RESEARCH FOR BUSINESS DECISIONS

AN INTERDISCIPLINARY APPROACH

Eli Peace Cox III
Editor

Assistant Professor of Marketing Administration
Graduate School of Business
The University of Texas at Austin

Bureau of Business Research
The University of Texas at Austin

POTENTIAL SOURCES OF ERROR IN BUSINESS RESEARCH PROJECTS

Eli P. Cox III
Assistant Professor of Marketing Administration
The University of Texas at Austin

A great deal of attention has been directed toward the decision-making process. This process has been characterized in numerous ways, but essentially it involves the following steps: problem or opportunity recognition, search for alternative courses of action, evaluation of these alternatives, choice of a course of action and implementation, and assessment of the results of the decision. The importance of relevant, accurate, current, and economical information at the various stages in this decision process cannot be over-emphasized. By the time that a problem or opportunity is universally apparent, it is often too late to do anything about it. Thus, in order to make changes successfully, a manager must have some sort of intelligence system whether he be in personnel, marketing, or finance. Once a problem or opportunity has been discovered, it is often the case that not all of the viable courses of action are known to the manager, so information must also be gathered at this stage. Certainly a manager's dependence upon information in the stage of evaluating alternatives is obvious. Such a process involves making forecasts of future events. As competitive pressures increase and as managers become more sophisticated, armchair decisions based upon a "feel" of the situation are being replaced by those which are based upon carefully fathered and analyzed information. Finally, as information is used to discover problems or opportunities, the same information can be used to assess the results of management's decision making.

Research methodology is the body of knowledge concerned with the techniques necessary for gathering quality information. Originally, research methodology was developed in the physical sciences. Recently it has been expanded to include the social sciences, including business administration. Research activities are traditionally characterized dichotomously: activities directed toward the generation of knowledge are described *as basic research* and *applied research* refers to those activities involved in gathering information which will be useful in solving a particular problem.

It must be emphasized that this dichotomy exists in the application of research methodology. It need not and should not exist in the research methodology itself. The same techniques that have brought rigor to basic research hi the physical and social sciences can be applied successfully in providing information for quality business decision making.

Although research methodology can be useful to businesses regardless of their size or the type of product or service they sell, not all businesses use research methodology with equal effectiveness or even use it at all. In fact, managers may be characterized as existing in one of four stages according to their attitude toward research methodology as a means of gathering quality information for decision-making purposes:

1. *The Stage of Ignorance.* Managers in this stage believe that research methodology is appropriately confined to the ethereal world of academia or at best to the technical research that may take place in the firm. These managers depend heavily upon intuition and experience, and the information they use is obtained informally.
2. *The Stage of Blind Faith.* Managers in this stage naively believe that the result of the application of research methodology is the good decision itself rather than the basis for making a good decision. They fail to see that good research can only reduce uncertainty; it cannot eliminate it. For a manager in this stage, the less he is capable of evaluating research because of its complexity, the more he is impressed with it. Show him some statistical analysis in a research report and he is very much convinced. Show him the same analysis on computer print-out and he is awe-struck.
3. *The Stage of Disillusionment.* This stage characterizes managers who were in the *Stage of Blind Faith* but now feel betrayed. These managers are very cynical because they have made costly mistakes even though the decisions were research-based. This vulnerability was due to the fact that they could not distinguish good from bad research, or that they felt that if the research was good, the decision had already been made,
4. *The Stage of Sophistication.* Managers in this final stage recognize the *potential* of research in improving the decision-making process. They recognize that a fortune teller is probably a better bargain than poor research. They also recognize that while good research does not eliminate the uncertainty involved in their jobs, it can be an economically warranted means of at least reducing that uncertainty.

If the effective application of applied business-research could eliminate uncertainty, rather than reduce it, there would hardly be a need for managers as skilled decision-makers. An experienced and sophisticated manager recognizes this fact. He sees the potential of good research in helping him to make good decisions, but at the same time he sees that some very real limitations exist regarding what such research has to offer.

