AN ANALYSIS OF PERFORMANCE EVALUATION USING SINGLE AND MULTIPLE ATTRIBUTES IN A SHORT-RUN AND A LONG-RUN SETTING

By

KEVIN DAYNA STOCKS

Bachelor of Science
Brigham Young University
Provo, Utah
1978

Master of Accountancy
Brigham Young University
Provo, Utah
1978

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1981

AN ANALYSIS OF PERFORMANCE EVALUATION USING SINGLE AND MULTIPLE ATTRIBUTES IN A SHORT-RUN AND A LONG-RUN SETTING

Thesis Approved:

Milton F. Usry
Thesis Adviser

J. Hall Turner

Gary B. Carter

William H. Rambo

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

This study is concerned with the decisions made by managerial personnel when evaluated under two different performance evaluation systems. The primary objective is to analyze the difference in the value of a firm which results from the use of a single short-term profitability measure versus a multiple composite measure consisting of both short-term and long-term attributes as the measure by which performance is evaluated. A Multivariate analysis of variance using Hotelling's T^2 and Bonferroni confidence intervals for pairwise comparisons are used in analyzing the results of decisions made in four environmental settings.

The author wishes to express his appreciation to his dissertation chairman, Dr. Milton F. Usry, for his sound advice and assistance throughout this research and to the other members of his committee, Dr. Lanny G. Chasteen, Dr. J. Scott Turner, and Dr. William W. Rambo. A special note of appreciation is given to Dr. Don R. Hansen for his constant encouragement and guidance as an unofficial member of this committee.

Appreciation is also extended to the School of Accounting at Oklahoma State University for the financial support which materially aided in making this study possible.

To my parents, I express my gratitude for their support and faith in me throughout my life. To my wife, Donna, I extend my eternal love for her patience and understanding which made all of this work worthwhile.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND OVERVIEW

Basic to the operation of all business organizations is the practice of evaluating performance. The degree of formality used in performing the evaluation as well as the scope of the evaluation may vary between settings. Likewise, the method employed to facilitate the evaluation and the factor(s) upon which the evaluation is conducted may also vary. Though organizations differ on the various facets of evaluating performance, there remains the need for performance evaluation (PE).

The Need for Performance Evaluation

Performance evaluation has been employed in all types of settings and through many different methods. The objectives that traditional management theory suggests may be accomplished through the use of PE commonly include attainment of organizational goals, examination of the potential for growth and advancement, maintenance of organization effectiveness and prediction of future performance (Miner, 1968; DeNisi and Mitchell, 1978; Beer, Ruh, Dawson and Kavanagh, 1978; Latham, Fay and Saari, 1979).

More recently developed organizational theories continue to indicate a need for an evaluation of performance. Agency theory (Jensen and Meckling, 1976) suggests that in any employer-employee relationship between two individuals, performance evaluation of some

type is necessary to enable the employer to influence the employee to function in the manner desired. This need is based upon the commonly accepted assumption that employees will perform so as to maximize their own self interest (benefit). The purpose of PE as used by the employer is to help align the activities which maximize the employee's self interest with those activities which maximize the employer's self interest.

This alignment of interests is accomplished by making the attainment of rewards by each employee contingent upon performance desired by the employer. The rewards available are both qualitative and quantitative in nature. Qualitative rewards include such things as praise, acceptance, acknowledgment and job security. Quantitative rewards include regular pay, benefits, bonuses, profit sharing revenues and gifts. In many if not most cases, rewards are of both a qualitative and quantitative nature. The "Employee of the Month" award is a tangible (quantitative) reward while it serves to indicate acceptance and praise (qualitative). Position advances include increases in pay and benefits (quantitative) as well as added prestige, job security and acceptance (qualitative).

The employer makes these rewards contingent upon performance by the employee which is of most benefit to the employer (Demski, 1972). Various factors or attributes are used as periodic measures of the extent to which the desired benefit maximization occurs. As with the employee's rewards, these factors are also qualitative and quantitative in nature. Qualitative factors include such things as company reputation, social and community acceptance, employee morale and capability. Quantitative factors include profit, market share, product development

and financial stability. As with the employee's rewards, many if not all of the attributes used by the employer are both qualitative and quantitative in nature. Maintaining a quality product indicates maintenance of the qualitative attribute of company reputation as well as the quantitative attribute of sales volume. Low employee turnover usually indicates good employee morale (qualitative) as well as low training costs and consistent production (quantitative).

Thus, the employer desires performance by the employee which may be measured in terms of both qualitative and quantitative factors and may be encouraged through qualitative and quantitative rewards. The difficulty that has developed over time is in the selection of a PE system that will provide the incentive for employee behavior to maximize the employer's benefits.

Nature of the Problem

Prior to the industrial revolution, almost all employee-employer relationships were on a personal, one to one basis. Workers directly and personally interacted with their masters. Performance evaluations were very informal in structure, but very individualized and frequent and immediate in nature. Rewards ranged from bondage for unacceptable service to personal honor, status and wealth for favorable service.

As industrialization developed, employee-employer relationships moved to a more formal, yet still generally personal basis. The apprentice (employee) served and worked with the master craftsman (employer). The arrangement included a teacher-student as well as an employee-employer relationship. Performance evaluation was made on a continuing basis and in an individual and personal manner. Unacceptable

performance often resulted in dismissal or expulsion from the caste, deprival of physical comforts or monetary payments. Acceptable performance resulted in personal praise and acceptance, monetary rewards, acknowledgment within the caste and often partnership in the business.

As the industrial revoluation progressed, the employee-employer relationship became much more formal and distant. Individuals now became the employees of a firm, not employees of an individual. The personal employee-employer relationship with its individualized performance evaluation system was replaced. Scientific management (Taylor, 1915) introduced the notion that given sufficient monetary incentives, the performance of an individual could be altered to accomplish almost any desired task. In essence, with the departure from the personal employee-employer relationship, scientific management suggested consolidating all qualitative and quantitative rewards available to an employee into the single reward of monetary remuneration. Not only were the various rewards consolidated into monetary terms, but the attributes used to measure performance of employees also tended to be consolidated into one quantitative factor. Industrialization specialized the task required by each employee. Thus, production employees were paid X number of dollars for each time they accomplished activity Y.

Managerial evaluation was also involved in this consolidation. Managers are those individuals who link the production employee with the employer. In small businesses, the employer (owner) often is also the manager who operates the day-to-day affairs of the company. In larger concerns, individuals are hired by the employer to oversee the daily operation. In the consolidation previously indicated, managers typically received monetary remuneration as their reward and the

maximization of profit was used as the performance measure of interest. Many studies have indicated that this type of PE system for managerial personnel has not been as effective as predicted because the PE method employed failed to control for dysfunctional behavior. For example, maximizing profit is typically identified with superior performance. However, equally of interest in most situations is the method by which profit is maximized. Control over the methods employed in maximizing profit is not available under a single measure system (see Ridgeway, 1956; Schneider, 1973; Downey, 1974; Mayfield, 1975; and Atkin and Conlon, 1978).

It is often suggested that a multiple measure performance evaluation system would exhibit more control over dysfunctional behavior. This is accomplished by explicitly incorporating factors which limit dysfunctional behavior into the measure by which performance is evaluated. It is this suggestion which serves as the basis of this study. This study attempts to examine empirically the relative performance of individuals in a profit setting evaluated on a single measure approach versus those individuals evaluated on a multiple measure approach. The results of this empirical testing may be useful in establishing that PE method which will best control dysfunctional behavior.

Scope and Limitations

This study restricts the issue of evaluating managerial performance to a profit making situation. Examination is further restricted to the position of a division manager who is in charge of an investment center. Extension of this work to include not-for-profit and lower level managerial positions is deferred to a subsequent study.

In this examination, two rewards are present for each and every participant. One of the rewards consists of periodic ranking with peers (qualitative) while the other is periodic remuneration in dollars (quantitative). Remuneration is based on the relative ranking position obtained during each period of evaluation. The use of different types of rewards (e.g., plaques, trophies, publicity, praise, etc.) as well as the use of a different basis by which to assign rewards (e.g., as a linear relation to periodic performance, on performance cumulated over the experimental period, etc.) might provide the basis for additional research.

Also of a limiting nature is the use of only two different methods of evaluating performance. One method uses a periodic profit measure of rate of return on initial investment. The other method uses a multiple composite of four performance factors; production efficiency, market effectiveness, profitability and capacity maintenance. Each of the four components of the multiple composite measure are weighted equally. The use of other measures of profitability, (e.g., residual income, periodic net income, etc.) and/or multiple composite measures consisting of alternative components or different weight combinations is also deferred to a subsequent study.

Other limitations are inherent in the organization of this study. In this examination, a simulated business setting is created using a computerized business game. College students are used as subjects and are asked to make simulated managerial decisions within the business game setting. The validity and usefulness of the results and conclusions of this study are based on the degree to which the activities in this study reflect those activities found in an actual business environment.

Summary of Content

Chapter II examines the current accounting literature on the topic of performance evaluation. Studies which examine the use of various performance measures are reviewed. Also examined is the issue of single versus multiple attribute measures as discussed in the accounting literature.

Chapter III builds a theoretical framework for examining PE. Included in this chapter are the theoretical arguments for and against the use of the single attribute of profit as a performance measure and those arguments for and against the use of multiple attributes as performance measures. The chapter concludes by suggesting specific testable hypotheses to be examined by this study.

Chapter IV and V discuss various issues of using multiple attributes in evaluating performance. Chapter IV focuses on the selection of attributes for a multiple measure. Also discussed in this chapter is the topic of scaling the selected criteria. Chapter V discusses the concept of attribute weighting. Various methods of weighting each attribute are discussed on both a practical and theoretical basis. Chapter V also presents a decision rule for use in this study. This rule is explained, illustrated and compared to alternative methods available. Both of these chapters conclude by summarizing the decisions and methods discussed in that chapter as employed in this study.

Chapter VI examines the experimental design and methodology used in testing the hypotheses suggested in Chapter III. The nature of the business simulation constructed, selection of the subjects, administration of the experiment and treatment of the experiment results are discussed in this chapter.

Chapter VII contains an analysis of the study's results. This analysis includes general comments about the actions and attitudes of the subjects as observed during the experiment as well as both graphical and statistical evaluation of experimental results.

Chapter VIII summarizes the data derived from the experiment. Conclusions that may be inferred from the study are stated as well as recommendations for future study.

CHAPTER II

LITERATURE REVIEW

Previous studies on the use of single measure and multiple measure PE systems are examined in this chapter. The major works developing arguments for and against each system are discussed. This discussion will be followed by a summary and an indication of the need for this study.

Performance Measures

Single Measure

In his discussion of the traditional view of a firm, Caplan (1971, p. 17) states "the role of the business manager is to maximize the profits of the firm." Such a basic notion has led to measurement of a manager's success in terms of the amount of profit accumulated. Solomons (1965) in his well known work on measurement and control of divisional performance devotes almost all of his discussion to the form (accounting net income, return on investment or residual income) of profit which should be used for managerial PE.

To be considered as a meaningful measure of performance it must be assumed that maximizing profit incorporates consideration of all of the factors necessary to elicit decisions which promote the prosperity and survival of the firm in the long run (and hence provide maximum benefit to the owner). Whether profit or any other single measure of performance

incorporates all factors for long term survival and prosperity is questioned in the literature. By only employing one criterion on which to measure performance, considerable pressure is placed on the manager to optimize that factor. However, when all effort is focused on the maximization of a single measure, other factors of importance may be ignored. Stated in another way, "Empirical evidence on single measures of performance indicates that employees place emphasis on this single factor and neglect others important in overall results" (Rosen and Schneck, 1967, p. 176).

Ridgeway (1956) in his discussion of dysfunctional consequences of performance measurements cites several examples wherein the use of a single measure of performance proved dysfunctional. In one case, employment interviewers were appraised by the number of interviews conducted. This resulted in the completion of as many interviews as possible with inadequate time devoted to finding jobs for the clients. The organization's primary function of placing clients in jobs was not given primary consideration due to the interviewer's attention to interview numbers.

Many similar results of dysfunctional behavior resulting from the use of a single performance measure have been cited in the literature. Blau (1955) found the imposition of a quota of eight cases per month for investigators in a federal law enforcement agency resulted in the length of the case being given priority rather than the urgency of the case as impartiality would require. Argyris (1952) found that to meet a quota, managers tended to "feed the machines all the easy orders" toward the end of the month rather than finish them in the sequence in which they were received (see also Granick, 1954; Warren, 1966; Steers, 1975; Berliner, 1976; and Vogel, 1981).

Multiple Measures

To compensate for the ineffectiveness of single measures of performance, the use of multiple measures has been recommended. Theory suggests that all desired aspects of performance will thus receive attention and emphasis.

A consensus was reached at a round-table discussion of business and professional people, that although return on investment (profit) is important, additional criteria are essential for an adequate performance appraisal (Newman and Logan, 1955). Solomons (1965) ends his discussion on the use of various profit measures by suggesting that in maximization of the long run objective of firm value, there is "plenty of room" for other measures of performance. Solomons identifies productivity in terms of output per unit of input or input per unit of output and market effectiveness in terms of market position as two potentially useful measures. Drucker (1954) suggested several measures for PE. These measures include: market standing, innovation, productivity, physical and financial resources, profitability, manager performance and development, worker performance and attitude and public responsibility. Bass (1952) carries the multiple measures into the psychological realm by suggesting that measures of the value of the firm to the individual and the value of the firm and its members to society should also be considered.

Multiple measures are intended to focus attention on the many facets of a particular position. Ridgeway (1956) summarizes this position as follows:

> The use of multiple criteria assumes that the individual will commit his or the organization's efforts, attention and resources in greater measure to those activities which promise to contribute the greatest improvement to over-all performance. There must exist a theoretical condition

> under which an additional unit of effort or resources would yield equally desirable results in over-all performance, whether applied to production, quality, research, safety, public relations or any of the other suggested areas (p. 245).

In order for a manager to commit to those activities which "promise to contribute to the greatest improvement in over-all performance," knowledge of the relative importance of each activity is essential. The knowledge of each activity's importance may be transmitted from the owner or individual in charge. In cases where this information is not given, managers and decision makers must impute their own value judgments (see Keeney and Raiffa, 1976, p. 67). This application of personal value judgments may lead to unexpected results.

Granick (1954) cites a study in the Soviet Union where multiple measures of performance have been introduced. In implementing the system, there was no indication of the relative importance of each measure. Subsequent actions by the managers suggest that they assumed meeting production quotas to be of prime importance. However, performance evaluations indicate that other factors were more important than simply meeting the quotas. Some managers were acclaimed for satisfying production quotas while violating labor laws while others were removed from office for violating quality and assortment plans while fulfilling production quotas. These results indicate that meeting production quotas was more important than obeying labor laws, but was not as important as complying with quality and assortment plans. The imputed value judgments of the managers did not coincide with those of their superiors.

Given the knowledge of the importance of each performance measure, there still remains the issue of how to make comparisons of performance. Schmidt and Kaplan (1971) review the two methods for making multiple measure comparisons. The first method involves combining the separate

measures in some way to form a composite measure. This measure then may be interpreted as an indication of the overall "success" or "value-to-the-organization" of each individual. Such a composite score may then serve as a basis of comparison between individuals (see Toops, 1944; Thorndike, 1948; Brogden and Taylor, 1950; and Nagel, 1953).

The alternative to a composite measure is simply a comparison by each individual measure. The argument presented is that measures of different variables should not be combined. Cattell (1957, p. 11) states this reasoning as follows: "Ten men and two bottles of beer cannot be added to give the same total as two men and ten bottles of beer." Proponents of this view argue that if the multiple measures display low positive, zero, or low negative correlations with one another, then they are obviously measuring different variables, and weighting them into a composite measure results in a score so ambiguous as to be uninterpretable (Schmidt and Kaplan, 1971, p. 420). (See also Ghiselli, 1956; Guion, 1961; and Dunnette, 1963.)

In analyzing this issue, Schmidt and Kaplan (1971) suggest that if the goal of the performance measure is "practical and economic", then multiple measures can be, and in fact, at some time must be weighted in a composite measure. If all performance measures are considered to be measures of a single underlying economic construct, then a resulting composite may unambiguously represent the economic construct and would be interpretable as such. However, if the primary purpose of PE is increased understanding, weighting of the multiple measures into a composite is unnecessary.

Though there are numerous works which discuss the use of single and multiple measures of performance, there has been little written on the

comparative performance of the two methods. A review of this material follows.

PE Studies

Stedry (1960) looks at three approaches in reviewing the relationship between performance evaluation information and the behavior of the individual being evaluated. His first approach is a mathematical model of individual behavior in a budget or goal attainment situation. His model assumes the presence of various performance levels and theorizes that the effects of budgeting on an individual's behavior is a function of various factors. These factors include the personality differences of those participating, the incentive structure used by management and the initial state of mind of each individual. The results of the model suggest that an understanding of the motivational structure of the individuals involved is necessary for proper application of a budgetary process.

Stedry's second approach uses a laboratory study designed to indicate the degree of relationship between performance, aspiration levels and externally imposed budgets. Performance is examined using four sample groups, each subdivided further into a "low", "medium", and a "high budget". The "low budget" satisfies the often cited principle of "attainable but not too loose." The "medium budget" is set at a higher level of performance and the "high budget" is "mathematically incapable of attainment over all trials . . . " (Stedry, 1960, p. 69).

Stedry concludes that the budget level imposed in his experiment has a significant effect on performance. He also concludes that the influence of aspiration levels on performance varies depending on

whether aspiration levels are established before or after the budget is received and on the budget level employed.

Stedry's third approach uses a linear programming formulation to examine the effects of having subgroups within an organization (i.e., departments or divisions) attempt to maximize their performance devoid of overall organizational consideration. Stedry concludes from this experiment that having each subunit attempt to concentrate on its own goals will not necessarily lead to the best possible overall organizational performance.

It is important to note that in the introduction to his study, Stedry (1960, p. 2) states, "As it will be treated here, the objective of budget control is to increase long-run profit at the fastest possible rate; or alternatively, at a given output, to reduce costs at a fastest possible rate." In other words, satisfying the budget in a period is the same concept as maximizing profit during the period. The use of one budgeted factor as seen in all three of Stedry's approaches is tandamount to maximizing a single measure of performance. Thus, the conclusions drawn from these approaches may be applied to the maximization of the single performance measure of profit.

Charnes and Stedry (1964) examine three models of managerial behavior in relation to budget systems. The first model considers a supervisor with one goal, typically a level of profit attainment. Rewards to the supervisor are directly associated with attainment of this goal and it is assumed that this relationship will aid in the internalization of the budget by the supervisor.

From the analysis, Charnes and Stedry conclude that budget changes which affect performance such that aspirations increase rather than

decrease or remain unchanged result in higher performance. The implication is that budgets should be based on trends and improvements rather than on a fixed level of attainment.

The second model deals with the behavior of a supervisor faced with multiple budgeted tasks and evaluated under a type of "management by exception" reward system. The process by which effort is allocated to the various tasks is dependent on the size of the difference between actual and aspired performance levels and the rewards attached to each budget. Charnes and Stedry conclude that the allocation of effort by the subordinate may be quite different from the allocation embedded in the goals of the superior. Charnes and Stedry indicate that three options are available to the superior. The superior may (1) use a different reward structure, (2) use a different stated goal or set of goals from that desired, or (3) attempt to train supervisors to behave in a nonself-optimizing manner.

The third model employed by Charnes and Stedry is developed from the theory of search and allows for diminishing returns to scale with increasing effort. This model addresses the problem of how to increase effort to a particular area of a multiple budget task setting. Charnes and Stedry conclude from this model that increasing the difficulty of reaching budget in an area can, up to a point, improve performance; but after a critical level is reached, performance decreases sharply.

In a subsequent study, Charnes and Stedry (1966) discuss the same issues under four motivational assumptions. One of the assumptions considers the maximization of expected reward where the reward is proportional to expected profit. The other assumptions relate to attainment, nonattainment and over attainment of a budgeted level of performance in a multiple budget task setting. For the second set of

assumptions, the problem is stated "in terms of designing a set of goals which will produce the same effort allocations--and performance--as profit maximization" (Charnes and Stedry, 1966, p. 475).

Charnes and Stedry develop heuristic approximations to some of the optimization procedures. They conclude by indicating that the heuristic approximation models developed may be useful for providing predictions of individual behavior.

Demski (1970) discusses the implementation effects of alternative performance evaluation methods. A simulation of the production smoothing decision is constructed in a single product firm. It is assumed that the decision maker will attempt to implement a given production schedule in a manner that will optimize his performance measurement. The decision maker is evaluated on a profitability indicator calculated under direct, absorption and an ex post costing system. Results of the simulation suggest that varying the evaluation method could significantly alter the decision maker's activities and thus the profit generated by the firm.

Demski (1971) essentially reworks the 1970 study using a multi-product firm. As in the single product firm, the decision maker is evaluated on a profitability measure calculated under various costing methods. A simulation is constructed of a firm with three main products, each with subassembly alternatives. Also included in this simulation are advertising decisions for each of the three products. The results indicate that altering the performance measuring methods significantly affect the activities and performance of the decision makers.

Both of Demski's simulation experiments assume that decision makers preceive their rewards as a linear function of the profits contributed by their subunit. The performance measures used in both experiments

simply consist of profit calculated under different accounting conventions. Thus, these two studies indicate that varying the method by which profit is calculated affects the activities and performance of the decision makers.

Swieringa and Demski (1971) use a simulation study to examine the effects of selected performance measures and incentive alternatives on the behavior of the decision maker. In their simulation, a linear programming model is used to centrally determine all marketing and production decisions which are to be implemented by the supervisors. Behavior of the supervisors is assumed to be affected by the measures used to evaluate their performance as well as the incentive alternatives engaged. Four performance measures (i.e., direct, absorption, ex post, and a combined costing system) under two incentive reward systems (i.e., linear function of reported profit and linear function of reported profit only outside of an established range) are employed.

Swieringa and Demski (1971) conclude that:

> . . . production supervisor's implementation behavior resulted in significantly higher profit levels under each of the two incentive system alternatives when his performance measure and implementation activities were not limited to variations within the production departments only (p. 441).

Also noted is that reported profit levels were sensitive to restrictions placed on the scope of the supervisor's information seeking and implementation activites.

Irish (1970) endeavored to address how the measurement of the division manager's performance can be best expressed in terms of accounting data. Irish interviewed 35 executives (mostly controllers) at various management levels and from various Fortune 1000 firms. Those interviewed strongly agreed that profit as a measurement criterion, or return on investment cannot be used exclusively as they do not gauge

how well the manager is fulfilling long-term objectives. They conclude that the operating performance of the manager should be measured with respect to all divisional objectives.

Summary

A review of the literature presents considerable discussion on the appropriateness of a single, multiple or multiple composite measures of performance evaluation. Profit has traditionally been identified as the measure used in a single measure PE system. It is suggested by many that a single measure of profitability in any of its many forms may not be the appropriate measure by which owners may influence beneficial managerial behavior. Though the idea of using multiple measures has been discussed and debated for some time, little if any empirical data have been gathered comparing the results of using a single as opposed to a multiple measure PE system. This study addresses that issue.

CHAPTER III

A THEORETICAL FRAMEWORK FOR PERFORMANCE EVALUATION

In describing the role of management accounting in influencing management behavior, Caplan (1971, p. 3) suggests the following tasks: the setting of goals; informing individuals what they must do to contribute to the accomplishment of these goals; motivating desirable performance; evaluating performance; and suggesting when corrective actions must be taken. Essentially, Caplan is indicating the intregal part that management accounting has in the evaluation of managerial performance.

Recognition of the role of the management accountant in PE leads to the need for a theoretical basis upon which to perform the evaluation. The basic point of interest is not if a PE system should be used, but which of the many alternative types of PE systems should be selected and how may it be employed so as to encourage those being evaluated to conform to the objectives of the evaluation.

The theoretical background for employing a PE system is discussed in the following sections. Discussion based on the theoretical development is then centered on the appropriateness of short term profit as a single measure of PE versus use of multiple PE measures. A statement of testable hypotheses suggested by the discussion follows.

Theoretical Background

Consider initially an organization where the owner and manager are the same person in a single period environment. In this setting, Demski (1972) concludes that performance evaluation issues are not present. That is, if an owner-manager, after acquiring information, selects what is felt to be an optimal alternative, there is no reason for retrospective analysis of this decision by that same individual.[1] Only if the owner functions in a multiperiod environment or delegates choice of action to another individual or group of individuals does a retrospective analysis of the chosen decision become meaningful.

In a multiperiod environment, retrospective analysis provides information concerning results of decisions made by the owner during the past period. This information may then be of use to the owner in making decisions in future periods. When choice selection is delegated by the owner to another individual, the purpose of retrospective analysis changes substantially.

Assume the owner delegates choice of action to a second individual (i.e., the manager). Retrospective analysis (PE) allows the examination of the manager's choice selection in light of the preferences of the owner. Thus, one purpose of PE is to motivate the manager to behave in a manner deemed optimal by the owner (Demski, 1972; Jensen and Meckling, 1976).

---

[1] Such an evaluation would consist of analyzing whether a decision one makes given a set of information is the same decision the same individual would have made under the same circumstances and given the same information.

The above concepts can be stated formally.[2] Assume that both the owner and manager are rational in the sense of satisfying a set of utility axioms.[3] Thus, if the owner and manager are the same person, the owner faces the problem of selecting an alternative from among a set of alternatives such that the selection maximizes the owner's expected utility (Demski, 1972; Ross, 1973).

In the case where the owner and manager are not the same person, the problem the owner faces changes from choosing the optimal alternative to the selection of a PE system. That is, the manager is now delegated the responsibility of choosing among alternatives while the owner influences the manager's decision by selecting the PE system employed to evaluate the manager's performance. The PE system consists of two components, the performance measure and the incentive function. These two components will be selected so as to maximize the expected utility of the owner.[4] Hence the owner's problem can be formulated as:

$$E(U_0|g^*,m^*) = \max_{g(\cdot),m(\cdot)} \int_s U_0[p(s,\alpha(\cdot), g, m) - g(m(\cdot), s)]\, \Phi_0(s) \qquad (1)$$

where U_0 = the owner's utility function,

$g(\cdot)$ = a specific incentive function dependent on the actual performance statistic,

---

[2] For a detailed discussion see Demski (1972).

[3] Different sets of axioms that imply the existence of utilities with the property that expected utility is an appropriate guide for consistent decision making are presented in von Neumann and Morgenstern (1947), Savage (1954), Luce and Raiffa (1957), Pratt, Raiffa and Schlaifer (1965), and Fishburn (1970).

