1986). The concept of self assessment as a complementary aspect of formal appraisals is also gaining momentum. Organizations must replace formal appraisal by superiors alone with multirater techniques so as to ensure more accurate evaluations of managerial performance.