First, the manager and the researcher face some very real time constraints. Social psychologists can afford to wait years and conduct dozens of research studies before they feel that anything conclusive can be said about the nature of the relationship between an individual's satisfaction with his participation in a group and associated variables. However, if the personnel director sees that turnover has reached an intolerable level among the clerical staff, then he is not in a position to wait very long. He can appraise the situation and try to make a remedial decision on the spot, or he can try to hedge his bet by systematically examining the situation before be takes action (i.e. by conducting research).

Second, the manager is faced with some very real financial constraints regarding the information he can gather. There is no internally specifiable budget limitation in basic research, because it is impossible to place a dollar value on knowledge that is gathered for its own sake. This is certainly not the case for applied research, where the information is obtained in order to deal with real problems, which at least in business situations can usually be evaluated economically. In such instances research is justified only so long as the costs of conducting it do not exceed the benefits derived from it.

Third, the manager is dealing with a decision-making environment of the utmost complexity. A chemist or physicist can reasonably conduct research in a controlled laboratory where he can systematically eliminate much of the complexity which confronts him. A sociologist or social psychologist is justified in creating situations which are amenable to study in order to draw tentative conclusions. A business manager can base his decisions upon information which has been gathered in a manner similar to that of me physical scientist or social scientist, but he is forced to confront the limitations of that research—he must live with the results of decisions which have been based upon imperfect information.

Because of the three factors just discussed, the business manager is faced with an important task before he can utilize the information which has been gathered through research. He must be able to appraise the quality of the information which has been gathered for his use. He should not be in a position where he is forced to accept blindly the results of applied business research at face value. He should recognize that every research project has its limitations. By inspecting these limitations a manager should be able to differentiate good from bad research. By examining the limitations which exist even within good research and adjusting for them, the manager can take quality information which has been gathered through research and make it even better as a tool for sound business decision making. Figure 1 (see page 17) presents an outline of the most common types of error. The three basic types: Errors of Definition, Errors of Estimation, and Errors of Explanation, will be discussed in turn.

POTENTIAL SOURCES OF ERROR

The remainder of this paper presents a discussion of the bases for the limitations of applied business research— the types of error which can reduce the quality of information gathered through research— and attempts to suggest some ways in which they can be dealt with after study has been conducted, as well as avoided if the research project is still in the design stages. Figure 1 presents an outline of the most common types of error. The three basic types, Errors of Definition, Errors of Estimation, and Errors of Explanation, will be discussed in turn.

Errors of Definition

Errors of definition exist when the researcher is researching the wrong topic. While the other types of error are the responsibility of the researcher, these errors are equally the responsibility of the researcher and the manager because they involve a problem of communication between the two parties, the two types of errors of definition are Misstatement of the Problem and Misstatement of the Relevant Variables.

Misstatement of the Problem

It is not unheard of that a well-conducted research project is completely useless because the wrong problem has been researched. In order for a manager to be able to utilize research, as well as be able to evaluate the research project, the problem at hand must have been clearly and explicitly defined. If this problem definition is effectively communicated to the researcher, then he is in a position to be able to direct all of his efforts toward its solution rather than searching in the dark.

It is possible that the manager has observed some of the symptoms of a problem---falling sales, declining profits, increased employee turnover—without knowing what the basic problem is. In such a case, it is quite legitimate that the research objective communicated by the manager to the researcher concern itself with problem definition. But even in this instance the manager must work closely with the researcher in order to guarantee the relevance of the research efforts.