[4] The manager's problem is to select the action which will maximize the expected utility of the manager given the performance measure and incentive function.

$m(\cdot)$ = a specific performance measurement method,

α = alternative selected by the manager given $g(\ )$ and $m(\ )$,

s = state of nature,

$\Phi_0(s)$ = owner's subjective probability distribution over the states of nature, and

$p(\cdot)$ = outcome measure expressed in monetary terms (such as profit).

Several comments should be made about the above model. First, it is a single period model. Next (g^*, m^*) represent the optimal incentive-measurement pair and depend on both the manager and the decision problem faced. Also notice that the monetary return to the owner is the actual monetary return $p(\cdot)$ less the financial compensation to the manager $g(\cdot)$.

Extension of this model into multiple periods requires no alteration if the results from one decision have absolutely no influence on future decisions. In the situation in which prior choices and outcomes may influence future choices and outcomes, slight modification of the single period model is necessary.

An index "t" will be used to denote the period where $t = 1, \ldots, T$ is treated as a strictly ordered sequence of known decisions that must be made (e.g., sequential periodic decisions). Outcome and probability functions as well as the action and state sets are labeled with this period index. It is assumed that decisions or period are linked by common knowledge of past actions $\bar{a}^t = (a^{-1}, a^1, a^2, \ldots, a^t)$ and performance evaluation information $\bar{r}^t = (r^{-1}, r^1, r^2, \ldots, r^t)$. Also assumed is an additive multiperiod utility function for both the owner and manager.

Essentially what is developed is an adaption of the single period model which requires recognition of the effect of prior actions and

outcomes on future period outcomes. Notationally the owner's problem can now be formulated as:

$$E(U_0|g^*,m^*) \equiv \max_{g(\cdot),m(\cdot)} \int_{s^t} \{U_0^t(p^t(s^t,\alpha^t(\cdot), \bar{a}^t, g, m) - g(m(\cdot), s^t)) + f_0^{t+1}(\bar{r}^t, m(\cdot)), (\bar{a}^t, a^t(\cdot)), g, m)\} \quad (2)$$

$$\Phi_0(s^t/\bar{r}^t) \qquad t = 1, \ldots, T$$

Short Run Profit as a Performance Standard

Single Period Framework

Examining the use of short-run profit maximization as the performance standard in light of the above models provides useful insight. Short-run profit maximization assumes that the planning horizon extends for a single period (Williamson, 1964). Thus short-run profit maximization has a single period perspective similar to equation (1). However, profit maximization is not necessarily equivalent to maximization of expected utility of the owner. For example, let m be a profit measure and let g = f(m) be an incentive function (usually linear) with the property $f'(m) > 0$. Assume the owner is risk neutral and the manager is risk adverse. Under these conditions it is conceivable that the manager would reject projects that the owner would accept.[5] Generally, whenever the risk preferences of the owner and manager differ, there may be situations in which the actions of the manager adversely affect the interests of the owner.

---

[5] See Demski (1972) for an example.

Consider the situation where both the manager and owner are risk neutral. Here the certainty equivalent of both the manager and owner will be equal to the expected monetary value.[6] Furthermore, maximization of expected utility for each is equivalent to wealth maximization. Thus, whenever linear approximation to utility functions over the feasible range is assumed, short-run profit maximization may be adequate as a performance standard in a single period environment.

Certainly, in a pragmatic sense some operational standard is needed. Utility maximization is properly viewed as the objective of the owner and manager. However, a workable proxy objective very well may be wealth maximization. For example, even if both manager and owner are risk averse (as is the most likely case) the assumption of risk neutrality and the resulting wealth maximization objective may serve as a reasonable operational approximization. Moreover, in a single period environment profit maximization is essentially equivalent to wealth maximization. In this study maximization of wealth is treated as maximization of the owner's utility.

Wealth maximization is defined in the traditional economic sense of "well-offness". With regard to wealth an owner derives from a business enterprise, well-offness is a function of cash flows produced during the period as well as the expectation of future cash flows resulting from the terminal position of the firm (see Solomons, 1965, p. 277). In a single period environment, no future cash flows are

---

[6]The certainty equivalent of a lottery is defined as a dollar amount, CE, such that the decision maker is indifferent between playing the lottery or accepting the amount, CE, for certain. Formally, $U(CE) = E\,U(X)$ where X represents the uncertain amount obtainable from the lottery.

expected and thus the wealth of the firm is simply the cash inflow during the period (profit).

Multiple Period Framework

Unfortunately, financial decisions are intrinsically multiperiod problems. For a multiperiod environment, one may retain the proxy objective of wealth maximization. However, short-run profit maximization is no longer necessarily equivalent to wealth maximization. To illustrate the difficulty, consider the definition of objectives and attributes given by Keeney and Raiffa (1976). An objective indicates the direction in which one should move while an attribute is an identifiabile measure which indicates the degree to which alternative activities meet the objective.

In terms of using profitability as a surrogate for wealth in an owner-manager relationship, the objective of interest is maximization of wealth while the attribute by which this objective is measured typically has been short-run profitability. Whether or not profitability should be the objective by which wealth is measured now becomes the point of interest.

Keeney and Raiffa (1976) list five desirable properties for any set of attributes:

1. Completeness. Knowledge of the level of the attributes provides a clear understanding of the extent the associated objective is achieved.

2. Operational. Attributes are operational if they are meaningful to the owner and manager and facilitate explanation to others.

3. Decomposable. Multiattribute utility assessment can be broken into parts of smaller dimensionality.

4. Non Redundancy. The attributes should be defined to avoid double counting.

5. Minimal. The set should be only large enough to ensure completeness.

In view of the properties suggested for a set of attributes, causal evidence exists suggesting that use of the single attribute of short-run profitability violates the property of completeness. Matz and Usry (1976) summarize this point as follows:

> A single measure of performance (e.g., return on capital employed) may result in a fixation on improving the components of the one measure to the neglect of needed attention to other desirable activities--both short- and long-run (p. 880). [also see Newman and Logan, 1955; Warren, 1966]

Numerous examples of dysfunctional behavior in an environment using short-run profitability as the single standard of performance evaluation may be found in the literature.

Granick (1954) cites a study in the Soviet Union in which the single criterion of profitability was rejected because its use led to a reduction in experimental work and de-emphasized the importance of production quantity, quality, and assortment. Later reforms shifted emphasis away from this single criterion.

Warren (1966) cites the following account of a division manager in a large chemical company:

> He pointed out that his sales manager, on one occasion, has laid off three missionary salesmen to reduce costs for the last quarter of a given year so as to meet his budgeted profit goals. He knows that replacements for these three men would be needed in the spring as business increased, and it was estimated that the cost of finding, screening, and training new men was considerably higher than the payroll cost involved in keeping the three men. Although in principle he disapproved of the action, the division manager permitted it for the following reasons: (1) the year's profit objective would be met; (2) the firing and subsequent hiring would not be revealed in the summary data he presented to top management; and (3) because the profit goal had been met there would be no problem getting approval for sufficient funds in next year's budget to rehire three salesmen.
>
> The division manager commented that he was competing with more than 16 other divisions for recognition and funds and

> based on his experience his division would, he felt, be better off in the 'long run' if he showed a solid profit picture for the year. Such a picture, in his opinion, was the only way to get the funds he required to 'continue the work needed to build the long-run success of my division.'
>
> A subsequent analysis of the division's performance, however, showed a history of repeated short-run oriented behavior, such as that noted above. The more support and funds this manager received, the more he felt the need to do an even better job next year, more often than not, at the expense of the future. How long he could have kept on 'mortgaging the future' became an academic question for him as his 'excellent results' led to his promotion. His successor was the man who had to 'face the music' (p. 65).

Vogel (1981) indicates that the same situation continues presently:

> The typical Chief Executive Officer now holds office for an average of five years, compared to 10 years a generation ago. Since Chief Executive Officers tend to be judged and judge themselves by the profits reported while they are in charge, they are understandably reluctant to pursue long-term projects that depress current earnings and won't pay off until after they retire. Executives frequently criticize politicians for making decisions on the basis of short-term considerations in order to ensure their reelection. But the same indictment applies even more to managers in the private sector. Elected officials have a time horizon of at least two years, but executive promotions within companies often are based on earnings calculated on an annual basis. Akio Morita, the chairman of Sony, recently observed: 'The annual bonus some American executives receive depends on annual profit, and the executive who knows his firm's production facilities should be modernized isn't likely to make a decision to invest in new equipment if his own income and managerial ability are judged based only on annual profit . . . I have heard many American managers say, ''Why should I sacrifice my profit for my successor?'' (p. 22).

Though Vogel is speaking specifically of the Chief Executive Officer, similar behavior may be seen in all levels of management.

The above examples clearly indicate that the use of the single criterion of short-run profitability can lead to dysfunctional behavior. A likely hypothesis for this result is that short-run profitability is an incomplete attribute with respect to the owner's objective of wealth maximization. Short-run profitability appears to

be an incomplete attribute because it encourages behavior beneficial to the manager in the short run but detrimental to the owner in the long run.

This behavior is possible because managers who do not expect to remain in a position for more than a short period of time may be able to enhance their short-run performance--and hence their own wealth--by avoiding long-run considerations. Neglecting long-run considerations will ultimately resuilt in detrimental effects to the company. However, as Warren points out above, due to the time lag necessary for these effects to be felt, the successor will suffer the consequences and not the present manager.

In recognition of this very problem, General Electric adopted a multiple attribute emphasis. They identified eight "key result areas", upon which performance evaluation depended (Solomons, 1965; Caplan, 1971; Anthony and Dearden, 1980). These key areas were: profitability; market position; productivity; product leadership; personnel development; employee attitudes; public responsibility; balance between short-range and long-range goals. Even though each of the first seven key areas had both short-range and long-range goals, the eighth area--balance between short-range and long-range goals had been specifically identified to make sure that the long-range health of the company would not be sacrificed for short-term gains (Anthony and Dearden, 1980, p. 109). Formally, this is equivalent to identifying a set of attributes which is complete with respect to wealth maximization.

Employment of a set of attributes to analyze performance with respect to wealth maximization involves the issue of multiple or composite PE measures. As summarized in Chapter II, when the goal of the analysis is practical and economic, use of a composite measure is

appropriate. As the purpose of the PE measures used in this study is to rank and reward performance, a composite measure is used in all multiple attribute cases.[7]

Statement of Hypotheses

The above analysis suggests two situations of interest. First, assume that a manager enters a long term appointment. Given the knowledge of the long-run nature of the appointment, it may be suggested that maximization of personal welfare over the term of the appointment would be the objective of the manager. That is, the manager will act in a long-run manner so as to maintain over the long-run period of employment the evaluation measure of short-run profitability. Thus, one could speculate that the use of the single short term PE system with profitability as a standard should result in no difference in managerial behavior (and hence owner welfare) than will the use of a multiple PE system which contains both short-run and long-run attributes. For a long term appointment, it can be argued that the best interests of the manager are served by exhibiting long-run behavior. Thus, it would not matter whether a single measure or a multiple measure PE system is used. This situation leads to the first hypothesis of interest.

---

[7] If multiple attribute analysis is used, then a vector of attribute measure scores for each case instead of a single composite score must be analyzed. Vector analysis will result in a division of the cases into two groups, dominated and nondominated. The vectors in the dominated group are inferior in all attribute scores to at least one other vector. Those vectors in the nondominated group each have at least one element score of the vector superior to the corresponding element in each of the other nondominated vectors. Beyond this two group division, further ranking is simply infeasible without combining the multiple measures into a single composite measure through the use of some decision rule.

Hypothesis I: If a manager anticipates a long-run appointment, there will be no difference in the welfare of the owner resulting from the use of a PE system using the single attribute of short-run profitability or a PE system using a composite of short-run and long-run attributes.

Next, assume that a manager enters a short term appointment. If the manager knows that performance will be evaluated only over a short-run period, it may be suggested that the method by which the manager is evaluated will significantly affect behavior. Specifically, if the manager in this setting is evaluated through the use of a short-run profitability measure, it would be to the manager's self interest to maximize this short-run standard to the exclusion, indeed complete sacrifice of all long-run considerations. Such behavior would maximize short-run profit, but as the effect of sacrificing long-run variables began to be felt, the overall posture of the firm (and hence the welfare of the owner) would be adversely affected. Such was the situation in the case cited previously by Warren (1966).

If, in place of a short-run profitability measure of performance, a multiple composite measure consisting of both short and long term attributes is used, the behavior of the manager may shift from a short-run orientation to a more balanced short- and long-run orientation. This is because the self interest of the manager is best served by considering long-run variables since these variables are included in calculating the manager's reward. Thus, if the manager knows that performance and hence reward are based on both long-run and short-run attributes, then attention will be devoted to both types of variables. Such behavior would eliminate many of the adverse effects resulting from single emphasis on short-run measure and as such would improve the posture of the firm and hence the owner's welfare. This leads to the second hypothesis of interest.

Hypothesis II: If a manager anticipates a short term appointment, then the welfare of the owner will be greater if a multiple composite measure PE system consisting of both short and long term attributes is used than if a PE system using a single short-run profitability measure of performance is employed.

In each of the hypotheses stated, the length of appointment anticipated is held constant and performance results under two different measurement systems compared. Building on the logic used to develop Hypothesis I and Hypothesis II, two additional hypotheses of interest may be examined by holding the measurement systems constant and comparing performance results under two different tenure anticipations.

Hypothesis III: If a PE system using a single short-run profitability measure of performance is used, the welfare of the owner will be greater if a long-run appointment is anticipated than if a short-run appointment is anticipated.

Hypothesis IV: If a PE system using a composite of short-run and long-run attributes is used, there will be no difference in the welfare of the owner resulting from anticipation of a long-run or a short-run appointment.

Summary

This chapter has presented a theoretical development of PE. This development suggests that for a single period where the owner and manager are the same individual, there exists no need for PE. For an owner-manager in a multiple period setting, retrospective analysis may enhance future decision making. When the owner and manager are different individuals, PE allows for an examination of the manager's choice selection in light of the preferences of the owner. This setting is the focus of this study. In this situation, the owner is faced with choosing the PE system which will encourage the manager to choose those alternatives most beneficial to the owner. This discussion is examined in both a single and a multiple period setting.

The issue of short-run profit as the single measure by which performance is evaluated is discussed. Maximization of wealth is suggested as a suitable proxy for maximization of owner's benefit (utility). Wealth is defined in the traditional economic sense of "well-offness" and in a single period setting, maximization of short-run profit is deemed to be equivalent to maximization of wealth.

When the analysis moves to a multiple period setting, short-run profit is found to be an incomplete measure of wealth because it fails to exact attention on long run economic factors. Several illustrations of dysfunctional behavior resulting from the use of a single measure of performance are cited.

As a result of this discussion, four testable hypotheses are developed. The first suggests that in a long term appointment, performance (decision) should not be significantly affected by the use of the two different performance measures. The second suggests that in a short term appointment, performance is expected to be more beneficial to the owner of the firm under the use of a composite short-run and long-run performance measure than under a single short-run profit measure. The third suggests that if a single short-run profitability PE measure is employed, anticipation of a long-run appointment will result in more benefit to the owner than will anticipation of a short-run appointment. The fourth suggests that if a composite short-run and long-run performance measure is employed, anticipated length of tenure should not affect performance.

CHAPTER IV

CHOOSING AND SCALING ATTRIBUTES

The use of multiple measure models requires four basic activities.[1] The first of these entails the selection of attributes (identifiable measures) which indicate the degree to which alternative cources of action meet the desired objective. The second activity then scales those attributes selected to transform single physical measures into measures of value or suitability. The third activity assigns importance in the form of weights to the multiple attributes. The fourth activity is the selection of a decision rule to combine attributes and rank alternatives in one step.

A discussion of choosing and scaling attributes for a multiple measure model is found in this chapter, while the remaining two activities are discussed in Chapter V. General procedures for choosing attributes are discussed, followed by an examination of the attributes chosen for use in this study. A general discussion on scaling reviews methods available and presents an explanation of the various value functions that can be developed. Discussion then focuses on the specific scaling procedures used in this study.

---

[1]The discussion in this chapter generally follows the discussion found in Hobbs, B. F., "Analytical Multiobjective Decision Methods for Power Plant Siting: A Review of Theory and Application," Division of Regional Studies, National Center for the Analysis of Energy Systems, Brookhaven National Laboratory, Upton, New York, August, 1979.

Choice of Attributes: General

An activity is undertaken with some purpose in mind. This purpose or objective indicates the way in which one should move. In order to select from among a number of alternative PE systems, one must define the objective of interest and then identify specific characteristics (attributes) of the various alternatives. Identification of these attributes then allows for comparisons of the relative value of each alternative in accomplishing the objective. There are two basic approaches to choosing attributes: "bottom-up" and "top-down". A brief examination of each of these two approaches follows.

Bottom-Up Approach

The bottom-up approach simply lists important attributes as suggested by experience, literature, and experts in the field. Because of the different orientation and views of those selecting attributes, the resulting list may be long, perhaps too long. This list might be trimmed by eliminating variables which are unlikely to make a difference in the final measurement. One must be cautious of dropping attributes simply because it is not easy to obtain information or because they are difficult to quantify. This can introduce bias in measurement analysis toward "hard" considerations or attributes which are readily quantifiable, (e.g., profit) and away from softer or nonquantifiable attributes (e.g., public responsibility). This bias could significantly affect performance and hence attainment of the objective of interest.

> . . . It is easily demonstrated that even extreme inaccuracies in [decision] models will often have less effect on the probable quality of the solution than does unintended deletion from the problem description of possible outcomes or important attributes (Huber, 1974, p. 453).

One method of deciding whether it is worthwhile to gather data for a given attribute consists of examining the difference in the accomplishment of the objective the attribute might make compared with the cost of data acquisition. The inherent difficulty is that in the bottom-up approach, no specific objective is identified a priori, and as such, the appropriateness of an attribute in evaluating an alternative action is essentially intuitive. Thus, the bottom-up approach does not provide any assurance that only important attributes are suggested nor does it assure that all relevant considerations are included as the concepts of important and relevant are dependent upon the unspecified purpose of the alternative actions.

Top-Down Approach

The top-down approach begins by identifying the overall objective of interest. Knowledge of the objective should then be used as a guide in selecting specific attributes and in making decisions (Rutherford et al., 1972). Between objectives and attributes there may be sub-objectives. For example, the attribute "sales volume" might be under the sub-objective "maximize market share", a component of the objective "maximize wealth". By starting at the top (the objective) and working downward, a hierarchy can be developed. This development should encompass all relevant considerations and include only those attributes which are important to the objective of interest.

In constructing such a hierarchy the properties of completeness, operational, decomposable, non-redundancy, and minimal suggested by Keeney and Raiffa (1976) should be kept in mind. Satisfying these properties in constructing a hierarchy is not a simple task. One is

forced to think systematically about relationships between objectives and attributes which may make later value judgments easier.

Choice of Attributes for PE

The purpose of PE from an owner's point of view is to encourage the manager to select those activities (decisions) which will provide the maximum benefit to the owner. The objective of interest as previously identified is the maximization of owner's wealth. Various attributes are the alternatives from which a selection must be made. Having identified the objective of interest, the top-down approach is used to select the four attributes used in this study.

The attributes used were selected under the guidance of the attribute properties suggested by Keeney and Raiffa (1976). These attributes include a measure of productive efficiency and a measure of market effectiveness as suggested by Solomons (1965). The productive efficiency measure (PM) is found in terms of input units to output units in period t as compared to the same relationship in period t-1. A high PM factor indicates improvement of the input-output relationship over the prior period. This improvement is reflected in the amount (cost) of inputs per output or, in other terms, in a decrease in the variable cost ratio.

The market effectiveness measure (MSM) reflects a measure of market share. This measure is calculated as the firm's periodic sales value divided by the total market sales in the period. All firms begin with the same sales volume. In this study, market sales volume is initially set at $500,000 and increases $5,000 per period. The larger the MSM factor, the larger the portion of the market claimed by the firm.

The third attribute used in the study is that of profit (RORM). This measure is calculated as the firm's periodic net income divided by the initial investment of $80,000. A high RORM factor represents a large net income. A low or negative factor represents a low periodic net income or a loss.

The fourth attribute used in this study is a measure of productive (capacity) maintenance (CMM). This measure is calculated in terms of periodic capacity. A high CMM factor represents maintenance of productive capacity.

The four attributes selected--productive efficiency, market standing, profitability and productive capacity--form a multiple measure which has the five desirable attributes of completeness, operational, decomposible, non redundancy and minimal suggested by Keeney and Raiffa (1976). In selecting these four attributes of performance, all basic functions of a manufacturing process are represented. The basic components of business survival include the capacity, equipment and accessibility necessary to facilitate production, ability to produce marketable products economically, the ability to sell those products produced and the ability to operate with a profit. The ability to function well in all four of these areas usually insures the success of the firm. Lack of any of these four components for any substantial period of time can result in the failure of the firm. Thus the set appears to be complete as well as minimal and nonredundant. That the attribute set is operational has been verified in the four operational measures developed for use in this study. The desired property of decomposibility is illustrated in Figure 1. Note that each of the four components mentioned may be further divided into subcomponents.

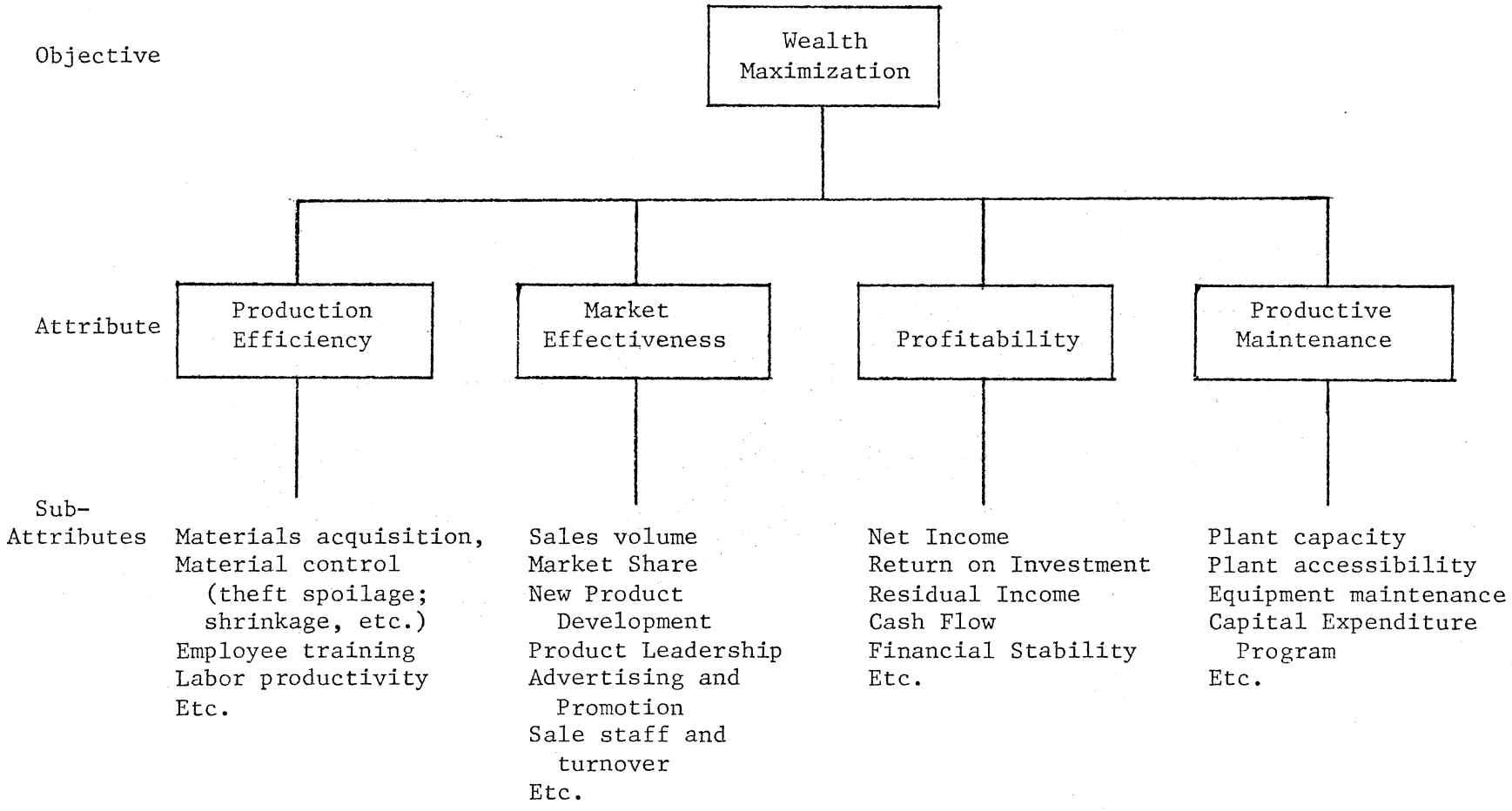


Figure 1. Multiple Attribute Composition of Wealth

This study suggests that in selecting a measure of wealth maximization, economic measures of productive efficiency, market effectiveness, profitability and productive maintenance constitute a complete yet minimal set of relevant attributes. The five measures indicated by Drucker (1954)--market standing, innovation, productivity, physical and financial resources and profitability--can easily be rearranged into the four measures suggested by this study. Likewise, the economic "key areas" suggested by General Electric (Solomons, 1965) of profitability, market position, productivity and product leadership are also amenable to those categories suggested in this study.

Scaling

Once attributes have been chosen, there must be a transformation to measures of value or suitability. This transformation is commonly referred to as value or utility scaling. The input to such a transformation is objective in nature (independent of the measurer) while the output is subjective in nature (a function of the measurer's judgment and preferences). Scaling methods may be categorized by three recognized types of measurement levels: ordinal, interval, and ratio (see Stevens, 1946). A brief discussion of each of the three levels follows.

Ordinal Scaling

In this general approach, levels of an attribute are classified into two or more categories. Examples include:

1. Acceptable and unacceptable
2. Poor, average, good, excellent

3. Hot, warm, cool and cold or

4. 1, 2, 3, 4 respectively.

Ordinal scaling analysis allows only for comparisons (<, =, >); the algebraic operations of addition, subtraction, multiplication and division are invalid. In essence, ordinal scaling is a simple ranking of attributes based on some factor.

The advantage of ordinal scaling is that it is rather simple to employ. There is no conceptual limit to the number of categories that can be ranked. The inability to define how much "better" or "worse" one level is in relation to another level is a disadvantage of ordinal scaling. Another difficulty is that risk and uncertainty cannot be taken into account rigorously.

In practice, judgment is used to create discrete ordinal value functions. A decision maker or group simply ranks or classifies the categories or levels of each attribute. Classification of attributes might be accomplished using the average of the responses from a Delphi or Nominal Group exercise. Thurstone's Law of Comparative Judgment (Thurstone, 1959) may also be used by groups to rank order categories of an attribute. This method uses a large sample of pairwise preference judgments to create an ordinal scale.

Interval Scaling

Interval scaling results in scales on which differences can be measured and are meaningful. An interval scale has an arbitrary zero point. Only the operations of addition, subtraction, multiplication and division by a constant are valid.

A typical example of interval scaling is the budgeting process. A budgeted amount is determined (an arbitrary zero point) and actual amounts are then referrenced to the budgeted level.

Interval scaling forces numerical rating which is more difficult than simple ranking for decision makers unused to making such judgments. Interval scaling does allow for risk and uncertainty in the construction of the various interval limits.

Ratio Scaling

Ratio scaling results in a scale whose intervals are meaningful and whose zero point signifies zero amount of the quantity. Profit in terms of dollars is a typical example. Zero profit indicates zero quantity of dollars. Under this method, statements such as "X is twice as large as Y" make sense.

This is the most flexible scaling method. Multiplication and division by interval- or higher-scaled variables can often be performed (in addition to the algebraic operations permitted with interval scaling), while still preserving at least an interval level of scaling.

The advantage of ratio scaling lies in the amount of information that it provides. Risk and uncertainty can be an integral part of the ratio scaling. A difficulty arises in that for many attributes there is no identifiable, absolute zero point.

Value Scaling of Attributes

The previous section discusses three basic measurement methods for scaling attributes. Each method examined defines the levels of the scaling established in a different manner. Ordinal scaling methods

result in a simple "better than", "equal to", or "worse than" type of comparison. No consideration of the degree of difference between two levels of an attribute is available. Consequently, no algebraic manipulation of attribute levels is valid.

Interval scaling methods do allow for comparisons of differences between different levels of an attribute. Algebraic manipulations of addition, subtraction, multiplication and division by a constant are valid. Concepts such as level 1 constitutes $1,000 net income per period less than level 2 are valid and useful. Since the zero amount in an interval scaled function is arbitrary, natural comparisons of value between levels of an attribute are not valid.

Natural comparisons between various levels of an attribute are permitted under ratio scaling. This is due to the nature of the nonarbitrary zero point of ratio scaled functions. For instance, $0 means an absence of profit and thus serves as a natural origin. Those interval scales which have a natural origin allow for direct comparisons between different levels of an attribute which make sense, e.g., $10 in profit is twice the value of $5. Obviously the interrelationship need not be linear nor constant over the total range of attribute levels.

It must also be noted that for most any attribute identified, a value function of any of the three types of levels discussed may be constructed. For instance, the attribute "sales volume" (an objective-independent of the measurer-input) can be converted into the following value scales (subjective--a function of the measurer's judgment and preferences-output):

Ordinal: High volume, Medium volume, Low volume.

Interval: ..., -2, -1, 0, 1, 2, 3, 4, ... where 100,000 units of sales constitute level 0 and levels increment by 10,000 units.

Ratio: 0, 1, 2, 3, 4, ... units sold.

Scaling of PE Attributes

Hobbs (1979) suggests four categories for evaluating the use of alternative methods. He suggests examining selection of alternatives with respect to the following:

1. Theoretical Validity. This refers to the extent to which the method chooses alternative X over alternative Y when X is actually preferred to Y by the decision maker.

2. Flexibility. This refers to the number of alternatives and attributes that can be handled, normative and descriptive characteristics of the method and theease with which implications of different perspectives can be determined

3. Results Compared to Other Methods. This refers to the extent to which alternative methods yield the same judgements.

4. Ease of Use. This refers to the time and cost involved, the degree of expertise required to understand the purpose, assumptions and general workings of the method and the extent to which decision makers believe that the method can reliably reflect their performance.

In this study, a scaling model which results in an interval level measurement is used. Examination of the four measures proposed for use in this study reveals a difference in the numeric range of each measure score. For instance, PM theoretically may vary from .950 to 1.680, MSM from 0.0 to $+\infty$, RORM from $-\infty$to $+\infty$ and CM from 0.0 to 1.0. These scale ranges for each attribute must be the same or implicit weighting of the individual attributes in constructing a composite score will occur.

In response to the need for equal scale ranges, the direct rating method was employed in constructing a standard continuum against which performance in each of the attributes is measured. This method uses a subjective continuum for each attribute. The continuum usually consists of a vertical or horizontal line with guideposts assigned numerical values. An expert or group of experts are then asked to rate how much each alternative contains of each attribute by identifying where it falls on the continuum (see Guilford, 1954).

In constructing these scales, the range of possible attribute measure values was categoried by the researcher into 11 components. The lowest component was assigned a value (points) of -1.0. Each successive level from the lowest to the best received an incremental .2 value. Thus, the lowest interval resulted in a value (points) of -1.0, the middle interval 0.0 and the highest interval +1.0 (see Appendix A).

This type of scaling model was chosen for several reasons. First, as shall be explained in the following chapter, the multiple attribute model used for comparing subject performances requires only an interval scaled value function. Thus, the added difficulties of obtaining a ratio scaled value function are unnecessary. The direct rating approach is theoretically valid and has been found to produce results similar to other more complex methods.

Perhaps of greatest practical importance in the choice of this scaling model is the flexibility and ease of use which it provides. Experience suggests that the practical usefulness of any program applied in a business is in part directly a function of its flexibility and simplicity. The direct rating approach is straight forward and easily understood. Introduction of this approach would be more acceptable to

the business community than would more complicated methods because of flexibility and simplicity. Other interval value approaches are available (i.e., deterministic questioning, gambling method, etc.), but tend to be more complicated than the direct rating method.

In addition, the nature of business changes over time. These changes require a scaling system which can be easily and quickly adjusted to the various ranges of attribute values that may occur. This flexibility is also necessary for the simulated environment constructed for this experiment. Due to the cumulative nature of the attribute measures used in the business game, the range of possible measure values changes from period to period. While maintaining the same number of value intervals and the same point allocation, the direct rating technique allows for period by period reassignment of interval limits. Thus, the same form of scaling is used for all attributes over all periods. Only the values of each attribute interval are reassigned as necessary.

Summary

This chapter has examined the issues of choosing and scaling attributes in a multiple attribute system. Two basic approaches to attribute choice are examined. The bottom-up approach lists all attributes of an objective as suggested by experience, literature, and "experts" in the field. Difficulties with this approach center on the inability to identify the attributes selected as relevant or important due to a clearly defined objective.

The top-down approach to attribute selection is explained and reviewed. This approach first identifies the specific objective of interest and then based on that objective selects a set of attributes

which encompass all relevant considerations with regard to achievement of the objective.

The top-down approach is then used for selection of a set of attributes to be used in managerial PE. Attributes selected include production efficiency, market effectiveness, profitability and productive maintenance. These attributes follow the four activities necessary for survival of a firm in the business community and satisfy the five properties suggested for a multiple attribute set by Keeney and Raiffa (1976).

Discussion then centers on the concept of scaling the various attributes to transform attribute scores into measures of value or suitability. Three types of scaling are discussed: ordinal; intermal; and ratio. Ordinal scaling simply ranks the attributes on the basis of some factor. No algebraic operations are feasible and direct comparisons between different levels of attribute values are meaningless.

Interval scaling results in scales on which differences can be measured and are meaningful. Algebraic operations with a constant are valid; however, since an interval scale has an arbitrary zero point, no direct comparisons of the degree to which two attribute levels differ are valid.

Ratio scaling is explained as essentially the same as an interval scale with a nonarbitraty zero point. Thus direct comparisons of the values of different attribute levels are valid in ratio scaling.

Various methods for creating ordinal, interval and ratio scaled functions are examined and discussed. Direct rating is selected for use in this study in creating an interval scaled function. This method is selected due largely to its flexibility and east of use, requirements necessary for any measure suggested for practical application.

CHAPTER V

A LINEAR COMPOSITE RULE

The discussion in the previous chapter centers on the selection and scaling of attributes for use in a multiple measure model. This chapter discusses the assignment of weights to the various attributes and selection of a decision rule to combine attributes and rank performances.

The chapter will first review the general problem of ranking performance using multiple attributes. A decision rule will be suggested for use in this study. Approaches for weight allocation will be reviewed and a weighting system for the attributes in this study selected. This discussion will then be summarized.

General Problem

In those situations where performances are compared on the basis of only one attribute, ranking and selection of the optimal performance is simple and straightforward. When comparison is based on multiple attributes which are noncommensurable, then ranking of performances and selection of the optimal performance is more difficult.

The problem of performance selection using multiple attributes can be defined formally through a payoff matrix. Let r_{ij} be the rating of the i-th criterion on the j-th performance ($i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$) as shown in Table I where more of r_{ij} is preferred to less.

TABLE I

PAYOFF MATRIX

Criteria	Performance 1	2	...	n
1	r_{11}	r_{12}	...	r_{1n}
2	r_{21}	r_{22}	...	r_{2n}
.	.	.	...	.
.	.	.	...	.
.	.	.	...	.
m	r_{m1}	r_{m2}	...	r_{mn}

In this setting, performance k can undisputably be said to be superior to all other performances if and only if $r_{ik} \geq r_{ij}$ for all i and j. Typically, the initial set of performances may be reduced in number by eliminating all dominated performances. That is, if $r_{iz} \geq r_{iy}$ for all i with $r_{ij} > r_{iy}$ for some i, then performance z is said to dominate performance y. Based on the attributes selected, performance z would always be choosen over y.

Elimination of dominated performances occasionally yields only one nondominated performance. In this case that performance would be the optimal performance and no further analysis would be required. More often, a set of nondominated performances result. Problems of this sort occur in many situations. For example, after eliminating 20 of 25 applicants, which of the 5 remaining candidates should be hired? In the context of PE, a ranking of all performances is desired. Thus, determining the worth of each performance is necessary.

In discussing the problem faced in selecting among a set of multi-attribute alternatives, Goicoechea, Hansen and Duckstein (forthcoming) make the following observation:

> . . . Determining the worth of alternatives that vary on many dimensions presents formidable cognitive difficulties. People faced with such complex decisions react by reducing the task complexity by using various heuristics. Unfortunately it has been observed that decision makers who rely on heuristic decision rules systematically violate the expected utility principle. Moreover, decision makers tend to ignore many relevant variables in order to simplify their problem to a scale consistent with the limitations of the human intellect. While such simplification facilitates the actual decision making, it clearly can result in suboptimal behavior (p. 250).

The above observation suggests that intuitive combination of information in an appropriate manner is extremely difficult. What is needed are analytical methods or decision rules which systematically help determine the worth of multiatttributed alternatives. A body of such decision rules exist in the literature. It is one of these many decision rules which has been selected for use in this study.

A Recommended Decision Rule

The weighted average decision rule is probably the most frequently used comparative procedure and is the decision rule recommended for use in this study. Essentially, numerical values for each r_{ij} (from Table I) are multiplied by weights assigned to each attribute. The resulting products are summed over each participant. The participant with the largest resulting summation is identified as the top performer.

This concept can be stated numerically as follows. Let $(w_i : i = 1, 2, \ldots, m)$ represent the set of weights assigned to each criterion (attribute). Let r_{ij} represent a numerical value for the ith

criterion and the jth performance. The worth of performance j can then be represented as follows.

$$u_{ij} = \sum_{i=1}^{m} w_i r_{ij} \qquad \text{(V-1)}$$

The weighted average decision rule is then used to rank each performance on the basis of its summed worth. In recommending the use of the weighted average decision rule for use in this study, a review of the method in light of the criteria set forth by Hobbs (1979) as discussed in the previous chapter is undertaken.

Theoretical Validity

Two basic conditions must be satisfied to ensure theoretical validity in the use of the weighted average method. Two simplifying assumptions are made concerning the multiple attribute utility function of the owner to satisfy these conditions.

The first condition is that each attribute's value or utility function is independent of the level of other attributes. This property is known as difference independence (Dyer and Sarin, 1977) or utility independence (Keeney and Raiffa, 1976). The second condition, called preference independence, is that the tradeoffs among attributes to be made by a decision maker are independent of the levels of any of the attributes (Keeney and Raiffa, 1976). This property means, for instance, that a decision maker is willing to give up a certain amount of production volume to increase quality control regardless of the base level of production volume or quality control.

The first simplifying assumption results in a linear single attribute utility function, i.e.,

$$u(r_{ij}) = a_i + b_i r_{ij} \qquad \text{(V-2)}$$

where $b_i > 0$. This equation implies that each attribute is monotonically increasing with respect to overall utility. For the plausible range of r_{ij} (usually in the center range of the utility function) it can be argued that the discrepancies between a nonlinear monotonic function and a linear function will not be significant. Many studies have indicated that unless the actual utility function is extraordinarily irregular, errors created by using a linear approximation tend to effect all alternatives in the same manner (Yntema and Torgerson, 1961; Fischer, 1972). Slovic and Lichtenstein (1971) and Tell (1976) survey studies which regress holistic (informal) judgments against attributes. They report that interactive terms rarely explain much of the variance, linear terms always dominate.

The second simplifying assumption is that the total utility for the j-th performance, u_j, is an additive function of $u(r_{ij})$ for all i. Notationally this is noted as follows:

$$u_j = \sum_{i=1}^{m} u(r_{ij}) \qquad \text{(V-3)}$$

Substituting equation (V-2) in (V-3) yields

$$u_j = \sum_{i=1}^{m} a_i + \sum_{i=1}^{m} b_i r_{ij} \qquad \text{(V-4)}$$

Since $\sum_{i=1}^{m} a_i$ is constant over all performances, the only parameters needed to estimate rank order on the n performances are the weights for the various attributes (letting $w_i = b_i$).

Flexibility

There is no limit to the number of alternatives or attributes

that can be handled by the weighted average decision rule. One study at the Oak Ridge National Laboratory evaluated over 20,000 alternatives on 29 attributes (Sarin, Dyer, and Nair, 1979).

Sensitivity analyses are easily done with weighted averages. The only parameter to be analyzed consists of the weights (b_i) assigned to each attribute. The implications of different weight sets have been examined in a number of studies (see Hobbs, 1979).

Results Compared to Other Methods

Several decision rules have been developed for use with attribute value functions of an interval level of measurement. Weighted average is perhaps the simplest of these rules. Others include decision analysis, goal programming, the Power Law, Hurwicz procedures and ELECTRE.

Hobbs (1979) reviews numerous studies comparing the results of the weighted average method with those of holistic methods as well as with the other rules cited above. Generally, results of the weighted average approach correlate very highly with the more complicated methods.

Ease of Use

The difficulty of the weighted average method depends on the attribute scaling and weighting methods used. The more complex the weighting methods, the more difficult the weighting summation method becomes. Still, this method is conceptually straightforward and with conventions of simplification becomes relatively simple to implement. It is more easily understood than most of the other multiattribute decision rules.

Use of Weighted Average

Of basic interest in this study is the use of a multiple attribute decision rule which can be easily used in practical business environments. For use in that type of setting, a decision rule must be flexible, simple to understand, and easy to implement. The weighted average method provides these desired characteristics while maintaining its theoretical validity. It is for this purpose that the weighted average method is used in this study.

As previously noted, the only parameter that must be identified to facilitate use of the weighted average method are the weights assigned to the individual attributes. The next section discusses the process of weight selection and assignment.

A Weighting Approach for PE

There are two fundamentally different approaches to weight selection: (1) observer-derived and (2) client-explicated. Basic to the understanding of the different weighting results produced by these different methods is the difference in the perceived purpose of decision rules between the observer-derived and the client-explicated approaches. Client-explicated approaches provide feedback and consistancy checks which serve to make decisions more logical and reliable than those made by purely subjective choice. The purpose of decision rules in the observer-derived approach is to predict (imitate) the decision makers' unassisted global evaluations. A discussion of each of these two basic approaches in the context of PE follows.

Observer-Derived Approach

Observer-derived techniques approach the weighting problem from a global or holistic evaluation view. Decision makers give overall evaluations of performance instead of attempting to assess the importance of each attribute. The global evaluations are then compared to the attribute scores by some decision rule (e.g., multiple regression, linear programming). This approach has been referred to as "policy capturing" and has been used in many settings (see Slovic and Lichtenstein, 1971; Huber, 1974).

A major disadvantage of this approach is the lack of control the decision makers feel over the weighting process. Rather than making some value assessment of the importance of each attribute and then viewing the combination of their attribute assessments into an overall assessment, decision makers now make a series of overall assessments and are told via some complex statistical procedure their views on the importance of each attribute. This "backwardness" becomes even more irritating when the statistical procedures assign negative weights to attributes that are a priori positively related to suitability.

The objective of using multiple attributes in PE is to focus attention on areas of activity which the attributes represent. The observer-derived approach would focus on an analysis of the overall evaluations of performance to assess what attributes received attention and in what relative proportion. Clearly, the observer-derived approach is inappropriate for selecting weights in a multiple attribute PE problem.

Client-Explicated Approach

This approach consists of methods in which a decision maker directly selects weights. This selection is often made via a series of preference or tradeoff questions. The purpose of this approach is to make evaluations more rational and consistent, thereby improving decisions.

The client-explicated approach is used in selecting attribute weights for this study. The purpose of the weighting done in this work is to influence the decisions of those being evaluated. The client-explicated approach allows direct selection of weights for each of the four attributes used in the multiple composite measure.

Under this approach, there are many methods available for determining the allocation of value: ranking, rating, ratio questioning, Metfessel allocation, the indifference tradeoff method, the Churchman-Ackoff technique, and decision analysis. Each method requires some form of subjective judgment on the part of the individual(s) setting the weights of the relative value of each attribute to the success of the objective. In this study, four attributes have been selected. Causal examination would indicate that achievement in all four areas is necessary for attainment of the desired objective of maximum long term wealth maximization. Since no one of the four attributes appears to be relatively more important in achieving wealth maximization, use of the ranking technique (or any of the other techniques) suggests equal weighting of the attributes used in this study.

Use of Equal Weights

It is recognized that different allocations of value among the four attributes chosen for this study might be made. In an actual business environment, experience or some other factor might suggest a slightly different weighting preference. The point of interest in this work is that a multiple measure PE system is used which forces participants to focus on each and all of the attributes of the measure. Placing equal weights on each of the four chosen attributes will accomplish this objective. Moreover, it should be noted that many studies have shown a high correlation of rankings produced by different weight sets. They conclude that unless significant value differences among the attributes appear, equal or even random weights are as useful as any of the methods discussed (see Dawes and Corrigan, 1974; Wainer, 1976, Einhorn and McCoach, 1977; Dawes, 1977).

Varying the weights by placing more emphasis on one or two attributes and less on the others will tend to dim or brighten the focus on an individual attribute. Taken to the extreme, allocating 100 percent of the weights on one measure and 0 percent on each of the others results in the single measure PE system against which the multiple measure system is being compared. The identification of the level of weight allocation where focus on an individual attribute no longer occurs is deferred to a subsequent study.

Exhibit IV in Appendix A illustrates the use of the weight allocation and decision rule discussed in this chapter. Note that at the bottom of the page the attribute measure scores from the fourth quarter of 19X0 are listed. Points are found by taking the attribute score and finding the point assignment on the appropriate scale.

Equal weighting is accomplished by simply summing the points received by each attribute. The weighted average decision rule then suggests that the best quarterly performance is achieved by that individual who obtains the highest total points.

Summary

This chapter begins by reviewing the general problem of comparing performance when multiple noncommensurable attributes are employed. In such cases, the need for an objective analytical decision model is discussed and confirmed.

The weighted average decision model is chosen for use in this study. This model takes the numerical performance score for each attribute and multiplies it by the weights assigned to each attribute. The resulting products are then summed. The participants' performances are then ranked based on their summed score.

The weighted average method is selected basically because of the ease in which it may be employed in a business environment. Its theoretical validity and performance compared to other decision rules are reviewed.

Two basic approaches for setting weights are introduced and discussed. The first approach, observer-derived, attempts to predict or imitate decisions. Decision makers give overall evaluations of performance which are then compared to the attribute scores by some decision rule (e.g., multiple regression, linear programming). Results from the evaluation are then treated as the value (weights) that must have been used by the decision maker in arriving at the initial overall

evaluation. This approach is found to be inappropriate for use in this study.

The second approach, client-explicated, consists of methods in which a decision maker directly selects weights. The ranking technique is applied to the four attributes in this study to produce weights of equal value. The concept of equal weighting is discussed with the conclusion that analysis of other weight allocations be deferred to subsequent studies. The use of the equal weights and the weighted average decision rule is illustrated.

CHAPTER VI

EXPERIMENTAL DESIGN AND METHODOLOGY

The purpose of this study is to examine the issue of PE with regard to a divisional manager in a profit making setting. Preceding chapters have developed the need for this study, reviewed current literature on this issue and provided a theoretical development leading to the two testable hypotheses stated in Chapter III. The purpose of this chapter is to develop and explain the methodology used for testing the four hypotheses of interest. This development covers the following areas: selection of experimental methodology; the KDSS game; experimental design; subject selection and sample size; PE measures employed and subject reward. This chapter concludes by giving a general overview of the actual process of the experiment.

Experimental Methodology

McGrath (1962) discussed four classifications of data-collection methods used in organization research: field studies; experimental simulations; laboratory experiments; and computer simulations. Placed on an ordering continuum, these methods may be thought of as proceeding from the concrete (at the field study end) to the abstract (at the computer simulation end). Alternatively, the continuum can be labeled as going from realism to artificiality, or from an open to a closed setting, or from loose to controlled conditions. Of importance is the

fact that the methods differ in terms of the advantages they offer and the limitations they impose. Those at the field study end contain the advantages of realism and the operation of inherent motivational forces; disadvantages consist of complexity and lack of control by the researcher. At the other end of the scale, the advantages and disadvantages are simply reversed.

In examining the hypotheses in this study, as much realism as possible in the examination was desired. However, the difficulty of maintaining control of critical variables precluded the use of a field study. Therefore, an experimental simulation was employed which provided the control necessary while attemtping to create as realistic a setting as possible.

The experimental simulation was operationalized through the use of a computerized business game, selected over a manual system for the computational speed and accuracy available in a computerized system. Also, use of a manual system might afford participants the opportunity to identify specifically the interrelationships of various game factors. As explicit identification of interrelationships in the operation of a business could not be examined from a written text, such an opportunity should not be available in this simulation.

The KDSS Game

Attempts made to select a game for use in this study suggested that a specialized game be constructed. Of interest in this examination was the issue of performance of individuals evaluated under two different PE methods. All of the canned programs examined provided a measure of net income. None of those games examined provided measures of production

efficiency, market effectiveness, and production maintenance as required for the multiple PE measure developed for this study. In addition, almost all of the games examined proved to be complicated enough to require several hours of thorough participant study before attempting to operate.

As a result of the difficulties identified in the use of a marketed business game, the KDSS game was constructed. This game placed the participant in the position of divisional manager of the Tulsa Unit of Hansen Company. This division was identified as a part of the small electronic games industry (see Appendix B).

Participants were required to make quarterly decisions on the level of expenditure for each of three areas. The first class of expenditures (E1) consisted of those which impact upon production capacity. The second class consisted of expenditures (E2) which impact upon production efficiency while the third class (E3) impact upon market position. The expenditure levels available for each area and the resulting effects of various levels of expenditure in each of the three areas were described and graphically illustrated.

The output provided by the game consisted of a quarterly income statement and four performance measures. The income statement consisted of sales, variable manufacturing costs, gross profit, quarterly expenditures made in each of the three areas and net income. Fixed costs were ignored and no inventory was considered.

The four performance measures indicated consist of a productivity measure (PM), a market share measure (MSM), a profit measure (RORM), and a capacity maintenance measure (CMM). Each of these four measures was explained in detail and the expenditures which affect each measure were identified.

Also included in the game material is an example of operations of the fourth quarter for 19X0. This example is given as an indication of the quarterly output available for each participant as well as an indication of the position of the firm at the initiation of the experiment.

Experimental Design

Of interest in the four hypotheses are the four situations previously mentioned. Illustrated in Figure 2, Situation I is characterized by a long term manager orientation where the manager is evaluated on the basis of a single short term measure of profitability (LRSM). Situation II consists of a long term manager orientation where the manager is evaluated on a multiple attribute measure (LRMM). Situations III and IV both assume short-run orientations with the manager evaluated on a single attribute of profitability (SRSM) and a multiple attribute measure (SRMM), respectively. Hypothesis I deals with differences between Situations I and II, Hypothesis II with differences between Situations III and IV, Hypothesis III with differences between Situations I and III and Hypothesis IV with differences between Situations II and IV. A completely randomized design was used for this study. This design allows for the relevant pairwise comparison of the results of the four situations.