Figure 1 **POTENTIAL SOURCES OF ERROR IN RESEARCH PROJECTS**

I. Errors of Definition
- A. Misstatement of Problem
- B. Misstatement of Relevant Variables

II. Errors of Estimation
- A. Measurement Error
  - 1. Error Due to Improperly Designed Scale
    - a. Ambiguity
      - i. Ambiguity in Question Transmission
      - ii. Ambiguity in Answer Transmission
      - iii. Ambiguity in Observation
    - b. Scale with Incorrect Mathematical Properties
    - c. Instrument Invalidity
      - i. Predictive Invalidity
      - ii. Content Invalidity
      - iii. Construct Invalidity
    - d. Instrument Unreliability
  - 2. Error Due to Improperly Used Scale: Inaccuracy
    - a. Inaccuracy Due to Inability
    - b. Inaccuracy Due to Unwillingness
      - i. Due to Time Costs
      - ii. Due to Perceived Loss of Prestige
      - iii. Due to the Desire for Privacy
      - iv. Due to Perceived Conflict with Researcher's Opinions
    - c. Inaccuracy Due TO Inability or Unwillingness of Researcher
- B. Frame Error
- C. Non-response Error
- D. Selection Error
- E. Sampling Error

III. Errors of Explanation
- A. Internal Invalidity
  - 1. History
  - 2. Maturation
  - 3. Testing
  - 4. Instrumentation
  - 5. Selection
  - 6. Morality
- B. External Invalidity
  - 1. Interaction of Testing and the Experimental Variable
  - 2. Interaction of Selection and Experimental Variable
  - 3. Reactive Arrangements
  - 4. Multiple Treatment Interference

Misstatement of Relevant Variables

It is also a danger that the relevant variables are not included for study. Thus, the manager and the researcher must agree upon the variables which may be causally related to the problem at hand. Such information may come from a preliminary investigation of the situation, past experience of the manager or researcher, or theory in the area of concern. It is equally important that the manager ensure that variables included for study are managerial relevant. A project which successfully examines a problem, but fails to imply a solution within the grasp of management, is not economically justifiable.

Errors of Estimation

While errors of definition are the joint concern of management and research, reduction of errors of estimation are primarily the responsibility of research. An error of estimation exists when there is a. discrepancy between the *true value* of a measure of some person or object and the *estimated valve of* that measure. Such a discrepancy can exist when a single variable is involved, such as when a researcher is trying to estimate a person's income, or a discrepancy can exist when a measurement instrument consisting of a number of variables is involved, such as when a researcher is trying to estimate a person's potential for employment on a particular job. The most frequently discussed types of error of estimation are: measurement error, frame error, non-response error, selection error, and sampling error. If a researcher were attempting to determine the average income in New York City by means of a survey, his efforts would be susceptible to all of these sources of error. (See Figure 2, Page 19)

Measurement Error

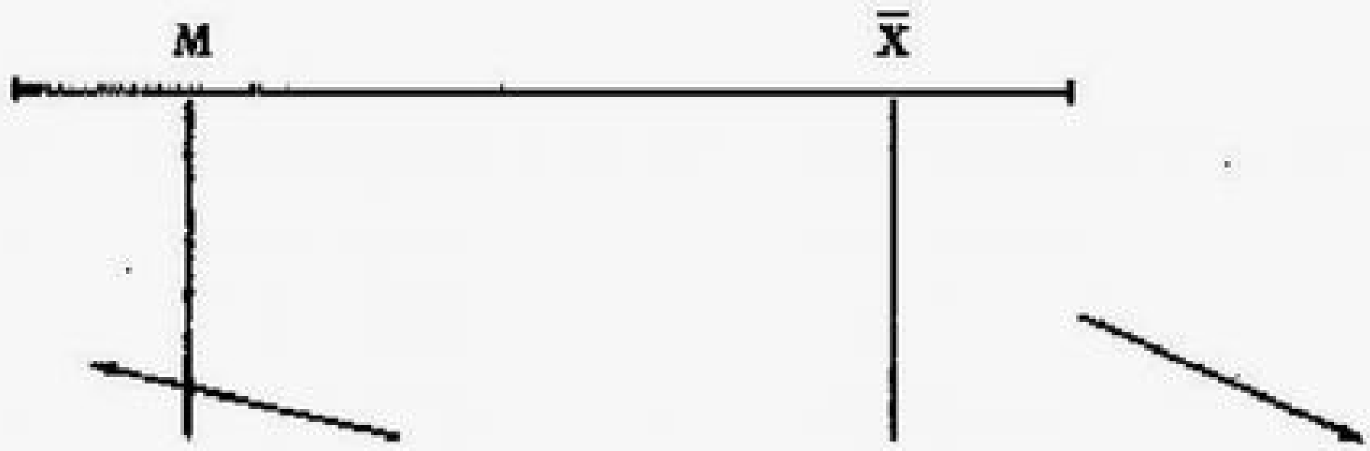
If a researcher were interested in obtaining information from a single person, be would not have to worry about frame error, non-response error, selection error, or sampling error. However, he would still have to contend with what is probably the most difficult error to deal with in the social sciences—measurement error.