Subject Selection and Sample Size

One hundred and twenty subjects were selected from the third, fourth and fifth year students in the School of Accounting at Oklahoma

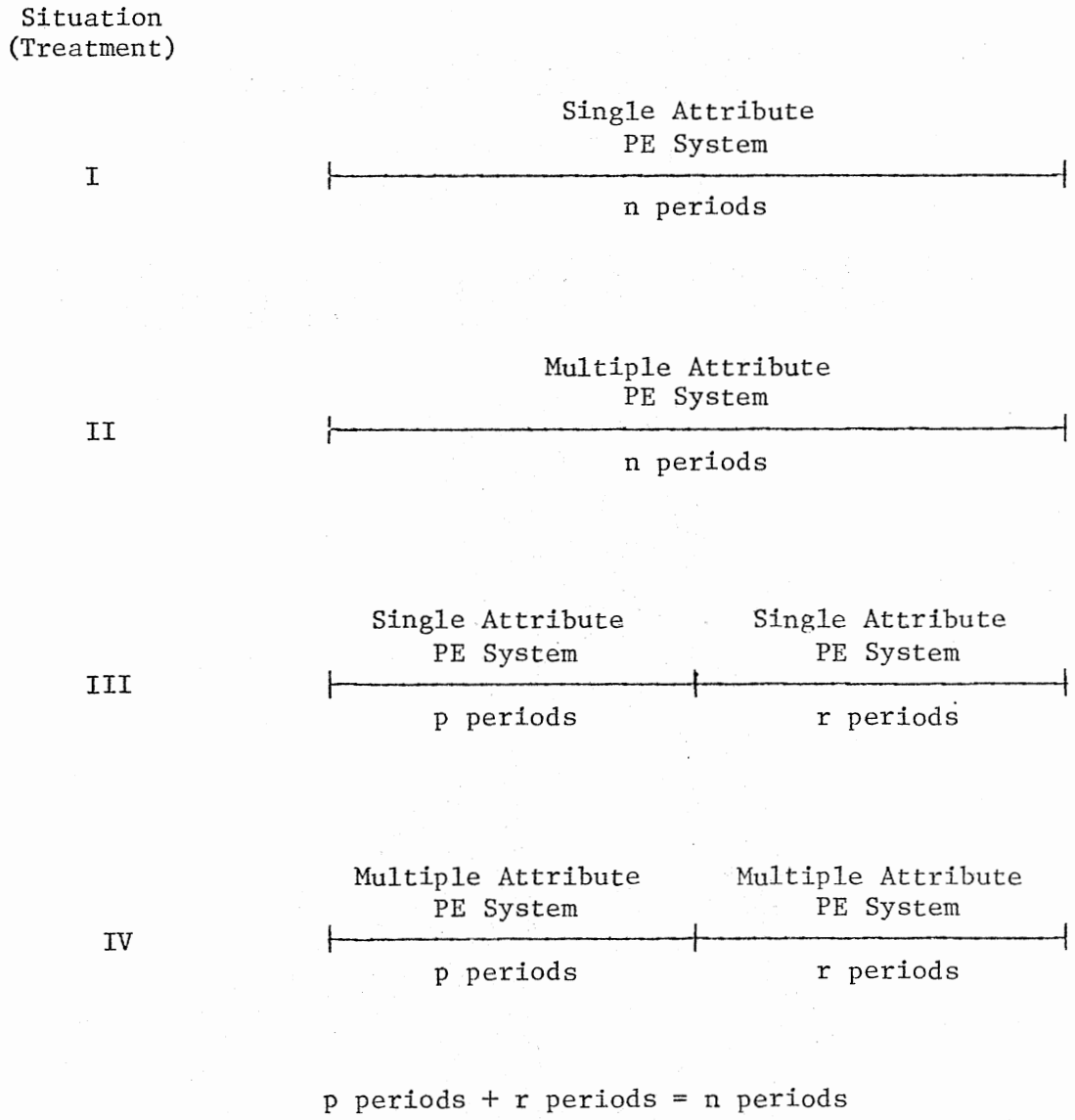


Figure 2. Illustration of the Four Environmental Settings Used in This Study

State University. Selection was restricted to accounting majors with a GPA of 3.0 or above.

Much has been written about the use of students as surrogates for managers in a business decision environment (Dickhaut, Livingstone and Watson, 1972; Copeland, Francia and Strawser, 1973; Abdel-khalik, 1974). The two major criticisms center on students' lack of experience in actual business settings and the gaming behavior often exhibited by students in experimental studies. Gaming behavior entails acting in a manner which would not be found in an actual environment.

The problem of lack of real world experience was to some degree offset by a knowledge of the business game and experience in its operation prior to the actual experimental session. The design of the experiment and an economically attractive reward structure were used to help limit gaming behavior.

Each student was placed in one of 20 groups of six participants each. Five randomly selected groups constituting a sample of 30 students were placed in a long-run single measure (LRSM) environment described in Situation I. Similarly, other samples of 30 students each were placed in long-run multiple measure (LRMM), short-run single measure (SRSM) and short-run multiple measure (SRMM) environments described by Situations II, III, and IV respectively. No attempt was made to distinguish between the third, fourth and fifth year students. General business knowledge was assumed to be comparable among all students. Randomization of group assignment eliminated any significant bias that might have occurred.

PE Measures Employed

Two PE systems were used in the study. The first was simply a quarterly comparison of performance based on the quarterly rate of return on original investment (RORM). The second consisted of a composite score of activity with regard to productive efficiency, market effectiveness, profitability and productive maintenance. Activity during a quarter was reflected in scores for each of the four performance measures. These scores were then rated according to the scales constructed for that quarter and points were then assigned for each measure. The total points for all measures became the quarterly composite measure which served as the basis for comparison of performance (see discussion in Chapter IV and V and an example in Appendix A).

Subject Reward

Each participant received a periodic performance score (either net income or the calculated composite score). A run sheet (see Appendix C) was used to tabulate the quarterly score. Each individual was then assigned a position from 1 to 6 according to the relative value of their performance score. The highest score was assigned position 1, second highest position 2 and so on. Identical scores by two or more individuals resulted in multiple positions being assigned to each. For example, if players #1 and #2 both scored 100, which is the top score in the group for the quarter, both players were assigned position 1-2.

For each period, 40 points worth $.10 each were allocated to participants within each group according to the schedule found in

Table II. Position ties simply divided evenly the points assigned to both positions (e.g., position 1-2 assigned (19 + 10)/2 = 14.5 to each of the individuals involved).

TABLE II

QUARTERLY DISTRIBUTION OF PAYOFFS TO PARTICIPANTS

Position	Points
1	19
2	10
3	5
4	3
5	2
6	1
Total	40

The point assignment for each quarter was then transferred from the run sheet to the group summary sheet (see Appendix C). This form provided a summary of the points earned by each participant over the course of the experiment. It also served as the basis from which monetary allocation of payoffs at the conclusion of the experiment were made.

Experimental Process

The following sections describe the process undertaken in operating the experiment. Details of the activity are described in the order in which they occurred.

Subject Enlistment and Introduction

The first step in conducting this experiment consisted of enlisting the participation of qualified subjects. Contact was made with students in various upper level accounting classes as well as in various locations in the College of Business Administration building. These students were told that they were needed for an experiment examining managerial level performance which would take a total of 2-3 hours of time scheduled at their convenience over the next two weeks. They were informed that payment for their participation would be contingent upon their performance and would average $5-$8 with payoffs of $18-$20 possible. Those indicating interest in participating were given the introduction and consent form. This form contained a brief written explanation of the purpose of the study and a consent form which each student participating in the experiment was asked to sign (see Appendix D).

Trial Runs

After collecting the consent forms, an instruction sheet and a trial run form (see Appendix E) in addition to a KDSS game packet was given to each student. The instruction sheet specified the time and place for turning in the trial run form and signing up for an experimental session. This sheet as well as direct verbal communication instructed the participant to keep all work on a personal and individual basis.

The participants were instructed that the trial run forms were for them to experiment with the operation of the game. It was emphasized that there was no objective or goal that they should work towards in

the decisions made for the trial run. The trial runs were simply available to aid their understanding of the mechanics of the game and to allow a pretest of the computer program for elimination or correction of any perceived deficiencies. Only one trial run of three quarters each was provided for each participant.

Returned trail run forms were then batch processed in the evening and made available the following day. When the participants returned for the results of the trial runs, any questions they expressed concerning the mechanical operation of the game were answered.

Session Sign-Up

Upon receiving the results of the trial run, each participant was asked to select (on a first come, first served basis) one of 20 two-hour periods in which to participate in the experiment. Each two-hour period was limited to six students. Prior to completion of the trial runs and sign-up of participants, each of the four environmental conditions was assigned in a random fashion to five of the 20 periods (see Appendix E).

Experimental Session: Introduction

After each group of six participants arrived for the experiment, a short review of the results of the two selected trial runs was conducted. Any specific questions as to the mechanical operation of the game were answered at this time. The purpose of this review as well as the trial runs was to minimize any first period(s) learning effects during the actual experiment.

Next, each person was given a decision form and identified as a player (see Appendix G). A description of the environmental setting

under which the group was to operate was given to each participant (see Appendix H). For those environments with the multiple measure setting, Exhibit IV (see Appendix A) was also given and explained to each participant. After discussion on Exhibit IV concluded, the scales for each quarter to be used in the experiment were given to each of these participants (see Appendix A). It was specifically noted and pointed out that while the interval limits of the productivity measure (PM) and the capacity maintenance measure (CMM) remain constant over the course of the game, the interval limits of the market share measure (MSM) and the profit measure (RORM) shift each period. This shift required increased scores of these measures from each succeeding quarter's operations in order to even maintain the present cumulative point level achieved on that measure.

At this point in the process, a distinction was made as to the groups which anticipated long-term job tenures and those which anticipated short-term job tenures. Those participants in the short-term settings were initially given only Part A of their environmental settings. Part A described a short-term position with promotion to another assignment expected in 3-5 quarters (see Appendix H). The short-run multiple measure groups also received only the first five quarters of the scaling forms for calculation of their performance. Each short-term group was explicitly told that they could only <u>expect</u> to be in their position for 3-5 quarters after which a superior would review their performance and make a decision on their subsequent assignment. Those participants in the long-term setting were told they would operate for 10-12 quarters before any review would occur.

Finally, the point distribution (payoff) for quarterly performance was discussed and illustrated (refer to Table II, p. 67). Included in this explanation was a discussion and illustration of both the run sheet and the summary sheet as they would be used for the group.

The group was then told to take as much time as they desired in individually studying the new material. The instruction of no communication between players was reemphasized. When each student was ready with their decisions for the first quarter, the actual operation of the game began.

Experimental Session: Operation

As each student presented quarterly decisions, the data were entered into the computer via a portable T. I. Silent 700 or DEC 10 remote terminal. Printouts of the decisions results were returned immediately on the terminal used and given to the student. Results (performance scores) of the quarter's activity for the student were tabulated and listed on a run sheet and the student was then allowed to prepare for the next quarter's decisions. When all the results of one quarter were received, payoffs to each participant were accumulated on the summary sheet.

At the end of the fourth quarter, those participants under the short-term settings received Part 2 of their performance measure which indicated that no advancement or movement of position was available. After reading the instructions and finding that they would continue in the same position under the same performance measure, these participants were instructed that the game would proceed for another 6-8 quarters.

At the end of a total of 10 quarters, all operations ceased. Participants of each group were assembled and thanked for their participation. The absolute necessity for silence with regard to their experience was discussed and all material relevant to the experiment was collected from each participant. Total remuneration due each person was summarized and dates for collection of the funds identified. Each group was then dismissed and told that results of the experiment would be available for examination two weeks after the final group was run.

It should be noted that the initial discussion with each group involved 15-20 minutes. The first quarter's decisions were ready for processing approximately 5-10 minutes later. Total duration of the experiment varied from 1 1/4 to 1 1/2 hours with one group delayed slightly by computer difficulties.

Summary

This chapter examines the experimental design and methodology used in this study. An experimental simulation was employed to provide as much realism as possible while maintaining necessary control. One hundred and twenty third, fourth and fifth year students from the School of Accounting at Oklahoma State University with GPA's of 3.0 or above were arranged into 20 groups of six participants each. The four environmental settings (LRSM, LRMM, SRSM, SRMM) were each randomly assigned to five of the experimental groups.

The chapter next discusses the payoff available for each participant. This remuneration was calculated quarterly using either a single PE system of profit or the multiple composite measure developed in this

study. The chapter concludes by reveiwing the process of the actual experimental activity.

CHAPTER VII

ANALYSIS OF RESULTS

The experimental results of this study are discussed in this chapter. First, the behavioral activities of the participants are reviewed. No quantification or statistical methods are used for this behavioral analysis. Rather quantitative observations of behavior relating to the four experimental hypotheses are discussed.

The second part of the analysis consists of a graphical representation of the quarterly expenditure decisions and resulting quarterly performance attribute scores of participants in each of the four environmental groups. A third portion of the analysis consists of a quantitative analysis of the performance of the participants in each environmental group with regard to the wealth of the firm at the terminal point of the experiment. A vector of attribute performance scores is used to create a surrogate measure of the value of each of the 120 firms at the end of the experiment. The resulting performance of participants in each of the four environments is analysed using multivariate statistical analysis.

Behavioral Analysis

In the four experimental hypotheses there are two conditions which may vary. Hypotheses I and II hold anticipated job tenure constant and vary the type of PE system employed (single vs. multiple). Hypotheses

III and IV hold the type of PE system employed constant and vary the anticipated job tenure (short-run vs. long-run). Analysis of the behavioral effects of the various PE systems employed and anticipated job tenures used is of interest in this study.

PE Systems Employed

The only behavioral effect that resulted from the use of a single vs. a multiple attribute PE system was the difference in time required to make the initial quarterly expenditure decisions. Those participants evaluated under the single measure of profitability made their initial sets of decisions in a matter of 3 to 4 minutes. Those participants evaluated under the multiple measures spend considerably more time (8-10 minutes) constructing their initial sets of decisions.

This behavior lends credence to the premise that a multiple measure PE system will focus attention on more attributes than the one attribute which receives full focus under the single measure PE system. Participants under the single measure focused only on profitability and required relatively little time to construct their decisions. Participants under the multiple measures had to focus attention on all four attributes (PM, MSM, RORM, CMM) and thus required additional time for decision making.

Anticipated Job Tenure

Two interesting behavioral activities resulted from the use of various anticipated job tenures. One activity concerned tenure and termination while the other concerned reorientation of individuals in the short term setting.

Tenure and Termination. Probably the most often asked question by the participants concerned the termination point of the experiment. Those participants in the long-run settings were told initially that they would operate for a total of 10-12 quarters. Concern was often expressed during the first few quarters of operation that early termination would occur. Towards the latter quarters (e.g, 8 and 9), probing by the participants was directed to identifying the particular quarter (10, 11, or 12) in which termination would occur. All inquiries received the response, "Termination of the experiment will be in quarter 10, 11, or 12."

For those operating under the short-run environments, initial instructions identified 3 to 5 quarters as the tenure of operation. This first part was terminated after period 4. The second part began with instructions of 6 to 8 additional operating quarters. This portion was terminated after 6 quarters, resulting in a total of 10 quarters of activity for those in the short-run settings. Probing as to the termination point in this setting was similar to that exhibited in the long-run settings.

Another aspect in this area of interest is the behavior exhibited upon termination of the operations. In both the long-run and the two parts in the short-run settings, a time span of 3 quarters was given for termination. In the short-run settings, termination of Part A was in the second of three possible termination periods. For the long-run settings and Part B of the short-run settings, operations were halted in the first of the three possible termination periods. Discussions with the participants universally received the comment, "If I had only known we were going to stop this period, I would have really cut my

expenses." This comment and the concern exhibited surrounding the terminal periods suggests that decisions made when termination is imminent differ from those made when tenure is continuing.

Reorientation. One of the most interesting behavioral aspects of this experiment deals with the reaction of the participants in the short-run settings when they found that they would remain in the same position for an additional 6 to 8 quarters. Those participants under the single measure PE system seemed very perturbed and a little confused as to their subsequent decisions. The most frequent comment at this point was, "Great, what do I do now?"

In contrast, those participants operating under the multiple composite measure PE system seemed almost unconcerned about their continuation. The most frequent comment was, "Oh, okay", after which input for the subsequent period was almost immediately available.

The difference in these two reactions appeared to be due to the respective positions of the firms after the fourth quarter. Relative to the initial position of the firm, those operating under the SRSM environment had made decisions which resulted in a poorer market share, decreased productivity, and reduced capacity position. Past decisions had essentially milked the future position of the firm. Thus, when faced with continuing on with operations, participants had to pay the price for their past work.

Those who operated under the SRMM environment made decisions in the first quarters that resulted in a position at the end of quarter 4 which was much better with regard to market share, productivity, and capacity than the initial position of the firm. As hypothesized, the multiple composite measure focused attention on long range decisions

and thus stopped any milking that might occur. Participants faced with continued operation of their firm were not hurt at all by any short-run orientation of prior decisions.

Graphical Analysis

From the experiment conducted, information is available with regard to levels of expenditure made by each participant for each quarter. Data from operations such as average production efficiency, market share, compounded values, and production capacity for each of the 120 operating firms at the end of quarter 10 are also available. A discussion on each of these areas follows.

Expenditure Levels

The level of average expenditures for E1, E2, and E3 for each of the four environmental settings (LRSM, LRMM, SRSM, SRMM) are illustrated in Figures 3 through 5. Of interest is the fact that the levels of expenditure in the multiple measure settings are, for all three expenses, higher than those under the single measure settings.

Also of interest are the dips in expenditure level of those under the SRSM environment. Large drops are seen in fourth quarter expenditures and decreases are seen again in quarter 10 levels. Much smaller decreases are seen under the SRMM. These drops do not appear under the long-run measures. This indicates that participants were sensitive to the terminal points, but that gaming with decisions at the end of the experiment was avoided.

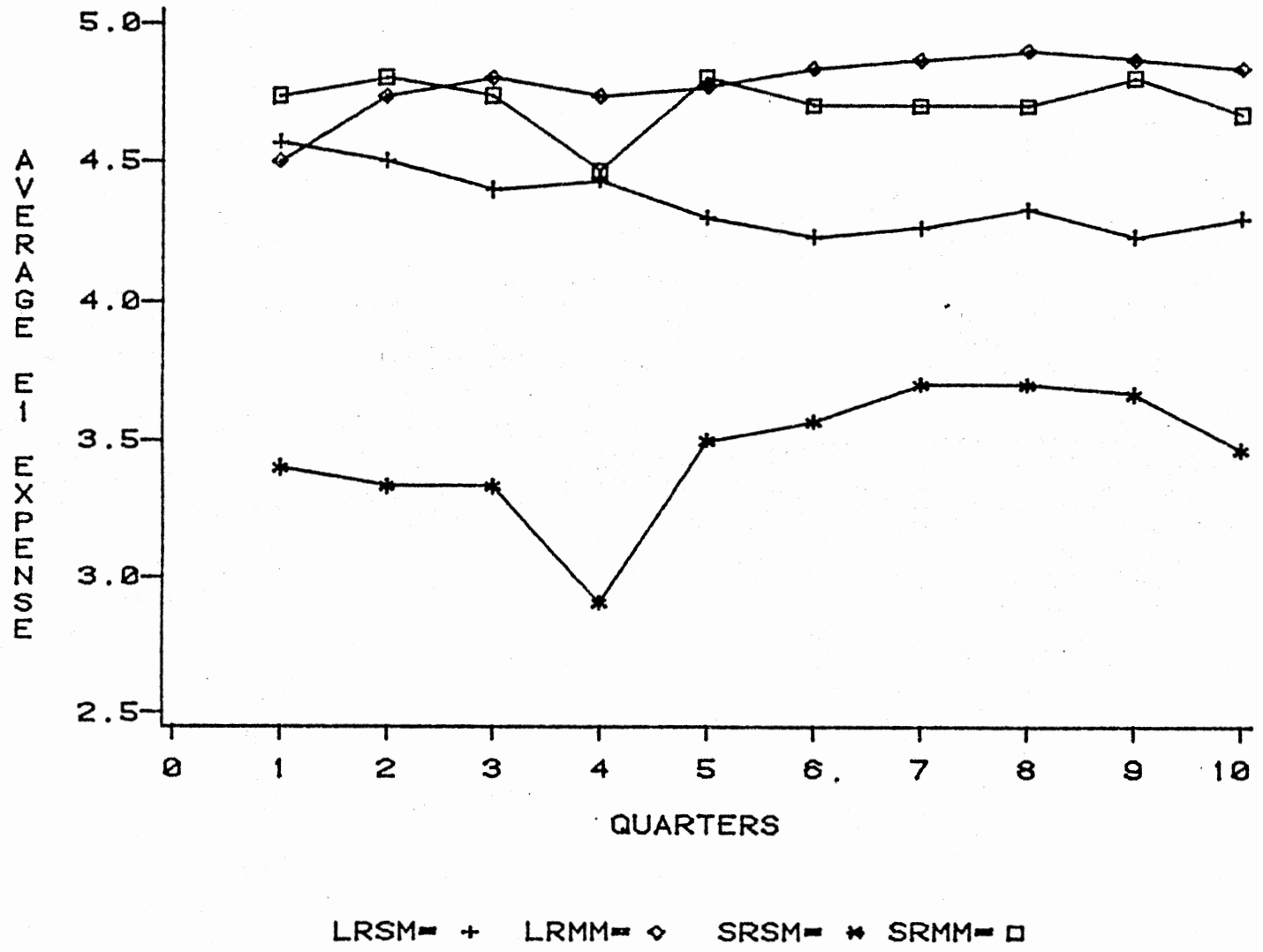


Figure 3. E1 Expenditure Levels

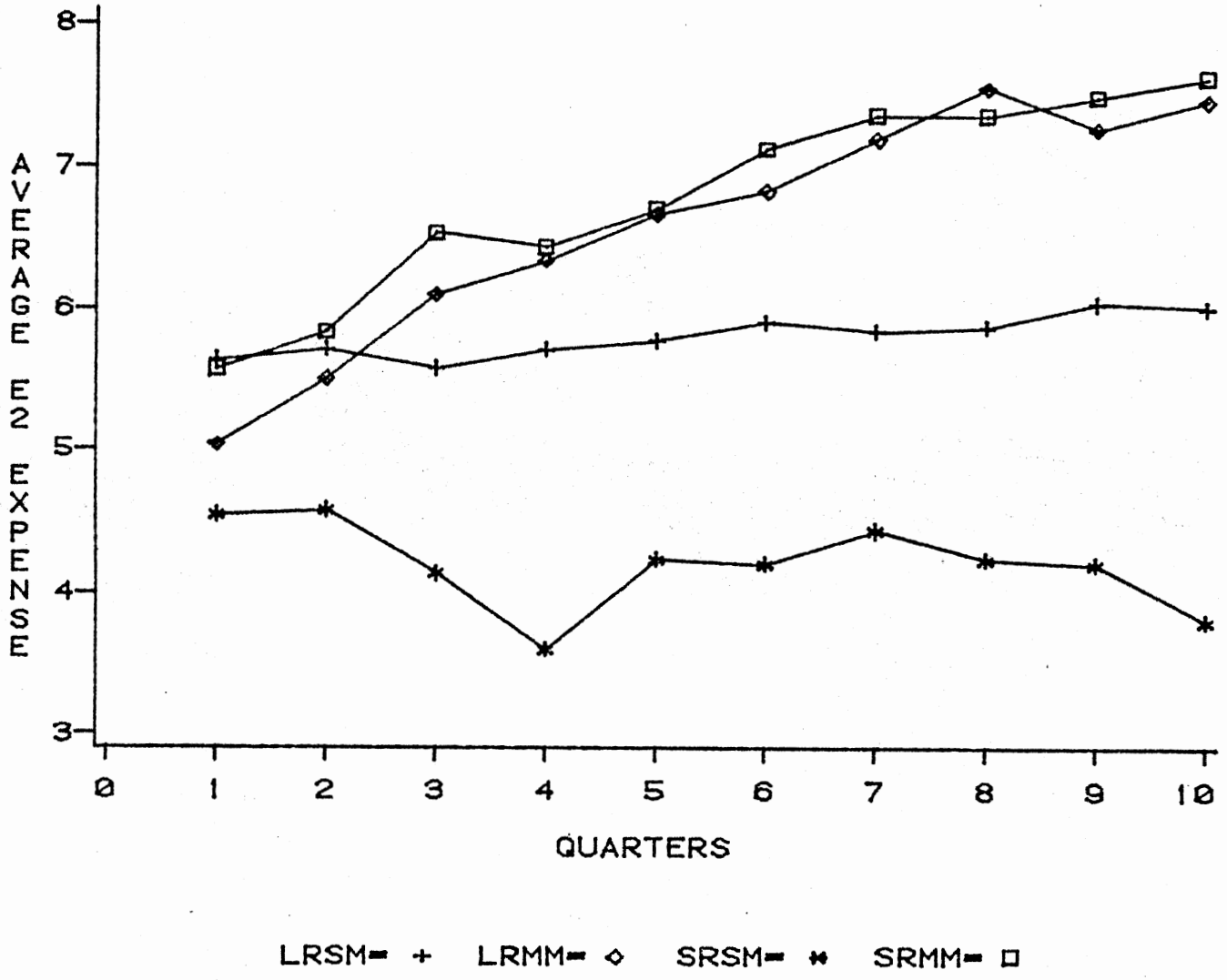


Figure 4. E2 Expenditure Levels

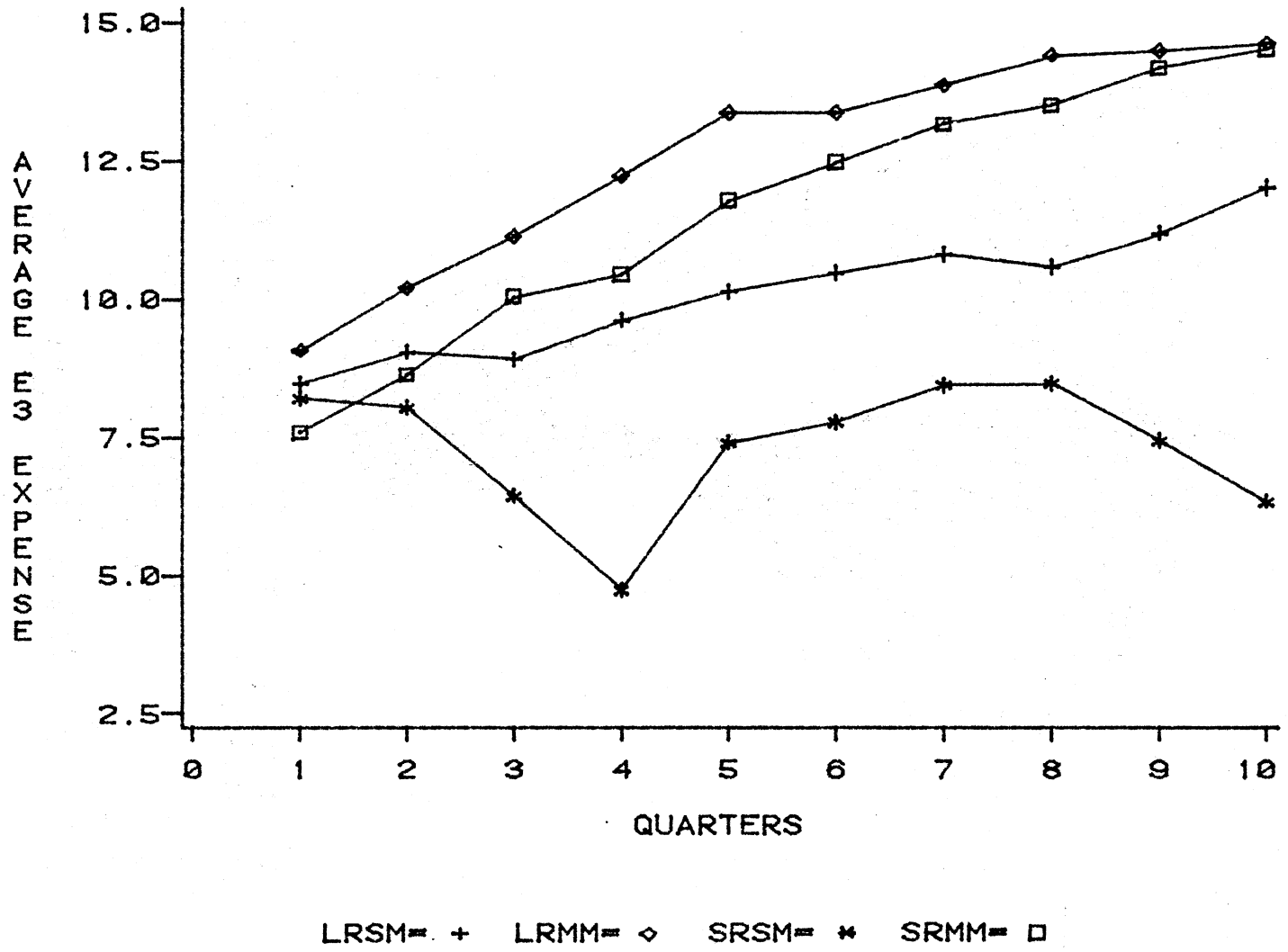


Figure 5. E3 Expenditure Levels

Data from Operations

In addition to the graphical analysis of the expenditures made by each of the four groups, graphical analysis of the quarterly performance results is also of interest. Four economic measures of performance have been selected as constituting a complete yet minimal set of attributes with regard to wealth maximization, i.e., measures of production efficiency (PM), market effectiveness (MSM), profitability (RORM), and productive maintenance (CMM).[1] An analysis of the average performance in each of these areas by each group follows.

Figure 6 depicts the average PM performance for each of the four groups over the 10 operating quarters where performance is measured with respect to Quarter 4, 19X0 (i.e., the base period). Average performance measured against the base period is found as the product of the PM score in quarter N with those of all prior periods, e.g., the base period PM for quarter 3 is the product of the scores for quarters 1, 2, and 3. Obviously, the higher scores indicate increased productive efficiency. Of note is the fact that performances by the multiple measure groups are extremely similar and far superior to those of the single measure groups.

Figure 7 depicts the average market share measure (MSM) for each environment over the 10 quarters. All groups start at 10 percent and drop initially due to the time delays built into the game. Of importance is the rise of the MSM's for the multiple measure groups. Note that the MSM raises to 13.6 percent and 11.7 percent for the two multiple measure groups while dropping to 9.9 percent and 6.5 percent for the single measure groups.

---

[1] See the discussion in Chapter IV.

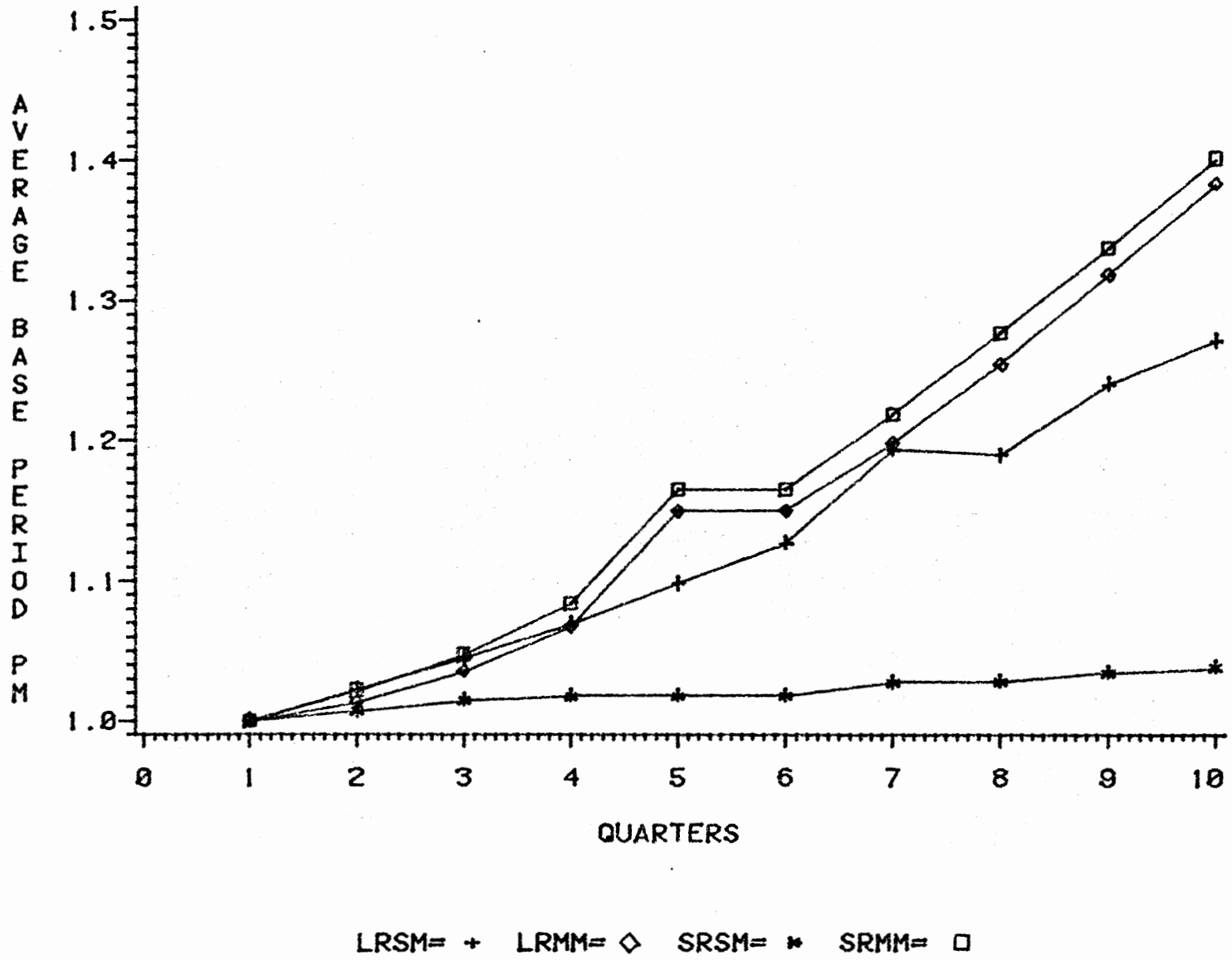


Figure 6. Average Base Period PM for Each Environment Over the Operating Period

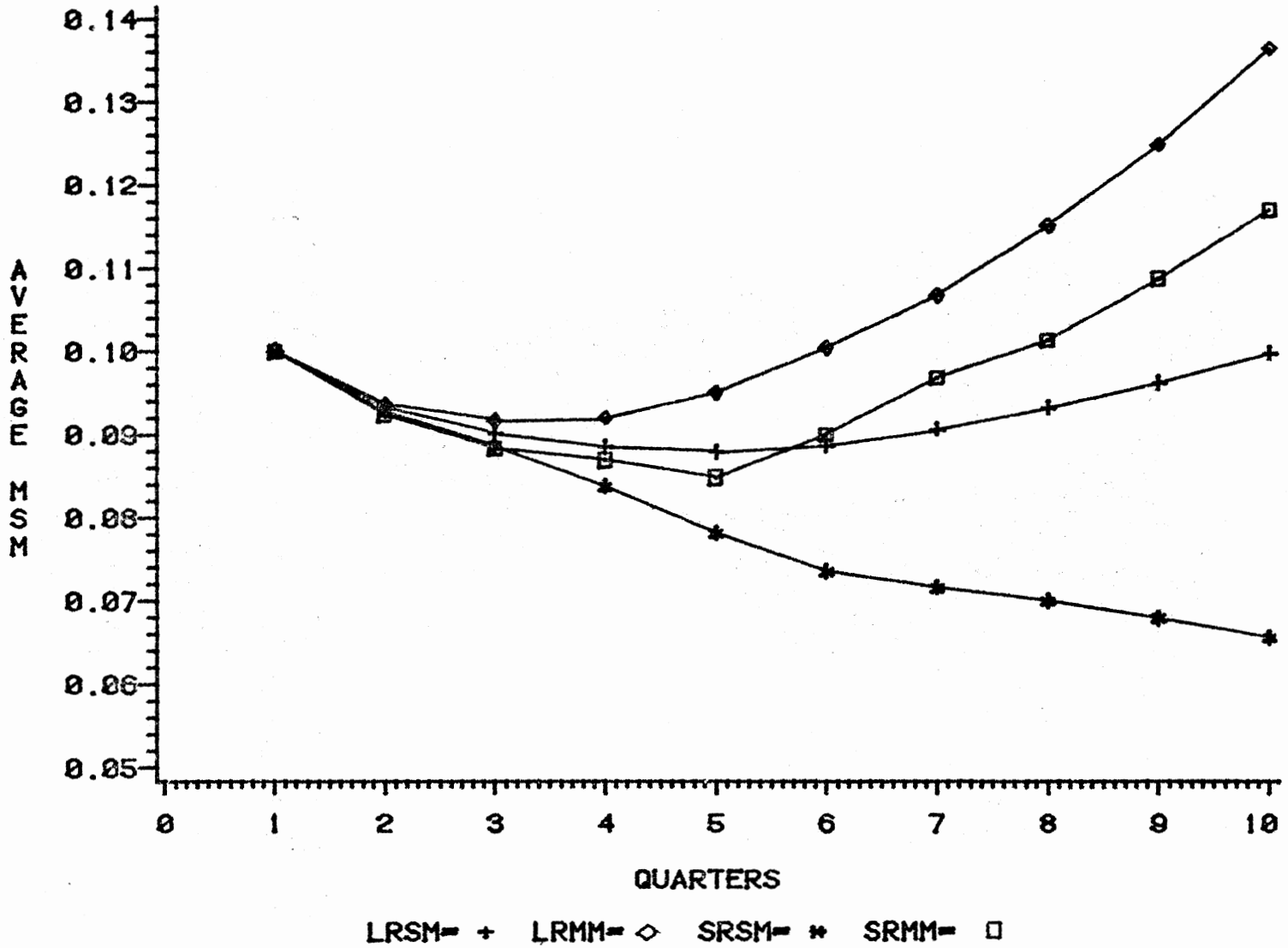


Figure 7. Average Market Share Measure (MSM) for Each Environment Over the Operating Period

Figure 8 depicts the average quarterly profit achieved by each of the four groups. Note that over the short-run (first four periods), the single measure groups achieve more profit than do the multiple measure groups. However, beginning after the fifth period the performances of the multiple measure groups are superior.

One further step can be made with respect to analyzing the average quarterly profit made by each of the four groups. Using the concept of the time value of money, the quarterly cash flows can be compounded to a single point in time and the resulting totals compared. Figure 9 depicts the compounded values of each groups' average net income flow. Used in this process are compound rates from 8 percent to 24 percent. Note the superiority of the performances of the two multiple measure groups over the single measure groups. Also note that the compound rate used does not affect the comparative relationships within the range of rates depicted. This would indicate that the comparative analysis of the average profits earned by each of the four groups is the same for all individuals regardless of risk attitude.

Figure 10 depicts the average capacity maintained by each of the four groups. Note that the two multiple measure systems maintained average capacity at 91 percent and 88 percent while average capacity of the single measure groups dropped to 78 percent and 52 percent.

Statistical Analysis

The wealth to the owner of a firm at any point of time is traditionally denoted as the compounded cash flow from past periods and the present value of all future periodic cash flow. In this experiment, past cash flows are available but future flows are not. The future

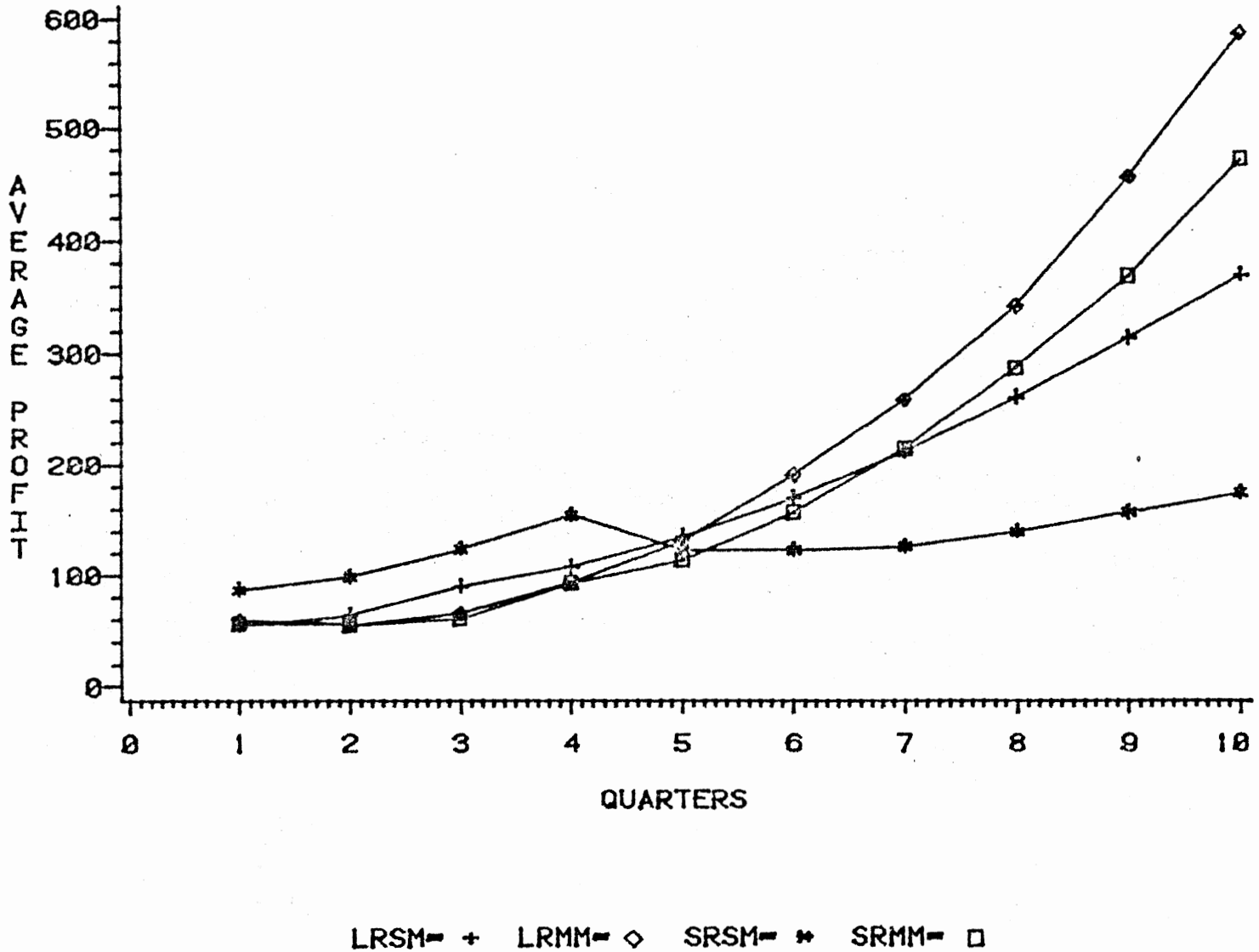


Figure 8. Average Profit for Each Environment Over the Operating Period

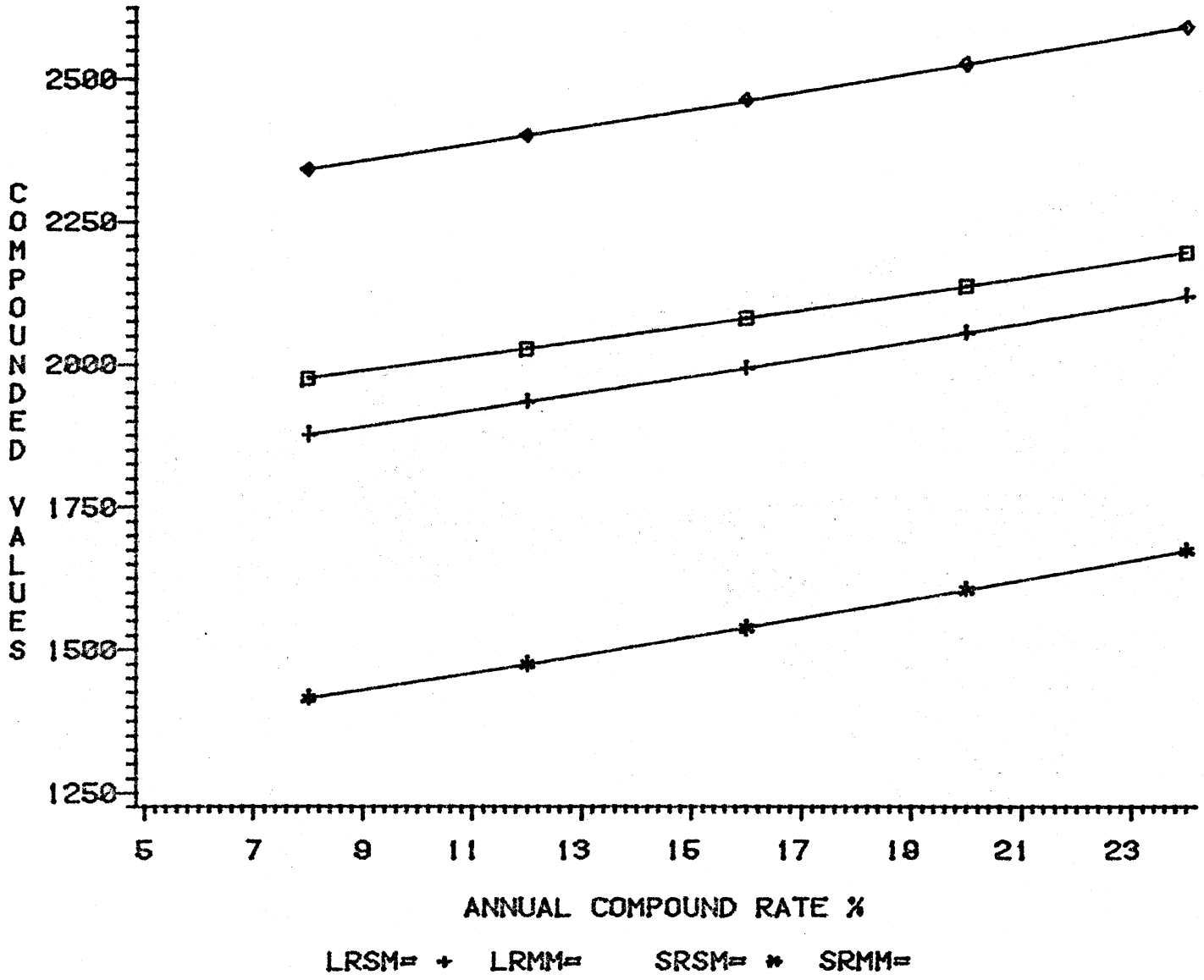


Figure 9. Relative Compounded Values Using Different Compound Factors

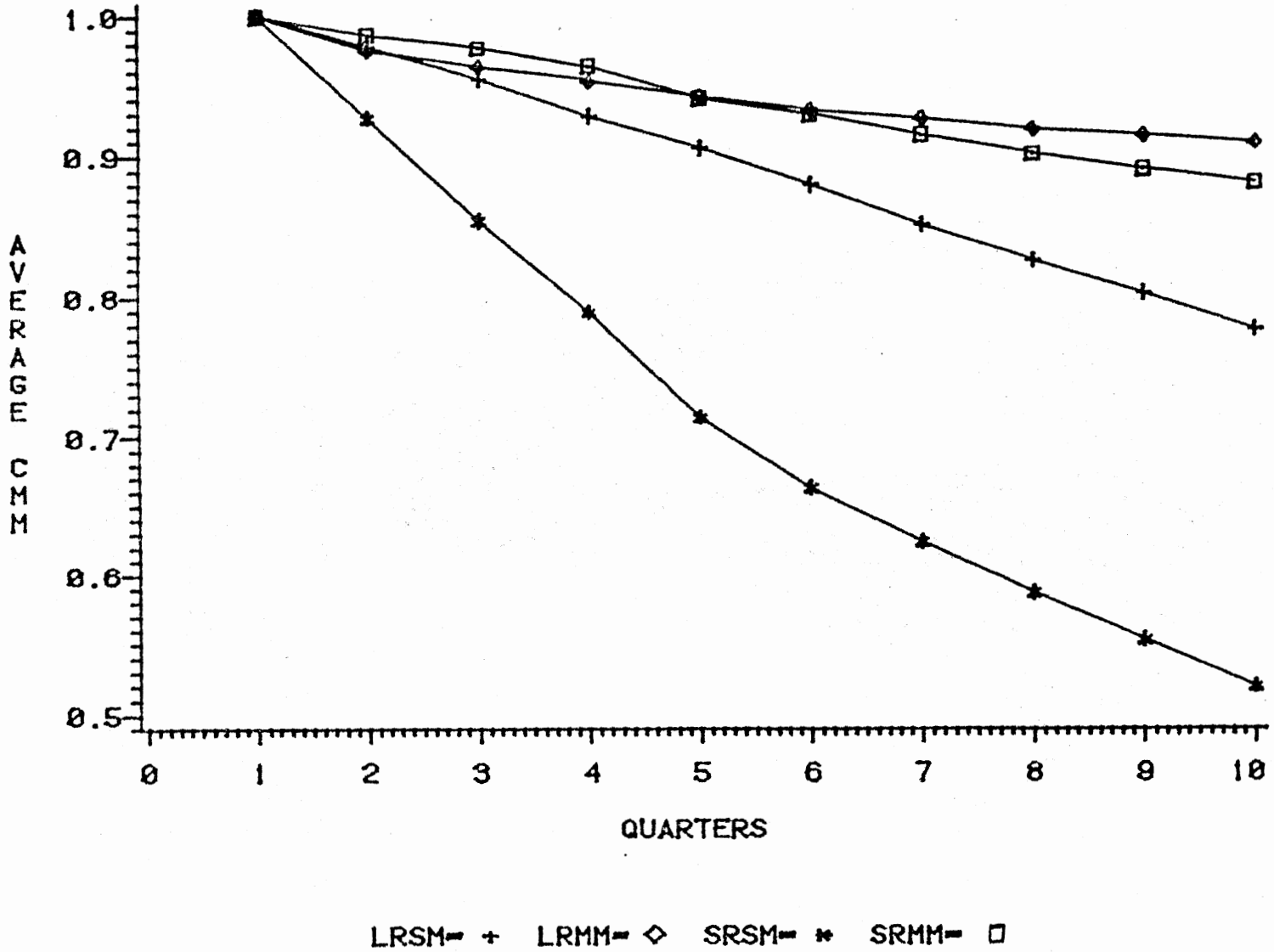


Figure 10. Average Capacity Maintenance for Each Environment Over the Operating Period

cash flows could be estimated, but such an estimate would be arbitrary at best. In this study, the wealth or benefit to the owners from the operation of the firms consists of two major partitions: past cash flow (compounded cash flow (CCF) to the end of quarter 10 at 16 percent) and three components reflecting the anticipated future cash flows. These three components are production efficiency (base period PM at quarter 10), market effectiveness (MSM at quarter 10) and productivity maintenance (CMM at quarter 10).

An average wealth vector is constructed for each of the four environmental groups. Then a statistical analysis of pairwise equality using a multivariate Hotelling's T^2 is conducted for each hypothesis. For those comparisons which result in a significance of 10 percent or less, a pairwise comparison of each of the four attributes using 90 percent Bonferroni confidence intervals is conducted in order to gain some feel for the cause(s) of the differences. The intervals created contain the range of difference values between the attributes of interest. Therefore, if the interval for D versus E contains the value of zero then one must say that no difference in the two attributes is indicated. If the interval is on the lower (negative) side of zero, the value of D is less than E. If the interval is on the positive side of zero, the value of D is greater than E.

Hypothesis I

Hypothesis I is as follows:

> If a manager anticipates a long-run appointment, there will be no difference in the welfare of the owner resulting from the use of a PE system using the single attribute of short-run profitability or a PE system using a composite of short-run and long-run attributes.

A pairwise comparison of the LRSM and LRMM wealth vectors using Hotelling's T^2 results in an overall significance level of 0.0001. This significance level suggests that the hypothesis that the LRSM and LRMM wealth vectors are equal may be rejected with a 99.99 percent probability of being correct. Analysis of the performance of the two groups on each of the four attributes results in the following confidence intervals.

TABLE III

CONFIDENCE INTERVALS - LRSM VS LRMM

CCF:	[-897.0389221, +19.37292212]
PM:	[-0.234377476, +0.011429476]
MSM:	[-0.056548492, -0.018785508]
CMM:	[-0.228191191, -0.034208809]

Since the intervals for PM and CCF contain zero, the performance of the LRSM and LRMM groups on these two measures are comparable. However, since the intervals for MSM and CMM are on the negative side of zero, performance for members of the LRMM groups can be judged to be superior to performance of the LRSM groups on these two measures.

With regard to Hypothesis I, the results of this experiment indicates that if a manager anticipates a long-run appointment, there is a difference in the welfare of the owner resulting from the use of a PE system with a single attribute of short-run profitability versus use of a PE system with a composite measure of short-run and long-run

attributes. Specifically, the performances of the LRMM groups are superior to the performances of the LRSM groups with regard to market effectiveness (MSM) and productive maintenance (CMM).

Hypothesis II

Hypothesis II is stated as follows:

> If a manager anticipates a short term appointment, then the welfare of the owner will be greater if a multiple composite measure PE system consisting of both short and long term attributes are used than if a PE system using a single short-run profitability measure of performance is employed.

A pairwise comparison of the SRSM and SRMM wealth vectors using Hotelling's T^2 results in an overall significance level of 0.0001. Thus equality of the two measures may be rejected with 99.99 percent probability of being correct. Further pairwise analysis of the Bonferroni confidence interval for each attribute pair results in the following information.

TABLE IV

CONFIDENCE INTERVALS - SRSM VS SRMM

CCF:	[-860.8186828, -266.3146372]
PM:	[-0.472342347, -0.252840653]
MSM:	[-0.066009609, -0.036923731]
LMM:	[-0.457807779, -0.271525621]

Since the intervals for all four measures are on the negative side of zero, performance of the SRMM groups is termed superior over all

attributes to performance by the SRSM groups. Thus, Hypothesis II cannot be rejected. The results of this study indicate that is a manager anticipates a short term appointment, then the welfare of the owner will be greater if a multiple composite measure PE system consisting of both short and long term attributes are used than if a PE system using a single short-run profitability measure of performance is employed.