It is necessary to define several other terms before defining measurement error. A *scale* is a. device used to record the degree in which some person or object possesses an attribute. A scale can be a physical device if it is measuring a physical attribute (i.e., a person's weight), or a scale can be a question on a questionnaire measuring some psychological attribute of a person (i.e. attitude). *Scaling* is the process of developing a scale which is capable of reflecting the properties of the attribute being measured. (i.e., there are many ways to measure a person's attitude toward his job; some are certainly better than others.) A *measure* is a score on a scale which should reflect the extent to which a person or object possesses the attribute. *Measurement* is the process of determining that score. Finally, *measurement error* exists when the score chosen does not successfully reflect the extent to which a person or object possesses the attribute. It occurs when the scale has either been improperly designed or improperly used.

Measurement Error Due to an Improperly Designed Scale

One of the most important problems faced in designing an accurate scale is that of ambiguity. *Ambiguity* involves miscommunication between the researcher and respondent in a personal interview or on a written questionnaire. This ambiguity can occur in the transmission of the question if the question is poorly worded and respondent fails to understand the question and provides an answer to what he incorrectly thought to be the question. This form of ambiguity is likely to occur if the researcher wishes to obtain a very specific piece of information from the respondent {i.e., students often find objective examinations very ambiguous). Ambiguity can also occur in the transmission of the answer, if the meaning of a vague response is misinterpreted. This is particularly a problem if

Errors of Estimation Involving the Determination of Average Income in New York City



M= Actual Average Income in New York City

$\bar{X}$= Estimated Average Income in New York City

e= $M - \bar{X} = e_1 + e_2 + e_3 + e_4 + e_5$= Total Error of Estimation Involving a Single Variable

e_1= Measurement Error

e_2= Frame Error

e_3= Nonresponse Error

e_4= Selection Error

e_5= Sampling Error

the researcher allows the respondent a great deal of latitude in answering a question (i.e., teachers often find the answers to essay questions highly ambiguous).

Unfortunately, ambiguity is not restricted to direct communication between the researcher and respondent. If the researcher is observing the behavior of an individual and using this behavior to infer unobserved aspects of the individual, then this may be done incorrectly due to ambiguity. For example, the brand of automobile a person drives is not always a good indication of a person's income or social class.

Significant steps can usually be taken in order to reduce the danger of ambiguity in the collection of data. In the case of personal interviews; only trained interviewers should be used, to ensure that the interviewer is capable of probing to obtain clear responses without shaping the nature of those responses. Written questionnaires should be pre-tested, that is, administered to a small but representative sample preliminarily. After the questionnaire has been completed the respondents can be asked to evaluate it and this information can be used to make revisions which reduce ambiguity. Even when observational studies are being conducted, ambiguity can be reduced. For example, the researcher may wish to observe several aspects of an individual in order to infer a single unobserved aspect.

A researcher can encounter measurement error due to an improperly designed scale if he selects a scale which does not nave the appropriate mathematical properties. If the properties of the attribute of the object or person that the researcher is attempting to measure do not conform with the mathematical properties of the scale he is using then measurement error occurs. In dealing with this potential problem there are four classes of scales from which the researcher can choose: ratio scales, interval scales, ordinal scales, and nominal scales. These scales are illustrated in Figure 3 in terms of the three properties that distinguish them. (See page 21.)