Hypothesis III

Hypothesis III is stated as follows:

> If a PE system using a single short-run profitability measure of performance is used, the welfare of the owner will be greater if a long-run appointment is anticipated than if a short-run appointment is anticipated.

A pairwise comparison of the LRSM and SRSM wealth vectors using Hotelling's T^2 results in an overall significance level of 0.0001. This level of significance indicates that equality of the LRSM and SRSM wealth vectors may be rejected with 99.99 percent probability of being correct. Pairwise analysis of attribute performance results in the following confidence intervals.

TABLE V

CONFIDENCE INTERVALS - LRSM VS SRSM

CCF:	[141.5261973; 769.6071359]
PM:	[0.111002627; 0.355430693]
MSM:	[0.020278223; 0.476551170]
CMM:	[0.142351908; 0.376448092]

Since the intervals for all four measures are on the positive side of zero, performance of the LRSM groups is termed superior over all attributes to performance by the SRSM groups. Thus, Hypothesis III cannot be rejected. The results of this study indicate that if a PE system using a single short-run profitability measure of performance is used, the welfare of the owner will be greater if a long-run appointment is anticipated than if a short-run appointment is anticipated.

Hypothesis IV

Hypothesis IV is stated as follows:

> If a PE system using a composite of short-run and long-run attributes is used, there will be no difference in the welfare of the owner resulting from anticipation of a long-run or a short-run appointment.

A pairwise comparison of the LRMM and SRMM wealth vectors using Hotelling's T^2 results in an overall significance level of 0.1599. Since this level of significance surpasses the adopted rejection range (90 percent), the hypothesis of the equality of the two vectors cannot be rejected. Thus, evidence in this experiment supports the hypothesis that a PE system using a composite of short-run and long-run attributes results in no difference in the welfare of the owner resulting from anticipation of a long-run or a short-run appointment.

Summary

The experimental results of this study are discussed in this chapter. Three types of analysis are employed: behavioral, graphical, and quantitative (statistical).

Several behavioral activities relating to the four hypotheses are discussed. One activity related to the use of different PE systems

is the time difference required to make decisions. Participants under the single measure PE system made periodic expenditure decisions very quickly. Those under the multiple measure PE system required much more time for their decisions. This behavior lends credence to the premise that a multiple measure PE tystem focuses attention on more attributes (and thus takes more time) than does the single measure PE system.

When the type of PE system employed is held constant and the anticipated job tenure is varied, two interesting behavioral activities result. The first relates to the concern expressed by the participants in relation to the terminal points of their activities. This concern suggests that decisions made when termination is imminent differ from those made when tenure is continuing.

The second activity deals with the reactions of participants in the short-run environments. In these environments, the initial operations were halted after the fourth quarter. Participants were then told the expected promotions were unavailable and that they would continue in their present positions for another 6-8 quarters. Because of the milking or short-run decisions made initially by those in the SRSM environment, the firms were not in a good position for continued operation. The participants in this environment seemed angry and confused at the prospects of continued operation.

In contrast, those participants in the SRMM environment seemed almost unconcerned with the prospect of continued operations. This behavior supports the hypothesis that the multiple measure PE system focuses attention on long range decisions and stops any milking of the firm in a short-run tenure setting.

Graphical analysis of the average quarterly expenditures is presented. Of interest is the fact that the levels of expenditures

in the multiple measure setting were, for all three expenses, higher than those under the single measure settings. Consequently graphical analysis of the four performance measures (PM, MSM, RORM, and CMM) indicates superior performance by the multiple measure groups over the period of operation.

Quantitative analysis is done with regard to each of the four hypotheses stated in this experiment. An average wealth vector consisting of compounded cash flow (CCF), cumulative production efficiency at period 10 (PM), market effectiveness (MSM), and productivity maintenance (CMM) is constructed for each of the four environmental groups (LRSM, LRMM, SRSM, SRMM). A Hotelling's T^2 test is conducted for selected pairwise comparisons of the four groups. For those comparisons with a significant difference of 10 percent or less, a pairwise comparison of each of the four attributes using 90 percent Bonferroni confidence intervals is conducted in an attempt to identify the cause(s) of the differences. Results of this analysis are found in Table VI.

TABLE VI

ANALYSIS SUMMARY

Group Comparison	Overall Significance Level	Attributes Causing Difference	Direction of Preferance
LRSM-LRMM	0.0001	MSM,CMM	LRMM > LRSM
SRSM-SRMM	0.0001	CCF,PM,MSM,CMM	SRMM > SRSM
LRSM-SRSM	0.0001	CCF,PM,MSM,CMM	LRSM > SRSM
LRMM-SRMM	0.1594	---	---

Thus, the experimental evidence suggests that use of the multiple measure PE system results in wealth to the owner which is superior to that from the use of the single measure PE system regardless of the anticipated job tenure duration. Also, an anticipated long-term tenure results in superior wealth when compared with a short-term tenure under the single measure PE system. However, under the multiple measure PE system, no difference in owner's wealth due to anticipated job tenure is observed.

CHAPTER VIII

SUMMARY AND RECOMMENDATIONS

The purpose of this chapter is to present a brief summary of the objective of this study. A summary of the background, theoretical development and methodology employed with regard to this objective is presented. This chapter also states the results of this examination and suggests possible areas for additional research.

Summary

The objective of this study is to examine the behavioral effects that result from managerial performance evaluation. Of specific interest is whether the measure(s) by which performance is evaluated and the duration of anticipated job tenure result(s) in different decisions (actions) by those being evaluated. This examination is limited to a divisional manager position in a profit making setting.

Examination of prior research indicates almost exclusive use of short-run profit as the measure by which managerial performance has been evaluated. These studies also indicate that fixation of a single short-run profit objective often leads to dysfunctional behavior or behavior which is detrimental to the long-run success of the firm.

As a reaction to the dysfunctional behavior resulting from the use of short-run profitability as a single PE measure, a few firms have attempted to employ multiple measure for PE which include both

short-run and long-run factors. The purpose of including long-run factors in the evaluation process is to focus attention on attributes which are essential for the long-run success of the firm. It is this issue of the use of a single versus a multiple PE system when anticipated job tenure is short-term or long-term that this study addresses.

Chapter III presents a theoretical framework for PE. Results of the development suggest that the purpose of PE with regard to the owner of the firm is to influence the manager to make decisions which will maximize the benefit (utility) of the owner. Maximization of wealth in the traditional economic sense of well-offness is presented as a suitable proxy for benefit (utility) maximization.

The theoretical development suggests that in a single period environment, maximization of short-run profit is equivalent to maximization of wealth. When the analysis moves to a multiple period setting, short-run profit is found to be an incomplete measure of wealth because it fails to exact attention on long-run economic factors.

As a result of this discussion, the following four hypotheses are suggested for testing in this study:

Hypothesis I: If a manager anticipates a long-run appointment, there will be no difference in the welfare of the owner resulting from the use of a PE system using the single attribute of short-run profitability or a PE system using a composite of short-run and long-run attributes.

Hypothesis II: If a manager anticipates a short term appointment, the welfare of the owner will be greater if a multiple composite measure PE system consisting of both short and long term attributes is used than if a PE system using a single profitability measure of performance is employed.

Hypothesis III: If a PE system using a single short-run profitability measure of performance is used, the welfare of the owner will be greater if a long-run appointment is anticipated than if a short-run appointment is anticipated.

Hypothesis IV: If a PE system using a composite of short-run and long-run attributes is used, there will be no difference in the welfare of the owner resulting from anticipation of a long-run or a short-run appointment.

Chapters IV, and V discuss the four basic activities necessary in establishing a multiple composite measure for use in PE. These activities consist of choosing the attributes to be included in the measure, scaling the attributes to transform physical levels of performance into measures of value or suitability, weighting the various attributes or allocating value to each attribute with regard to importance in the overall measure and determining a decision rule by which to combine the multiple measures into the composite measure.

Analysis and discussion in each of these four areas results in the multiple composite PE measure used in this study. This measure includes the attributes of production efficiency, market effectiveness, profitability and productive maintenace. These attributes are scaled on an interval value level and weighted equally. The weighting summation model is employed to create the composite measure.

Chapter VI reviews the methodology used in this study to examine the hypothesis of interest. An experimental simulation is employed in this study. This is facilitated through a computerized business game involving periodic decisions on three classes of expenditures: capacity maintenance; productive efficiency; and market effectiveness. One hundred and twenty subjects were randomly assigned to four

environmental settings: long-run tenure with short-run profit as the single PE measure; long-run tenure with the developed multiple composite PE measure; short-run tenure with short-run profit as the single PE measure; and short-run tenure with the developed multiple composite PE measure.

Each individual in each environmental setting made quarterly expenditure decisions for a total of 10 quarters. Analysis then centers on whether differences in performance exist between individuals operating under the four environmental settings.

Chapter VII presents an analysis of the experimental results. A wealth vector consisting of four attributes (discounted cash flow (CCF), production efficiency (PM), market effectiveness (MSM), and productivity maintenance (CMM)) is constructed for each of the four environmental groups (LRSM, LRMM, SRSM, SRMM). Pairwise comparisons of the various wealth vectors indicated by the four hypotheses is conducted using Hotelling's T^2. Further analysis of significant overall difference is done using a 90 percent Bonferroni confidence intervals for each of the four attributes.

Results of this analysis (summarized in Table VI, p. 95) indicate that a multiple composite PE system constructed of both short- and long-term factors is superior in all tenure settings to a single measure PE system based on short-run profit in maximizing the wealth of the firm. The results also indicate that if a single measure PE system based on short-run profitability is used, performance when long-term tenure is anticipated is far superior to performance when short-run tenure is anticipated. When a multiple composite PE system constructed of both short- and long-term factors is used, anticipated job tenure has no significant affect on performance.

Recommendations

In the course of this study, several areas for future research have been suggested. Among these are the following:

1. Examination of these same issues in a not-for-profit situation or in a lower level managerial position. The issue of a single measure PE system (perhaps not of strict profitability but some other single focus) versus a multiple measure PE system as well as the effect of anticipated job tenure is of interest in more than just the case of a divisional manager in a profit making setting.
2. Examination of the use of alternative reward structures. Perhaps reward tied directly to performance on individual attributes instead of ranking among peers would yield different results. Use of a nonlinear reward function or nonmonetary rewards exclusively could be examined.
3. Examination of different economic settings. This study assumed only an expanding market. Similar analysis in a steady or contracting market might be of interest.
4. Examination of the use of an alternative setting procedure, weighting technique and/or decision rule in constructing the multiple composite measure.
5. Extension of the analysis in this study to an actual business setting using managerial individuals as subjects.

Research in any or all of these areas might be of interest in examining the issue of performance evaluation.

SELECTED REFERENCES

Abdel-khalik, A. R. "On the Efficiency of Subject Surrogation in Accounting Research." The Accounting Review, 49, 4 (1974), pp. 743-750.

Agryis, C. "Human Problems with Budgets." Harvard Business Review, 31 (1953), pp. 97-110.

Anthony, Robert N. and John Dearden. Management Control Systems. 4th Ed. Homewood, Illinois: Richard D. Irwin, Inc., 1980.

Atkin, R. S. and E. J. Conlon. "Behaviorally Anchored Rating Scales: Some Theoretical Issues." Academy of Management Review, 3 (1978), pp. 119-128.

Boss, B. M. "Ultimate Criteria of Organizational Worth." Personnel Psychology, 5 (1952), pp. 156-173.

Beer, M., R. Ruh, J. A. Dawson, and M. J. Kavanagh. "A Performance Management System: Research, Design, Introduction and Evaluation." Personnel Psychology, 31 (1978), pp. 505-535.

Berliner, J. The Innovation Decision in Soviet Industry. Cambridge, Mass.: M.I.T. Press, 1976.

Blau, P. M. The Dynamics of Bureaucracy. Chicago, Ill.: University of Chicago Press, 1955.

Brogden, H. E. and E. K. Taylor. "The Dollar Criterion--Applying the Cost Accounting Concept to Criterion Construction." Personnel Psychology, 3 (1950), pp. 133-167.

Caplan, E. H. Management Accounting and Behavioral Science. Reading, Mass.: Addison-Wesley, 1971.

Cattell, R. B. Personality and Motivation: Structure and Measurement. New York: Harcourt, Brace, World, 1957.

Charnes, A. and A. Stedry. "Exploratory Models in the Theory of Budget Control." In W. W. Cooper, H. J. Leavitt and M. W. Shelly, Jr., (Eds.), New Perspectives in Organization Research. New York: John Wiley and Sons, 1964 , pp. 212-249.

_____. "Search-Theoretic Models of Organization Control by Budgeted Multiple Goals." Management Science, 12, 5 (1966), pp. 457-482.

Copeland, R. M, A. J. Francia, and R. H. Strawser. "Students as Subjects in Behavioral Business Research." The Accounting Review, 48, 2 (1973), pp. 365-372.

Dawes, R. "Predictive Models as a Guide to Preference." IEEE Trans. Systems, Man, and Cybernetics. Vol. SMC-7, 5 (May, 1977), pp. 355-357.

Dawes, R. and R. Corrigan. "Linear Models in Decision Making." Psychological Bulletin, 81, 2 (February, 1974), pp. 95-106.

Demski, J. S. "The Decision Implementation Interface: Effects of Alternative Performance Measurement Models." The Accounting Review, 45, 1 (1970), pp. 76-87.

_____. "Implementation Effects of Alternative Performance Measurement Models in a Multivariable Context." The Accounting Review, 46, 2 (1971), pp. 268-278.

_____. "Optimal Performance Measurement." Journal of Accounting Research, 10 (Autumn, 1972), pp. 243-258.

DeNisi, A. S. and J. L. Mitchell. "Analysis of Peer Ratings as Predictors and Criterion Measures and A Proposed New Application." Academy of Management Review, 3 (1978), pp. 369-374.

Dickhaut, J. W., J. L. Livingstone, and D. J. Watson. "On the Use of Surrogates in Behavioral Experimentation." Report of the Committee on Research Methodology in Accounting, The Accounting Review, Supplement to Vol. 47 (1972), pp. 455-471.

Downey, R. G. "Associate Evaluations: Nominations vs. Ratings." U. S. Army Research Institute for the Behavioral and Social Sciences, 1974.

Drucker, P. The Practice of Management. New York: Harper & Brothers, 1954.

Cunnette, M. D. "A Modified Model for Test Validation and Selection Research." Journal of Applied Psychology, 47 (1963), pp. 317-323.

Dyer, J. S. and R. K. Sarin. Measurable Multiattribute Value Functions. Discussion Paper No. 66, Management Science Study Center, Graduate School of Management, University of California, Los Angeles, California, June, 1977.

Einhorn, H. J. and W. McCoach. "A Simple Multiattribute Utility Procedure for Evaluation." Behavioral Science, 22 (1977), pp. 270-282.

Fischer, G. W. Multi-Dimensional Value Assessment for Decision Making. University of Michigan, Engineering, Psychology Laboratory, Technical Report, 1972.

Fishburn, P. C. "Methods of Estimating Additive Utilities." Management Science, 13, 7 (1967), pp. 435-453.

_____. Utility Theory for Decision Making. New York: John Wiley, 1970.

Ghiselli, E. E. "Dimensional Problems of Criterion." Journal of Applied Psychology, 40 (1956), pp. 1-4.

Granick, D. Management of the Industrial Firm in the USSSR. New York: Columbia University Press, 1954.

Goicoechea, A., D. R. Hansen and L. Duckstein. Multiobjective Decision Analysis With Engineering and Business Applications. New York: John Wiley and Sons, forthcoming.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.

Guilford, J. and H. Dingman. "A Valuation Study of Ratio-Judgment Methods." American Journal of Psychology, 67, 3 (September, 1954), pp. 395-410.

Guion, R. M. "Criterion Measurement and Personnel Judgment." Personnel Psychology, 14 (1961), pp. 141-149.

Halper, J. A., D. M. O'Regan and W. R. Miller. Computerized Methods in Power Plant Siting Studies. Second Annual University of Missouri-Missouri Energy Council Conference on Energy, Cranford, N. J.: Dames and Moore, 1975.

Hobbs, B. F. "Analytical Multiobjective Decision Methods for Power Plant Siting: A Review of Theory and Application." Division of Regional Studies, National Center for the Analysis of Energy Systems, Brookhaven National Laboratory, Upton, New York, August, 1979.

Huber, G. P. "Methods of Quantifying Subjective Probabilities and Multi-Attribute Utilities." Decision Sciences, 5 (1974), pp. 430-458.

Irish, R. R. "The Measurement of Divisional Performance in Terms of Accounting Data." (Unpub. Ph.D. dissertation, Library, The University of Texas at Austin, 1970.)

Jensen, M. and W. Meckling. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." Journal of Financial Economics, 3 (1976), pp. 305-360.

Keeney, R. L. and H. Raiffa. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: John Wiley and Sons, 1976.

Latham, G. P., C. H. Fay and L. M. Saari. "The Development of Behavioral Observation Scales for Appraising the Performance of Foremen." Personnel Psychology, 32 (1979), pp. 299-311.

Luce, R. D. and H. Raiffa. Games and Decisions. New York: John Wiley and Sons., 1957.

Matz, A. and M. F. Usry. Cost Accounting: Planning and Control. 6th Ed. Cincinnati, Ohio: Southwestern Publishing Co., 1976.

Mayfield, E. C. "Peer Nominations in the Life Insurance Industry." In R. G. Downey and E. F. Medland (Chairs.), "Peer Ratings: Beyond Validity Studies," paper presented at the American Psychological Association Convention, Chicago, 1975.

McGrath, J. E. "Toward a 'Theory of Method' for Research in Organizations." In W. W. Cooper, H. J. Leavitt, and M. W. Shelly, Jr., (Eds.), New Perspectives in Organization Research. New York: John Wiley and Sons, 1964.

Miner, J. B. "Management Appraisal: A Capsule Review and Current References." Business Horizons, 11, 5 (1968), pp. 83, 84, 86-96.

Nagle, B. F. "Criterion Development." Personnel Psychology, 6 (1953), pp. 271-289.

Newman, W. H. and J. P. Logan. Management of Expanding Enterprises. New York: Columbia University Press, 1955.

Pratt, J. W., H. Raiffa and R. O. Schlaifer. Introduction to Statistical Decision Theory. New York: McGraw-Hill, 1965.

Ridgway, V. F. "Dysfunctional Consequences of Performance Measurements." Administrative Science Quarterly, 1 (1956), pp. 240-247.

Rosen, L. S. and R. E. Schneck. "Some Behavioral Consequences of Accounting Measurement Systems." Cost and Management, 41, 9 (1967), pp. 6-16.

Ross, S. A. "The Economic Theory of Agency: The Principal's Problem." American Economic Review, Vol. LXII (May, 1973), pp. 134-139.

Rutherford, G., et al. "Goal Formulation for Socio-Technological Systems." ASCS Proceedings, Journal of Urban Planning and Development Division (September, 1972), pp. 157-169.

Sarin, R. K., J. S. Dyer, and K. Nair. "A Comparative Evaluation of Three Approaches for Preference Assessment." Woodward-Clyde Consultants, San Francisco, California, presented at 1979 ORSA/TIMS Joint Meeding, Honolulu, Hawaii, May, 1979.

Savage, L. J. The Foundations of Statistics. New York: John Wiley and Sons, 1954.

Schmidt, F. L. and L. B. Kaplan. "Composite vs. Multiple Criteria: A Review and Resolution of the Controversy." Personnel Psychology, 24 (1971), pp. 419-434.

Schneider, B. J. "Implicit Personality Theory: A Review." Psychological Bulletin, 79 (1973), pp. 294-309.

Slovic, P. and S. Lictenstein. "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment." Organizational Behavior and Human Performance, 6 (1971), pp. 649-744.

Solomons, D. Divisional Performance: Measurement and Control. Homewood, Ill.: Richard D. Irwin, 1965.

Stedry, A. Budget Control and Cost Behavior. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1960.

Steers, R. M. "Problems in the Measurement of Organizational Effectiveness." Academy of Management Journal, 18 (December, 1975), pp. 546-547.

Stevens, S. S. "On the Theory of Scales of Measurement." Science, 103 (June, 1946), pp. 677-680.

Swieringa, R. J. and J. S. Demski. "Performance Measure and Incentive Alternatives in a Multivariate Setting." In Behavioral Experiments in Accounting, Thomas J. Burns (Ed.), Columbus, Ohio: Ohio State University, 1962.

Taylor, F. W. The Principles of Scientific Management. New York and London: Harper and Brothers, 1915

Tell, B. A Compariative Study of Some Multiple Criteria Methods. Economic Research Institute, Stockholm School of Economics, Stockholm, Sweden, 1976.

Thorndike, R. L. Personnel Selection. New York: John Wiley and Sons, 1948.

Thurstone, L. The Measurement of Values. Chicago: University of Chicago Press, 1959.

Toops, H. A. "The Criterion." Educational and Psychological Measurement, 4 (1944), pp. 271-297.

Vogel, David. "America's Management Crisis." The New Republic (February 7, 1981), pp. 21-23.

von Neumann, J. and O. Morgenstern. Theory of Games and Economic Behavior. 2nd Ed. Princeton, N.J.: Princeton University Press, 1947.

Wainer, H. "It Don't Make No Nevermind." Psychological Bulletin, 82, 2 (1976), pp. 213-217.

Warren, E. K. Long Range Planning: The Executive Viewpoint. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1966.

Williamson, O. E. The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

Yntema, D. B. and W. S. Torgerson. "Man-Computer Cooperation in Decisions Requiring Common Sense." I.R.E. Transactions on Human Factors in Electronics, 3 (1961), pp. 20-26.

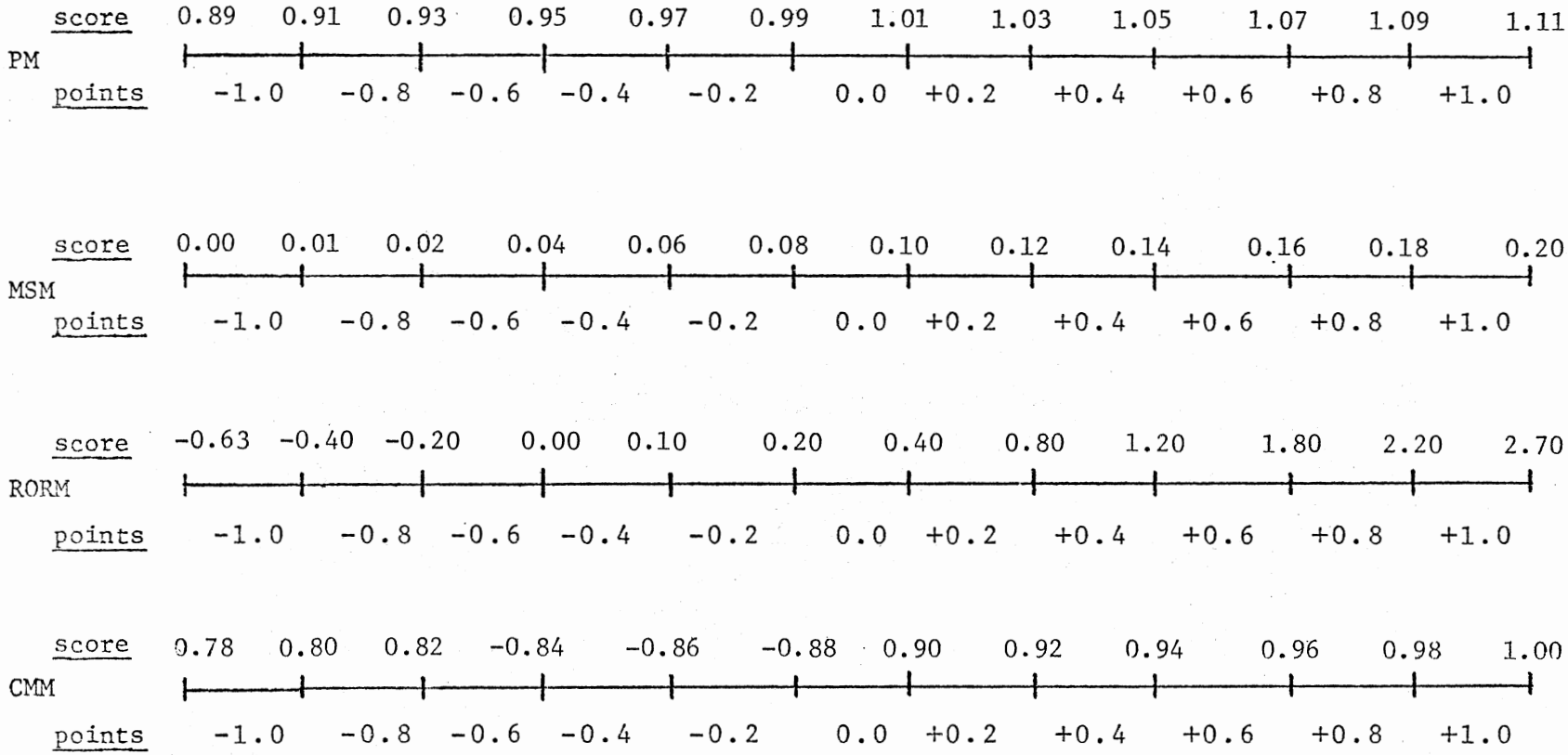
APPENDICES

APPENDIX A

MULTIPLE COMPOSITE MEASURE FORMS

Exhibit IV is an example given to each of the participants in the multiple measure groups to explain the process of evaluating quarterly performance. The example is based on the activity of the fourth quarter of 19X0 found in The KDSS Game in Appendix B. Exhibits I, II and III are found in the KDSS game. The other forms following Exhibit IV are used in the appropriate quarter throughout the operation of the game by those in the multiple measure groups. Note that the scale ranges for MSM and RORM change each quarter.

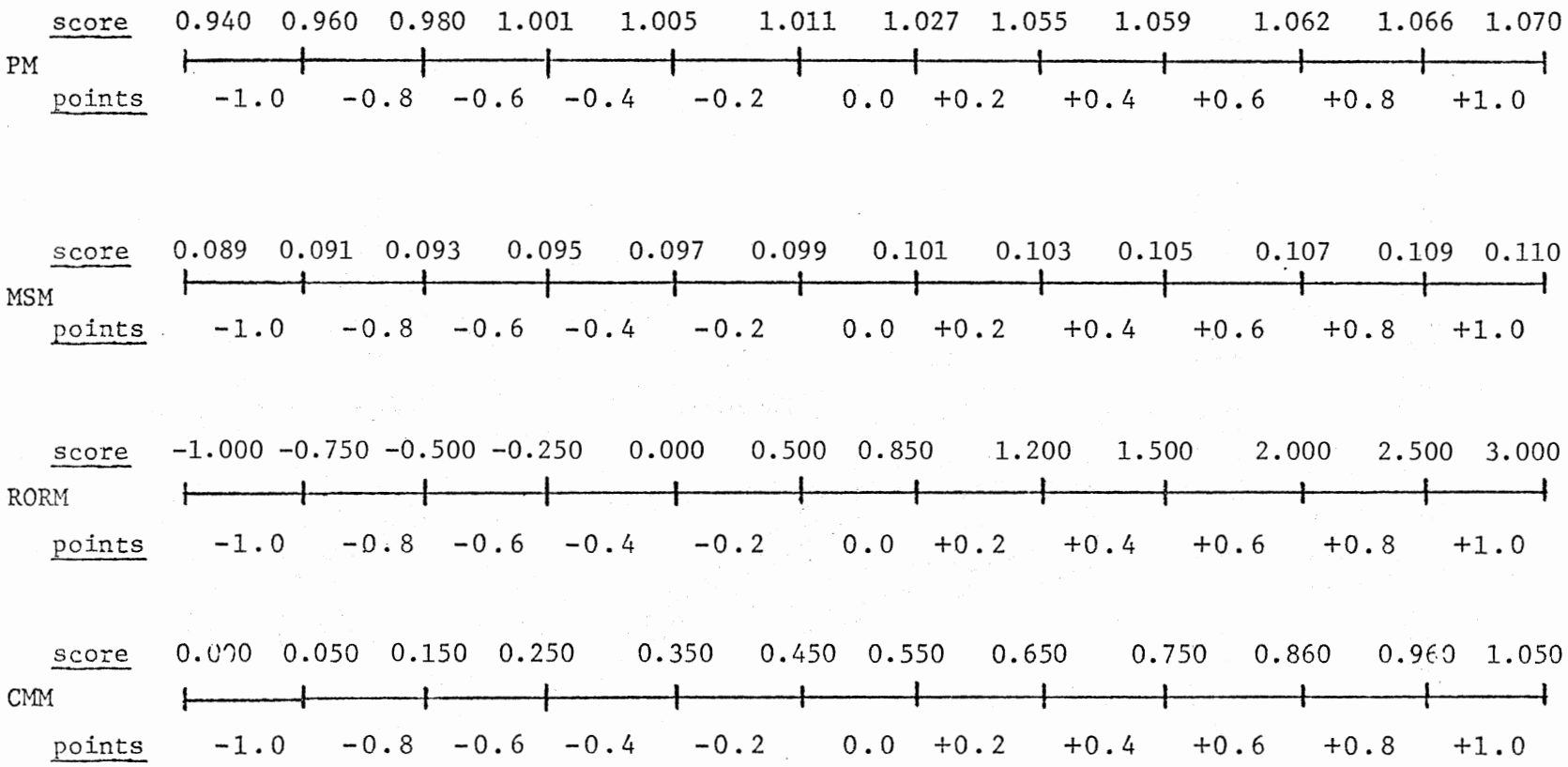
Exhibit IV



Scores for Example I, Quarter 4, Year 19X0

	Printout Score	Points
PM	1.000	0.0
MSM	0.100	0.0
RORM	1.000	0.4
CMM	1.000	1.0
Total		1.4

Quarter 1 Year X1



	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 2 Year X1

PM	score	0.940	0.960	0.980	1.001	1.005	1.011	1.027	1.055	1.059	1.062	1.066	1.070
	points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

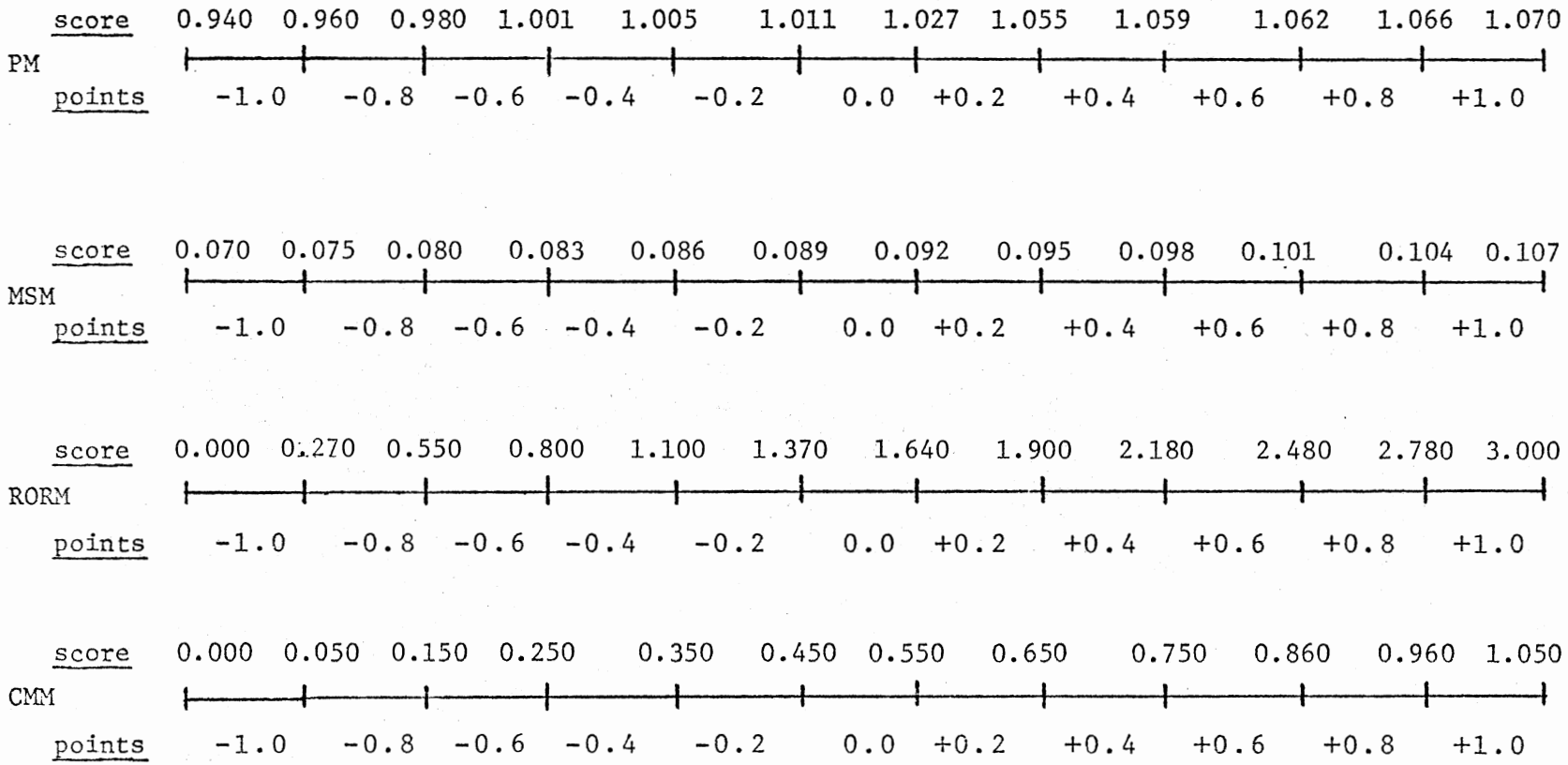
MSM	score	0.085	0.087	0.089	0.091	0.093	0.095	0.097	0.099	0.101	0.103	0.105	0.107
	points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

RORM	score	-1.000	-0.500	0.000	0.200	0.400	0.700	1.000	1.350	1.700	2.100	2.500	3.000
	points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

CMM	score	0.000	0.050	0.150	0.250	0.350	0.450	0.550	0.650	0.750	0.860	0.960	1.050
	points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

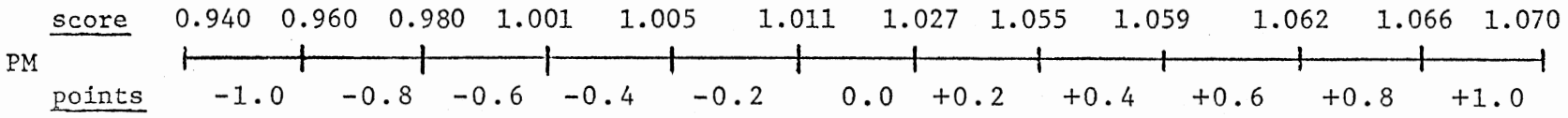
	Score	Points
PM	_____	_____
MSM	_____	_____
RORM	_____	_____
CMM	_____	_____
Total		_____

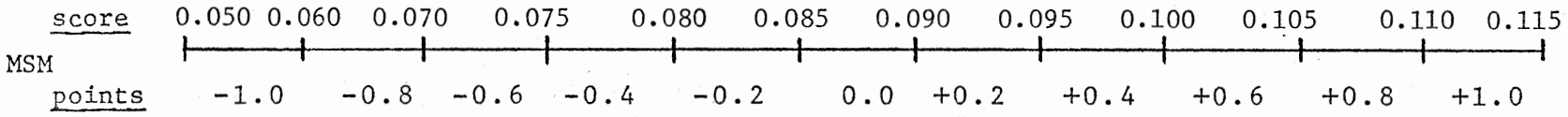
Quarter 3 Year X1

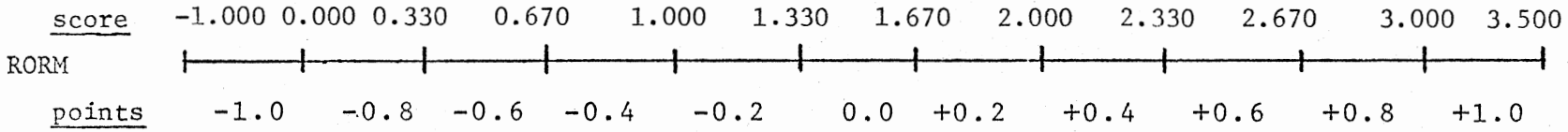


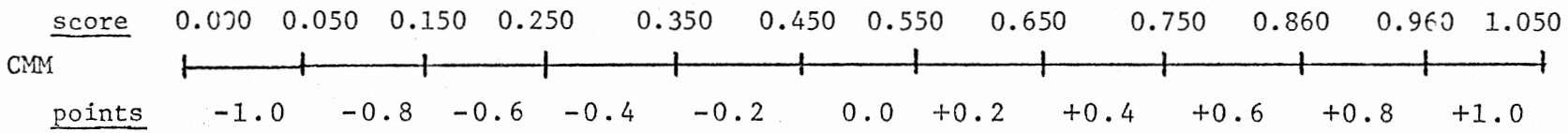
	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 4 Year X1



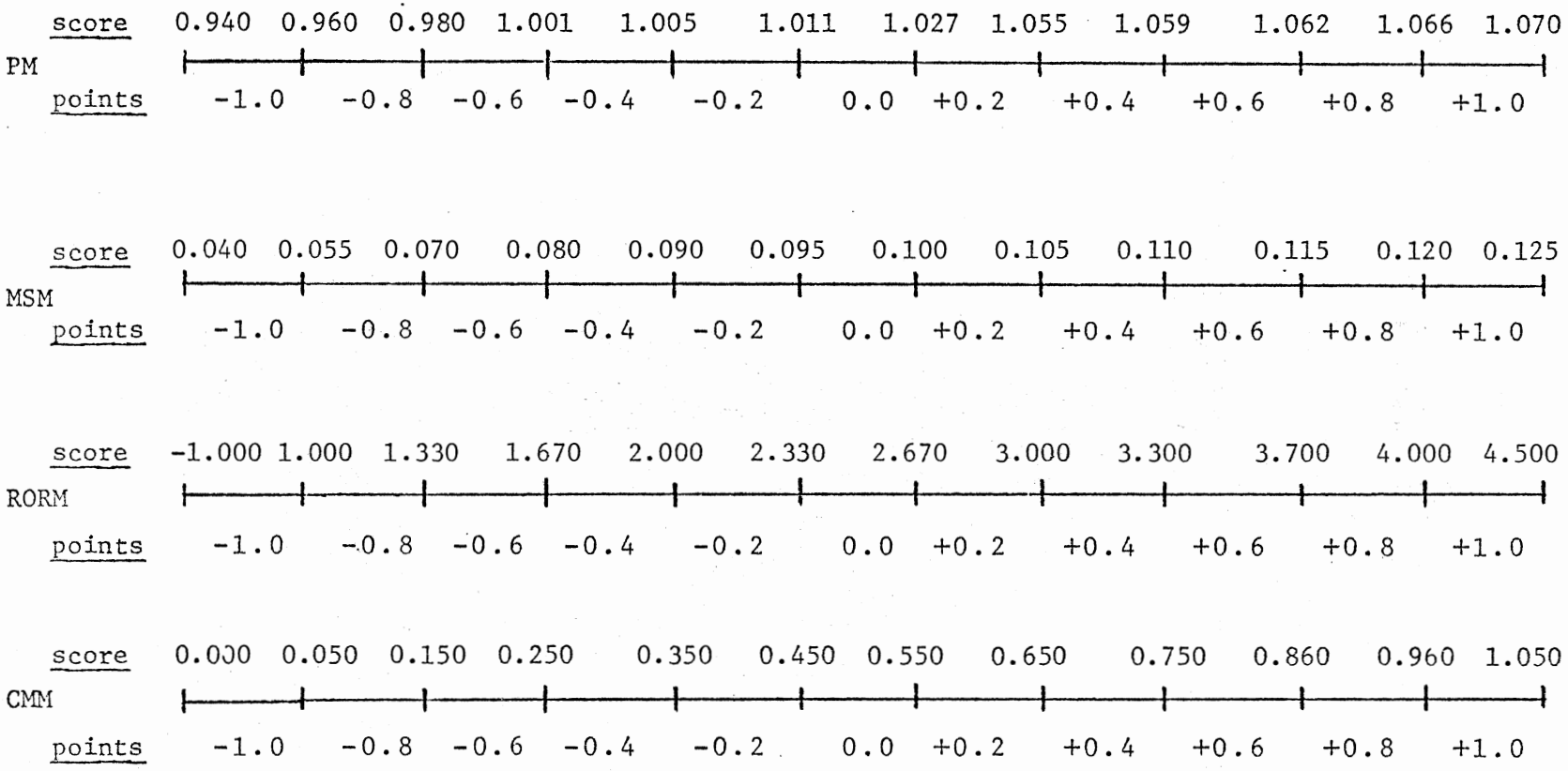






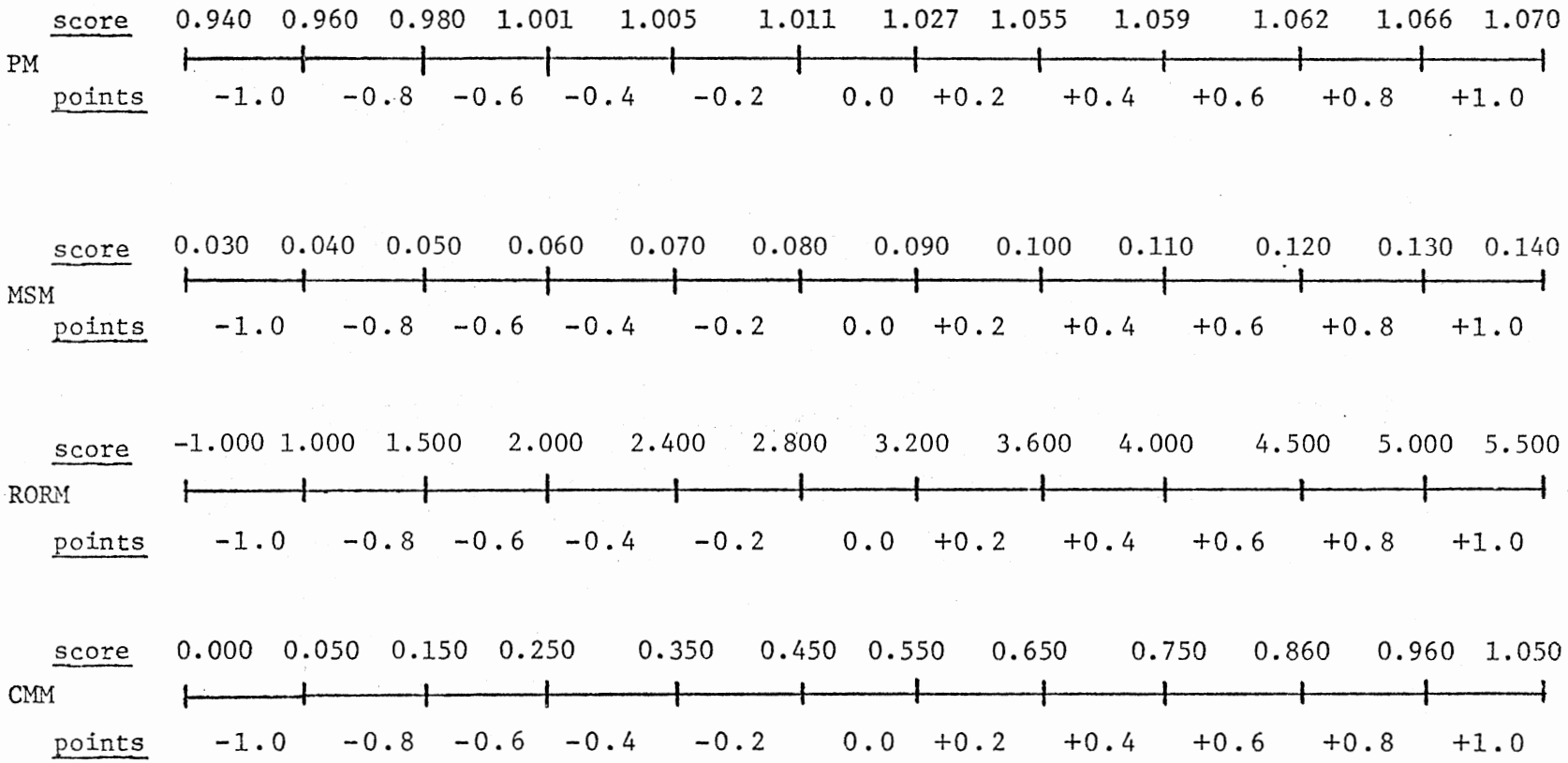
	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 1 Year X2



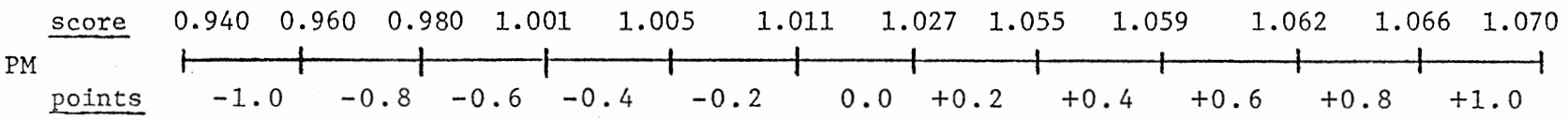
	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

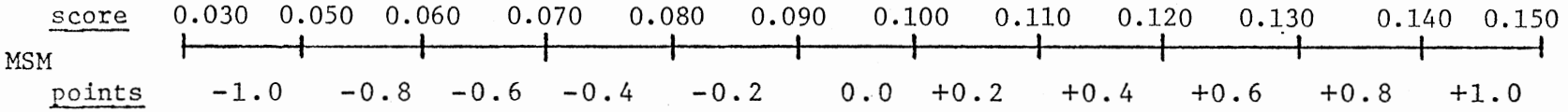
Quarter 2 Year X2

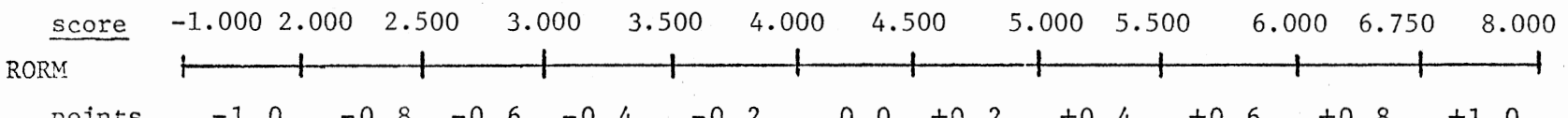


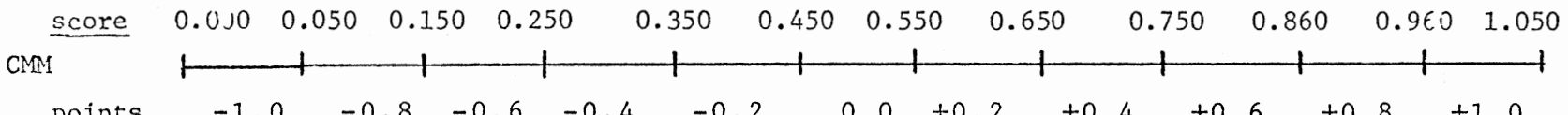
	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 3 Year X2



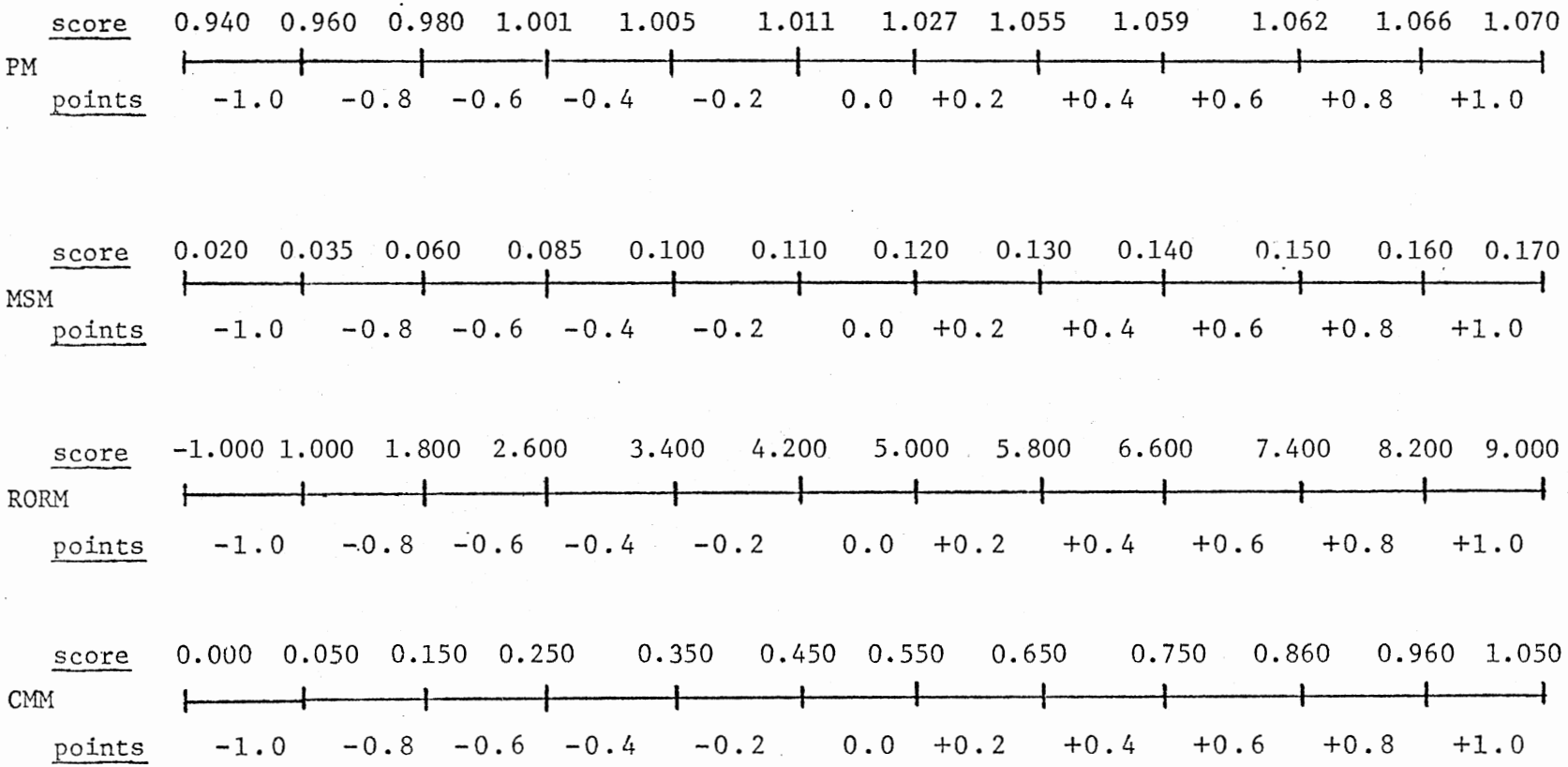






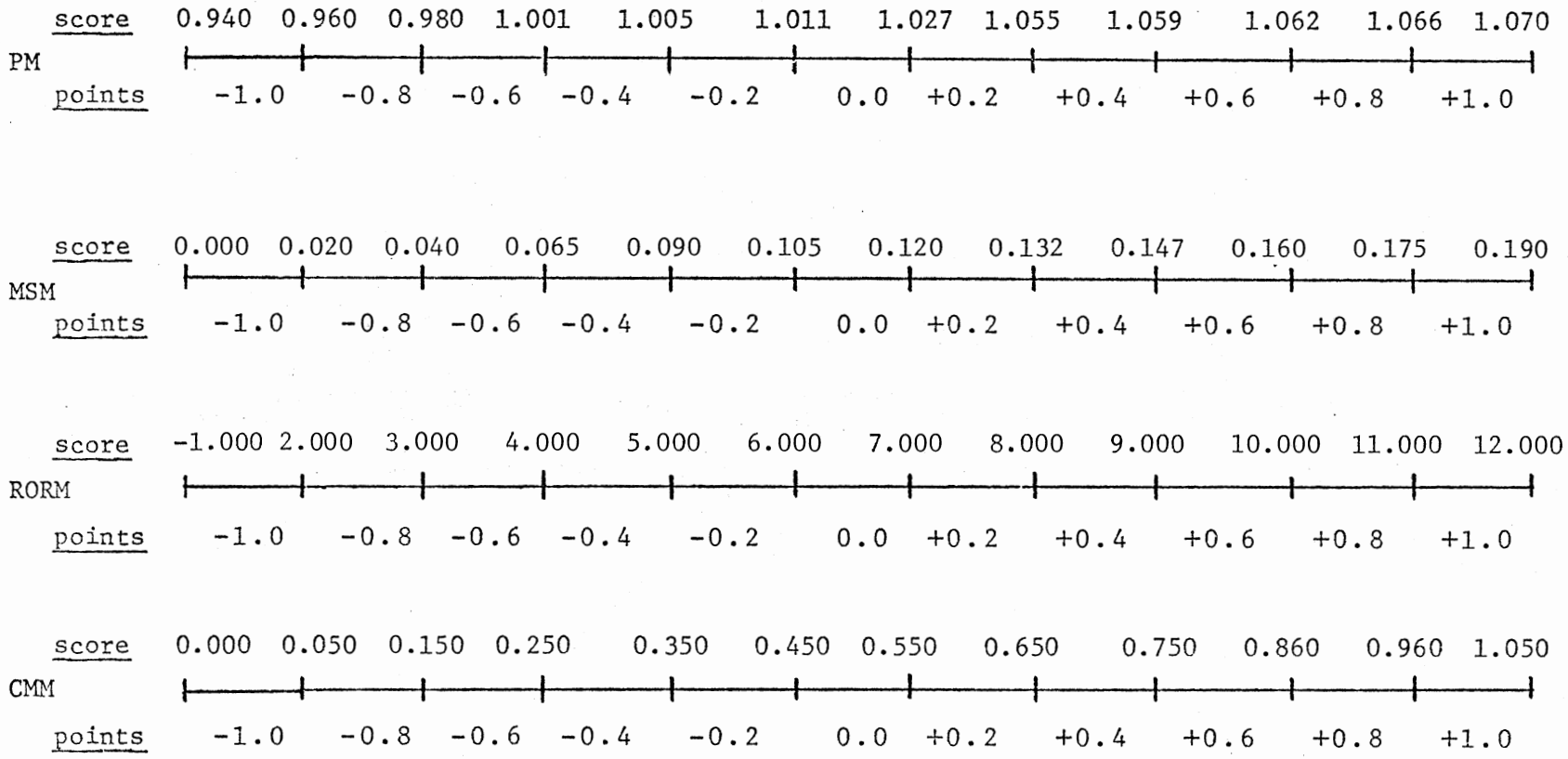
	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 4 Year X2



	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 1 Year X3



	Score	Points
PM		
MSM		
RORM		
CMM		
Total		

Quarter 2 Year X3

PM score	0.940	0.960	0.980	1.001	1.005	1.011	1.027	1.055	1.059	1.062	1.066	1.070
PM points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

MSM score	0.000	0.020	0.040	0.060	0.080	0.100	0.115	0.130	0.150	0.190	0.190	0.210
MSM points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

RORM score	-1.000	1.000	2.500	4.000	5.500	7.000	8.250	9.500	10.750	12.000	13.500	15.000
RORM points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

CMM score	0.000	0.050	0.150	0.250	0.350	0.450	0.550	0.650	0.750	0.860	0.960	1.050
CMM points	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	+0.2	+0.4	+0.6	+0.8	+1.0	

	Score	Points
PM	_____	_____
MSM	_____	_____
RORM	_____	_____
CMM	_____	_____
Total		_____

APPENDIX B

THE KDSS BUSINESS GAME

The KDSS game was developed specifically for use in this study. This appendix provides information concerning the mechanical operation of the game. Three types of expenditures are developed and explained. Output of the game consists of a quarterly income statement and four performance measures. An example of the operation of quarter 4 of 19X0 is given. Also, an indication of the general effect of different expenditure levels is given.