A *ratio scale* is so-called because the ratio of numbers taken from such a scale indicates that the same ratio describes the degree to which the two objects possess some attribute. For example, if one car is going 50 miles an hour and a second car is going 25, that is a ratio of 2 to 1 and thus the first car is going twice as fast as the second. The mathematical properties of such a scale can be described in terms of this example. First, the fact that one object is scored with a number which is larger than another indicates that it possesses more of something—speed. Second, the intervals along the scale imply equal differences in the attribute being scaled. That is, two cars have increased their speed by the same amount if one goes from 50 miles per hour to 51 and the second goes from 150 to 151. Third, there is a zero point on the scale which has a. real or natural meaning. Thus, a car going zero miles per hour is at rest. Other ratio scales measure length, weight, income, and age.

An *interval scale* orders objects according to the degree to which they possess some attribute, and the intervals along the scale are equal. An example of an interval scale is the Fahrenheit temperature scale. One hundred degrees is hotter than fifty degrees, and the increase in temperature between 100 and 101 is the same as the increase between 50 and 51. However, we cannot say that 100 is twice as hot as 50 just because it involved a ratio of 2 to 1 any more than we can say that 1 degree is a million times hotter than .000001 degree just because a ratio of 1,000,000 to 1 is involved. The reason for this is that interval scales do not possess a natural zero point. The Fahrenheit and Centigrade scales measure the same phenomenon, but their zero points indicate different degrees of coldness. The Kelvin temperature scale is a ratio scale and thus has an "absolute" zero.

An *ordinal scale* is a measurement device which orders objects according to the degree to which may possess some shared attribute. Examples of ordinal scales are the top ten records and rankings of graduating seniors and football teams. As the name implies, the ordering of the

C. H. Coombs, *A Theory of Data*, New York: John Wiley and Sons, 1964.

Figure 3 SCALES AND THEIR MATHEMATICAL PROPERTIES

Figure 3

SCALES AND THEIR MATHEMATICAL PROPERTIES

Scale	Natural Zero	Equal Intervals	Order Signifies Magnitude	Increased Preference by: Researcher Respondent
Ratio	+	+	+	↑ ↓
Interval	—	+	+	
Ordinal	—	—		
Nominal	—	—		

numbers on an ordinal scale does imply a similar ordering of the objects. However, the intervals along the scale are not necessarily equal. That is, the fact that a football team is ranked second does not suggest that it is superior to the team ranked third to the same degree that it is inferior to the team ranked first. Finally, as was the case with interval scales, a ratio of measures does not imply that two objects possess an attribute according to the same ratio, because there is not a natural *zero* in such a scale.

Finally, a *nominal scale* is actually more of a classification procedure than a scale, because the numbers designate membership in a particular class but do not indicate anything about the objects in those classes. Examples of nominal scales are: telephone numbers, football jersey numbers, and zip codes. The mathematical properties of a nominal scale can be described in the context of the example of telephone area codes. First, the fact that one part of the country has an area code of 812 and a second part of the country has an area code of 517 does not mean that the first area has any more or any less of something than the second area. In other words, the ordering of numbers in a nominal scale does not imply that the objects can be ordered similarly. Second, the fact that there are two pairs of area codes dial are equally similar numerically (i.e., 812 and 813, and 517 and 518) does not mean that the two pairs of areas are equally similar. In other words, the intervals along the scale do not indicate a specific unit of difference in terms of the property the nominal scale measures. Third, there is no natural zero point on a nominal scale; the range of telephone area code numbers is at the complete discretion of the person who chooses them.

If a researcher has his choice of scales to use he will pick a ratio scale, because it contains more information than the other three scales, as was suggested by the preceding discussion. Unfortunately, such a decision is not entirely at the discretion of the researcher, but is largely determined by the nature of the things being scaled. For example, it would be very difficult to use a ratio scale which reflects the absolute differences in quality of football teams; the problem is difficult enough with an ordinal scale.

As is suggested in Figure 3 (page 21), while the researcher prefers ratio and interval scales to ordinal and nominal scales because they contain more information, the respondent generally prefers to respond to nominal and ordinal scales. For example, most people can say with ease that they prefer a Ferrari to a Volkswagen (an ordinal statement), but would find it very difficult to make the precise statement that they prefer the Ferrari, say, 3.78 times as much as the Volkswagen (a ratio statement).

Thus the selection of the most appropriate scale is very important for two reasons. First, the researcher is always faced with the problem of choosing a scale which contains the maximum amount of information for his analysis purposes and at the same time does not make unrealistic demands upon the respondent. His problem is easily resolved if the researcher is asking the respondent about his objective characteristics: age, income, education, etc. However, it is a very difficult one to resolve if such subjective characteristics as attitudes, opinions, and preferences are involved.

The second reason why the determination of the appropriate scale is so very important is that the nature and number of variables determine the type of statistical analysis that can be done, Figure 4 (page 23) provides an illustration of the interrelationship. As can be seen, the types of statistical test available are determined by whether one or two sets of variables are involved, the number of variables in the sets, and the scaling properties of those variables.[2]

If a researcher has a lower-order scale, it is unacceptable to treat it as a higher-order scale, because he is assuming information which is not there. However, if a researcher has a scale of a

This classification was developed after examining several others: T.C. Kinnear and J. R. Taylor, "Multivariate Methods in Marketing Research: A Further Attempt at Classification," ***Journal of Marketing*** **34 (October 1971): 50-59; J. N. Sheth, "The Multivariate Revolution in Marketing Research,"** ***Journal of Marketing*****, 34 (January 1971): 13-19; S. Siegel,** ***Nonparametric Statistics*****, New York: McGraw-Hill Book Company, 1956; and B. W. Tuckman, Conducting Educational Research, New York: Harcourt, Brace and Javanovich, Inc., 1972.**

FIGURE & THE NATURE OF VARIABLES AND SELECTION OF STATISTICAL TECHNIQUES

Figure 4

THE NATURE OF VARIABLES AND SELECTION OF STATISTICAL TECHNIQUES

Type and Number of Variables in First Set (Dependent)

Type and Number of Variables in Second Set (Independent)

		INTERVAL		ORDINAL		NOMINAL		
		1	More Than 1	1	More Than 1	1	More Than 1	
		Mean, Standard Deviation, Average Deviation, t-Test	Factor Analysis, Cluster Analysis	Median, Percentile, Nonmetric Multi-Dimensional Scaling	Coefficient of Concordance, Cluster Analysis, Latent Structure Analysis	Mode, Binomial Test, Chi-Square	Cluster Analysis	
INTERVAL	1	Product-Moment Correlation, Simple Regression, Z-Test		Transform Ordinal Variable into Nominal, Use E-1		Discriminant Analysis		Row 1
	More Than 1	Multiple Correlation, Partial Correlation, Multiple Regression	Canonical Correlation	Transform Ordinal Variable into Nominal, Use E-2		Multiple Discriminant Analysis, Hotelling's T^2, Generalized Mahalanobis D^2		Row 2
ORDINAL	1	Transform Ordinal Variable into Nominal, Use A-5		Spearman Rank Correlation, Kendall Rank Correlation		Transform Ordinal Variable Into Nominal, Use E-5		Row 3
	More Than 1	Transform Ordinal Variable into Nominal, Use A-6		Kendall Partial Rank Correlation		Transform Ordinal Variable Into Nominal, Use E-6		Row 4
NOMINAL	1	t-Test, One-way Analysis of Variance, F-Test		Median Test, Mann-Whitney U, Kolmogorov-Smirnov, Runs Test, Moses Test		Cross-Classification, Chi-Square, Phi Coefficient, Fisher Exact		Row 5
	More Than 1	Analysis of Variance: Randomized Blocks, Latin Square, Factorial Design, MCA, AID	Multivariate Analysis of Variance	Friedman's Two-Way Analysis of Variance, Transform Ordinal Variable into Nominal, Use E-6		Multiple Cross-Classification Analysis, Chi-Square		Row 6
		COLUMN A	COLUMN B	COLUMN C	COLUMN D	COLUMN E	COLUMN F	