KDSS Game

Company Background

The Hansen Company is part of the very competitive electronic games industry. The Tulsa Unit is assigned the production and development of the hand held sports games. Due to restrictions beyond the firm's control, all the games from the Tulsa Unit sell for $20 each.

Tulsa Unit's initial manufacturing cost is $10/unit. Prior to each quarter's production, decisions must be made with regard to three classes of expenditures.

1. The first class of expenditures consists of those which impact upon production capacity. The quarterly expense on production capacity is denoted as E1 and includes expenditures such as maintenance, capital budgeting, etc.

2. The second class of expenditures consists of those which impact upon production efficiency. The quarterly expense on production efficiency is denoted as E2 and includes such expenditures as Research and Development, employee training, etc.

3. The third class of expenditures consists of those which impact upon market position. The quarterly expense on market position is denoted as E3 and includes such expenditures as advertising, public relations, quality control, etc.

The dollar expenditures for each of the three classes (E1, E2, and E3) will be subtracted from the gross profit in arriving at net profit. The effect of each of these expenditures on sales volume and/or manufacturing costs is described below.

Due to the high demand for these games, the Tulsa Unit maintains no inventory. All units produced in a quarter are sold during that quarter. No unfinished work in process nor any finished goods remain on hand at the end of any quarter.

A brief description of the variables from which you as the manager of the Tulsa Unit must make quarterly expense decisions follows.

Basic Business Setting

Hansen's economic forecasters have reported that business volume is expected to increase by approximately 2% per quarter for the next four years. Initial sales volume is 25,000 units per quarter. Results for activity during the fourth quarter of 19X0 are found in Exhibit I.

Decision Variables

Expense 1

The Tulsa Unit has a practical production capacity of 100,000 units. With a production capacity expense, E1, of $100,000 per quarter, practical capacity may be maintained. You, as manager of the Tulsa Unit must decide upon the level of E1 expense incurred. This expense ranges from $20,000 to $100,000 with increases in increments of $20,000. While the quarterly $100,000 expense will maintain capacity at $100,000 units, smaller expenditures will cause a decrease in practical capacity. The effect of quarterly expenditures is felt in the quarter following the expense and is cumulative in nature.

Expense 2

Production efficiency expenditures, E2, range from $10,000 to $100,000 with increases in increments of $10,000. E2 affects the quantity of inputs which are used to manufacture the products which the Tulsa Unit sells. Initially, 100 input units at $.10 each are used to construct one finished output unit. An expenditure of approximately $40,000 will maintain the amount of inputs for each unit of output produced. A lower expenditure will raise the number of required inputs (and hence the cost per unit sold) while an expenditure

greater than $40,000 will--up to a certain level--decrease the amount of inputs per output unit produced (and hence the cost per unit sold). The general trend of the increase or decrease may be found in Exhibit II. The amount of the increase or decrease in input units may range from +5% to -7% per quarter.

The effect of the quarterly E2 expense is felt the first quarter after expenditure. The quarterly effect is cumulative in nature over periods of operation.

Expense 3

The expenditure which impacts on market position, E3, ranges from $10,000 to $100,000 with increases in increments of $5,000. An expenditure greater than $30,000 will increase quarterly sales volume while an expenditure less than $30,000 will result in a decrease of quarterly sales volume. The general increase or decrease in sales volume that results from various levels of expenditure can be seen in Exhibit III. The amount of the increase or decrease in volume ranges from a possible +17% to -5% per quarter.

E3's effect on sales is not felt in the quarter of expenditure. Partial effect is felt in the first quarter after the expenditure while full effect is felt beginning in the second quarter after the expenditure. The effect of the quarterly E3 expense on sales volume is cumulative in nature.

Program Output

For each quarter, you will receive an output similar to Exhibit I. The income statement is self explanatory. Other information is as follows:

PM

PM is the productivity measure and is defined as the percent savings due to a change in productivity. The measure is calculated as follows:

$$1 + \frac{L_1/X_1 - L_2/X_2}{L_1/X_1} \quad \text{or} \quad 2 - \frac{L_2/X_2}{L_1/X_1}$$

where L_1 = Units of intput in quarter t-1,

L_2 = units of input in quarter t,

X_1 = units of output in quarter t-1, and

X_2 = units of output in quarter t.

The units of input used in calculating this measure are directly affected by the amount of quarterly production efficiency expense (E2).

MSM

MSM is the market share measure and is calculated as quarterly sales of Tulsa Unit divided by quarterly sales of the market. Initially, Tulsa Unit maintains 10% of the market. MSM is calculated as follows:

$$\frac{\text{Tulsa Unit sales in quarter i}}{\text{Market sales in quarter i}}$$

Total market sales increase approximately $500,000 per quarter. Quarterly sales volume is directly affected by the level of market position expense (E3). Sales volume may also be affected by production capacity expense (E2).

RORM

The Rate of Returm on Investment measure (RORM) is calculated as follows:

$$\frac{\text{Net income in quarter i/Investment}}{\text{Cost of capital}} \quad \text{or} \quad \frac{\text{Net income in quarter i}}{\$80,000}$$

The investment for the Tulsa Unit is $800,000. Hansen Company has calculated that the cost of capital for the firm is 10%. RORM is directly affected by the level of E1, E2, and E3 expense.

CMM

CMM is the Capacity Maintenance measure. CMM is defined as actual quarterly capacity/practical capacity. Tulsa Unit has a new plant facility with a practical production capacity of 100,000 units. The quarterly capacity level is directly affected by the level of quarterly E1 expense selected.

Exhibit I

Income Statement and Performance Measures
For Quarter 4, 19X0

(all numbers in thousands)

Player #1	Year 19X0	Quarter 4
Sales		$500
Variable Manufacturing Costs		250
Gross Profit		$250
Expenses		
Expense Class 1	100	
Expense Class 2	40	
Expense Class 3	30	
Total Expenses		$170
Net Income		$ 80

PM = 1.000 MSM = 0.100 RORM = 1.000 CMM = 1.000

Exhibit II

General Effect of Quarterly E2 Expense Levels on
Input Units Per Unit of Output

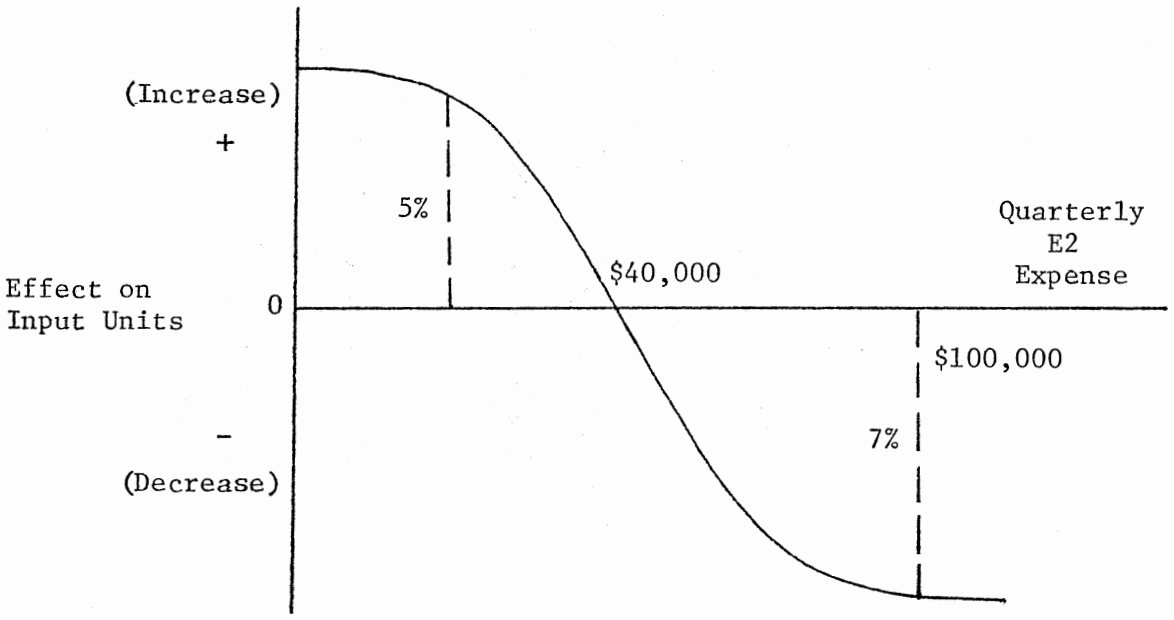
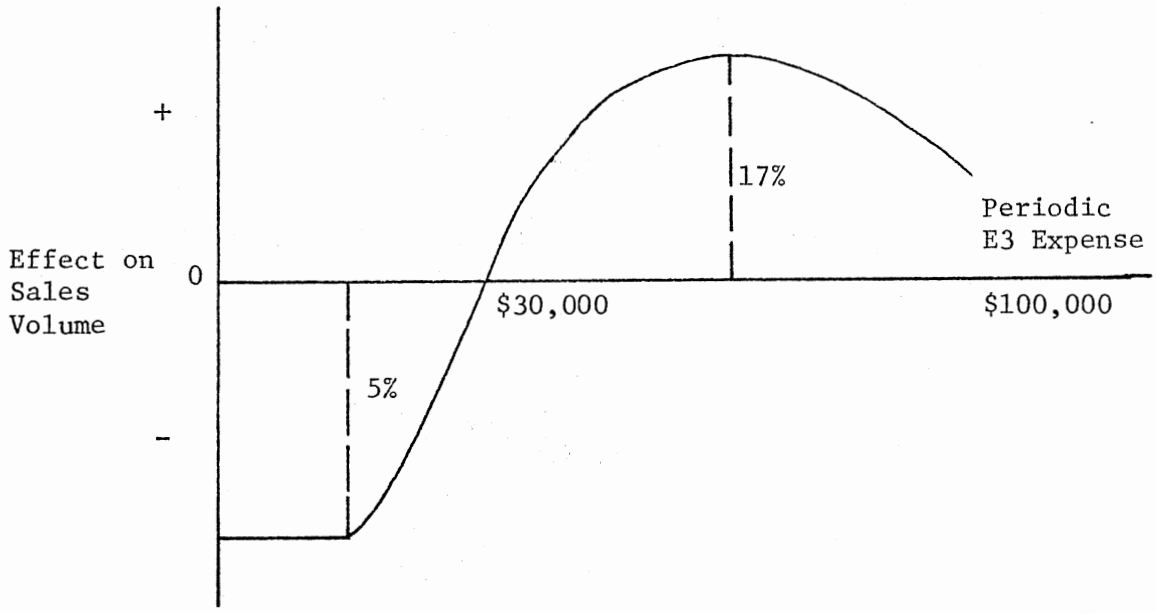


Exhibit III

General Effect of Quarterly E3 Expense
Levels on Sales Volume



Decision Form

E1--Capacity

Level	Expense
1	$ 20,000
2	40,000
3	60,000
4	80,000
5	100,000

E2--Production Efficiency

Level	Expense
1	$ 10,000
2	20,000
3	30,000
4	40,000
5	50,000
6	60,000
7	70,000
8	80,000
9	90,000
10	100,000

E3--Sales Volume

Level	Expense
1	$ 10,000
2	15,000
3	20,000
4	25,000
5	30,000
6	35,000
7	40,000
8	45,000
9	50,000
10	55,000
11	60,000
12	65,000
13	70,000
14	75,000
15	80,000
16	85,000
17	90,000
18	95,000
19	100,000

APPENDIX C

RUN AND SUMMARY SHEETS

A run sheet is used each quarter to tabulate the quarterly performance score for each group participant. Once scores are tabulated, position assignment is made and reward points are assigned.

The group summary sheet provides a summary of the points earned by each participant over the course of the experiment. This sheet also serves as a basis from which monetary allocation is made.

Run Sheet

Group # _____ Date _____ Time _____ Setting _____

Run # _____

	Score	Position	Points
Player #1			
Player #2			
Player #3			
Player #4			
Player #5			
Player #6			

Summary Sheet

Group # _____ Date _____ Time _____ Setting _____

	Name	Phone #
Player #1		
Player #2		
Player #3		
Player #4		
Player #5		
Player #6		

Points Per Player Per Run

Run #	1	2	3	4	5	6	7	8	9	10	11	12	13
Player #1													
Player #2													
Player #3													
Player #4													
Player #5													
Player #6													

APPENDIX D

INTRODUCTION AND CONSENT FORM

This form is given to each student indicating interest in participating in the study. It contains a brief written explanation of the purpose of the study as well as a consent form retained by the researcher.

Introduction and Consent Form

This study is undertaken to examine managerial performance. The KDSS game will be used in which you will be required to make decisions on three expense variables over a number of quarters. The type of industry, environment, and performance measure will be specified at the beginning of the activity. Participants will be placed into groups of six individuals. Monetary payments will be based on individual performance as evaluated by the performance measure specified. The amounts of these payments range from approximately $1 to $20.

Informed Consent by Subjects in Experiments

I, _____, have carefully read and
Print Name

fully understand the instructions for this experiment on Managerial Performance. I give my consent to serve as a subject in this requirement on _____. I am aware that I may ask
(Date)

questions or terminate the experiment at any point. I am also aware that discussing my activity in this experiment with any other persons will result in my removal as a participant without any payment.

_____
Signature

APPENDIX E

INSTRUCTION SHEET AND TRIAL RUN FORM

The instruction sheet specifies the time and place for turning in the trial run form and signing up for an experimental session. The trial run form simply provides a form for making expenditure decisions for three sequential quarters.

Introduction

The following is a description of the KDSS game, the quarterly decisions that must be made, and the results of the decisions made in the last quarter of 19X0. Please study this material carefully.

A period of time will be given for you to read this material and make a set of trial decisions for the quarters. These decisions may be turned in to Bus. 433-E on February 11-16 from 8 am to 5 pm. Results of the trial run will be available beginning the day after they are turned in. Sign up sheets for the experiment are in Bus. 433-E. The experiment will be run from February 16 through February 27.

Please place yourself in the position of a manager and approach this study in a serious and nontrivial manner. For experimental reasons, we must insist that you not discuss your involvement in this experiment with any other individual until after the experiment has been totally finished. Those found discussing their participation will be removed from the experiment with no monetary payment.

Name _____

Phone _____

Trial Run

Run	Quarter	Year	Player #	Level of Expense E1	E2	E3
00	4	X0	1	5	4	5
01	1	X1	1	____	____	____
02	2	X1	1	____	____	____
03	3	X1	1	____	____	____

Turn this in to Bus. 433-E on February 11-16 between the hours of 8 am to 5 pm. Results of your trial run will be available the next day.

APPENDIX F

EXPERIMENTAL SESSION DATES AND ENVIRONMENTAL ASSIGNMENTS

This appendix simply lists the dates and times of the 20 experimental sessions. The setting is assigned to each session is done so in a random manner.

Group #	Date	Time	Setting
1	Tuesday February 17	3 pm - 5 pm	LRSM
2	Tuesday February 17	6 pm - 8 pm	LRMM
3	Tuesday February 17	8 pm - 10 pm	LRMM
4	Wednesday February 18	3 pm - 5 pm	LRSM
5	Thursday February 19	3 pm - 5 pm	LRMM
6	Friday February 20	3 pm - 5 pm	LRSM
7	Saturday February 21	10 am - 12 noon	LRSM
8	Saturday February 21	2 pm - 4 pm	SRMM
9	Monday February 23	6 pm - 8 pm	SRMM
10	Monday February 23	8 pm - 10 pm	SRMM
11	Tuesday February 24	3 pm - 5 pm	SRSM
12	Tuesday February 24	3 pm - 5 pm	SRSM
13	Tuesday February 24	6 pm - 8 pm	SRSM
14	Tuesday February 24	6 pm - 8 pm	LRSM
15	Tuesday February 24	8 pm - 10 pm	LRMM
16	Wednesday February 25	3 pm - 5 pm	LRMM
17	Wednesday February 25	8 pm - 10 pm	SRMM
18	Thursday February 26	3 pm - 5 pm	SRSM
19	Thursday February 26	3:30 pm - 5:30 pm	SRSM
20	Friday February 27	12 noon - 2 pm	SRMM

APPENDIX G

EXPERIMENTAL DECISION FORMS

The form in this appendix is used by the participants in each group to indicate their quarterly expenditure decisions.

Decision Form

Run	Quarter	Year	Player #*	E1	E2	E3
00	4	Xo	1	5	4	5
01	1	X1	1	____	____	____
02	2	X1	1	____	____	____
03	3	X1	1	____	____	____
04	4	X1	1	____	____	____
05	1	X2	1	____	____	____
06	2	X2	1	____	____	____
07	3	X2	1	____	____	____
08	4	X2	1	____	____	____
09	1	X3	1	____	____	____
10	2	X3	1	____	____	____
11	3	X3	1	____	____	____
12	4	X3	1	____	____	____
13	1	X4	1	____	____	____
14	2	X4	1	____	____	____
15	3	X4	1	____	____	____

*1, 2, 3, 4, 5, or 6--one for each member of the group.

APPENDIX H

ENVIRONMENTAL SETTINGS

This appendix contains the four environmental settings used in this experiment.

Environment and Performance Measure

LRSM

Hansen Company has an opening for the manager of its Tulsa Unit. You have been interviewed by Mr. Hansen and have taken the job. You accept the position with the knowledge that your performance will be closely watched. This is a position you have worked for years to obtain. You now fully expect that a good performance record will keep you in this position for many years to come. In fact, due to your age and the organization of the company, you expect to remain in this position until your retirement in eight years unless you are fired. As no alternative position within or without the firm is available, to be fired would simply be a total loss of employment.

Mr. Hansen reminds you that the future of the Tulsa Unit is affected by your decisions. He trusts your ability and indicates that your performance will be evaluated quarterly, along with the performance of Hansen Company's other unit managers. Maintaining satisfactory evaluations is the key to remaining in this position. After 2-3 years, Mr. Hansen will drop by to assure himself that your performance has been satisfactory. This evaluation as well as the quarterly evaluations will be based solely on the amount of profit achieved.

Environment and Performance Measure

LRMM

Hansen Company has an opening for the manager of its Tulsa Unit. You have been interviewed by Mr. Hansen and have taken the job. You accept the position with the knowledge that your performance will be closely watched. This is a position you have worked for years to obtain. You now fully expect that a good performance record will keep you in this position for many years to come. In fact, due to your age and the organization of the company, you expect to remain in this position until your retirement in eight years unless you are fired. As no alternative position within or without the firm is available, to be fired would simply be a total loss of employment.

Mr. Hansen reminds you that the future of the Tulsa Unit is affected by your decisions. He trusts your ability and indicates that your performance will be evaluated quarterly, along with the performance of Hansen Company's other unit managers. Maintaining satisfactory evaluations is the key to remaining in this position. After 2-3 years Mr. Hansen will drop by to assure himself that your performance has been satisfactory. This evaluation as well as the quarterly evaluations will be based on the following areas:

25% - Productivity
25% - Market Position
25% - Profitability
25% - Capacity

Prior to each period, you will be given information regarding the point scale for each of the four performance measures (see Exhibit IV for an example). Your scores for each of the four measures will be taken from the computer printout and your points totaled. Payment will be based on your comparative point standing.

Environment and Performance Measure

SRSM Part A

Hansen Company has an opening for the manager of its Tulsa Unit. You have been interviewed by Mr. Hansen and have taken the job. You recognize that this is another step in your upward movement with Hansen Company. Your performance in this position will be a major factor in your career development. You will be evaluated on a quarterly basis along with the managers of Hansen Company's other units. It is understood that a good performance evaluation will lead to another quick promotion as well as a large increase in pay.

Mr. Hansen reminds you that the future of the Tulsa Unit is affected by your decisions. He trusts your ability and indicates that your next promotion may be possible within 3-5 quarters. Maintaining high quarterly evaluations is the key to obtaining the promotion. Quarterly evaluations will be based solely on the amount of profit accumulated. Quite obviously, the higher the quarterly profit, the more favorable the evaluation.

Environment and Performance Measure

SRSM Part B

You have done very well in your first four quarters with the Tulsa Unit. Mr. Hansen congratulates you on your work and announces a $5,000 raise. Regretfully, there is not a higher management position open at this point in time. However, in another 6-8 quarters a position will open up to which you may be elevated. For the interim. Mr. Hansen requests that you continue to operate the Tulsa Unit with the same success that you have experienced. Another evaluation will be made at the time the future position is available to assure the company of your ability. This evaluation as well as each interim quarterly evaluation will be based on the same set of variables as your previous evaluations.

Environment and Performance Measure

SRMM Part A

Hansen Company has an opening for the manager of its Tulsa Unit. You have been interviewed by Mr. Hansen and have taken the job. You recognize that this is another atep in your upward movement with Hansen Company. Your performance in this position will be a major factor in your career development. You will be evaluated on a quarterly basis along with the managers of Hansen Company's other units. It is understood that a good performance evaluation will lead to another quick promotion as well as a large increase in pay.

Mr. Hansen reminds you that the future of the Tulsa Unit is affected by your decisions. He trusts your ability and indicates that your next promotion may be possible within 3-5 quarters. Maintaining high quarterly evaluations is the key to obtaining the promotion. Quarterly evaluations will be based on the following areas:

25% - Productivity
25% - Market Position
25% - Profitability
25% - Capacity

Prior to each period, you will be given information regarding the point scale for each of the four performance measures (see Exhibit IV for an example). Your scores for each of the four measures will be taken from the computer pointout and your points totaled. Payment will be based on your comparative point standing.

Environment and Performance Measure

SRMM Part B

You have done very well in your first four quarters with the Tulsa Unit. Mr. Hansen congratulates you on your work and announces a $5,000 raise. Regretfully, there is not a higher management position open at this point in time. However, in another 6-8 quarters a position will open up to which you may be elevated. For the interim, Mr. Hansen requests that you continue to operate the Tulsa Unit with the same success that you have experienced. Another evaluation will be made at the time the future position is available to assure the company of your ability. This evaluation as well as each interim quarterly evaluation will be based on the same set of variables as your previous evaluations.

VITA 2

Kevin Dayna Stocks

Candidate for the Degree of

Doctor of Philosophy

Thesis: AN ANALYSIS OF PERFORMANCE EVALUATION USING SINGLE AND MULTIPLE ATTRIBUTES IN A SHORT-RUN AND A LONG-RUN SETTING

Major Field: Business Administration

Biographical:

Personal Data: Born in Logan, Utah, March 28, 1953, the son of Dr. and Mrs. Dayna L. Stocks.

Education: Graduated from Provo High School, Provo, Utah in May, 1971; received the Bachelor of Science degree magna cum laude with a major in Accounting and a Master of Accountancy degree from Brigham Young University in 1978; completed the requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1981.

Professional Experience: Graduate assistant, Institute of Professional Accountancy, Brigham Young University, 1976-1978; graduate teaching associate, Oklahoma State University, 1978-1980; Assistant Professor, School of Accounting, Oklahoma State University, 1980-present.

Professional Activities: Certified Public Accountant, Oklahoma, 1980; member American Institute of Certified Public Accountants; member American Accounting Association; member National Association of Accountants; AAA Doctoral Consortium participant, 1980.