higher order it is quite acceptable for him to it as a scale of a lower order. A researcher is often willing to make this sacrifice of information in order to increase the availability of statistical tests.

Finally, a researcher confronts measurement error due to an improperly designed scale when a number of questions or scales are combined in order to make a more complex measurement instrument or test. Examples of these complex instruments are I.Q. tests, job placement tests, and academic placement tests such as the A.T.G.S.B. (the Admission Test to Graduate Schools of Business). Such tests are designed to measure complex aspects of human potential. The two types of error which are usually discussed in the context of these types of tests are called instrument invalidity and instrument unreliability. However, to be consistent with their traditional use this discussion will direct itself towards the positive qualities of instrument validity and reliability.

Validity refers to the measurement power of an instrument, that is, whether it is successful m measuring what it is supposed to measure. One questions the validity of an I.Q. test when asking whether a set of questions, which are culture-bound and education-bound are truly capable of measuring a person's intellectual ability, or whether a personality test which we require of our potential employees is actually capable of indicating who should be employed. Such a measurement instrument can be evaluated in terms of its predictive validity, content validity, and construct validity.[3]

Predictive validity refers to the ability of a measurement instrument to forecast a person's behavior. If those students who have high scores on the A.T.G.S.B. succeed in graduate business school and those who receive low scores do not, then the instrument has predictive validity. Thus, if there is a high correlation between the criterion variable (success) and the predictor variable (A.T.G.S.B.) then predictive validity has been demonstrated. The assumption is that a human characteristic is capable of manifesting itself in several ways, so the ability to achieve which is demonstrated by success in an M.B.A. (Masters of Business Administration) program can also be demonstrated by a high score on a properly designed admission test.

Since the objective of a manager is to predict human behavior (i.e., that of customers or employees), predictive validity is certainly a very important consideration for obvious reasons. However, predictive validity should not be the sole basis for evaluating the quality of a measurement instrument. Content and construct validity are equally important, particularly if the behavior being predicted is complex.

Often it is very difficult to define a single criterion against which a measurement instrument can be evaluated. It is generally agreed that intelligence is not a single quality, but consists of several aspects such as mathematical ability, verbal ability, memory, and creative thinking. Success as a manager requires intellectual ability, interpersonal skills, experience, motivation, and so on. A measurement instrument which reflects the complex nature of the behavior or set of behaviors under consideration is said to have *content validity*. While predictive validity lends itself to empirical evaluation, content involves the judgment of an expert who is capable of determining whether the full complexity of the behavior being considered is reflected in the measurement instrument.

When a researcher is dealing with a very complex human characteristic such as entrepreneurial ability, it may be virtually impossible to define one or even a set of criteria, against which one can empirically test the validity of a measurement. An alternative is to define other characteristics of behaviors which are theoretically consistent with the complex characteristic, but which are measurable themselves. For example, we can say that risk propensity is a measurable attribute which is consistent with the theoretical construct of entrepreneurial ability. If a researcher is successful in the development of a measurement instrument which is a logical derivative of this complex human characteristic

3 E. Thiselli, *The Theory of Psychological Measurement*, New York: McGraw-Hill Book Company, 1964, Chapter 11.

then the test is said to have *construct validity*,

As was suggested earlier, a second type of measurement which is usually discussed in the context of complex measurement instruments is *instrument reliability.* Reliability refers to the ability of an instrument to provide a consistent measure of the human characteristic over time and among people. In one respect an I.Q. test is reliable if a person takes it and then takes it again and gets the same score. However, it is quite possible that a person will do better on the second test because of the experience gained from taking it the first time, rather than because of increased intelligence. In another respect, an I.Q. test is reliable if two people with the same level of intelligence receive the same score. However, they may differ in scores due to differing educational backgrounds.

What this discussion suggests is that a manager should use such complex measurement instruments with care, particularly if they have been declared valid and reliable by fiat. In fact, the development of such instruments is an important task of research in itself. The study of the nature of intelligence and the development of accurate I.Q. tests have been in process for decades.

Certainly a personnel manager cannot wait decades for the research staff to develop a sound placement test. However, he should have some idea of the accuracy of such an instrument before he puts it to use. This estimate of accuracy should be established empirically as well as by means of the evaluations of experts. In addition, the process of evaluating such instruments should continue after they are in use. In the case of the personnel director, this can be done by comparing employment records with test performance historically.

Error Due to an Improperly Used Scale

As was suggested in earlier comments, instrument error can occur either because of an improperly designed scale or because of a scale that is improperly used. *Inaccuracy* is the term commonly associated with the latter type, of measurement error, and refers to the instance where information has been received incorrectly from the respondent. If the respondent provides incorrect information about his current state of affairs, then this is termed *concurrent inaccuracy.* If the incorrect information concerns the future, then the term *predictive inaccuracy* is used.

Inaccuracy can occur either because of the *unwillingness or inability of the respondent* to provide the correct information. Thus we can have concurrent or predictive inaccuracy due to either inability or unwillingness on the part of the respondent. A man reveals that he has a certain amount of insurance, but has forgotten about a policy through his company-concurrent inaccuracy due to inability. A housewife indicates over a telephone interview that her favorite magazine is *Harper's*, while it really is *True Story*—concurrent inaccuracy due to unwillingness. A salesman incorrectly states what his sales for the next year will be simply because he has no real basis for knowing—predictive inaccuracy due to inability. In completing an application for an advanced training program, an executive falsely states that he is not going to be leaving the company in the near future—predictive inaccuracy due to unwillingness.

Although inaccuracy due to inability is straightforward, inaccuracy due to unwillingness can occur for a number of reasons. First, the length of the interview or questionnaire may be such that the respondent provides superficial answers to the questions. Second, the respondent may respond in such a way as to create a feeling of prestige or status rather than reflecting his true feelings. Third, the respondent may provide an incorrect answer based upon the feeling that the subject involved is a private matter and not the business of the researcher. Finally, the respondent may simply provide

4 P. E. Green and D. S. Tull, *Research for Marketing Decisions*, New York: McGraw-Hill Book Company, 1954, pp. 121-126.

answers which he feds will conform to the expectations of the researcher.

Inaccuracy can occur due to either the *inability or unwillingness of the researcher* himself. Inaccurate information collected in an in-depth interview is particularly susceptible to inaccuracy due to the inability of the researcher. The answer of the respondent may be quite clear to the interviewer, but the interviewer may have great difficulty in coding the answer so that it can be effectively compared with the other answers. Another danger is that the researcher has a stake in the outcome of the research and thus is not objective in the data he records for study.

The researcher and manager should always realize that some information is beyond their reach simply because it is beyond the reach of the respondent— inaccuracy due to inability. Thus it should always be realized that the more demanding the questions are, the more likely it is that the answers will be inaccurate.

However, it is possible for the researcher to reduce the problems of inaccuracy due to unwillingness. When pre-testing the questionnaire, the researcher should make sure that it is not so long that the respondent tends to answer superficially. In addition, questions of a sensitive nature should be phrased in such a manner that they do not threaten the respondent, and, at least in the case of marketing research, their anonymity can be guaranteed. Inaccuracy can also be reduced in personal interviews by using only a highly trained field research staff.

Frame Error

A frame is a master-list used to enumerate all of the elements in the universe, or population under study. A frame may be inaccurate either because it excludes elements that are a part of the universe or because it includes elements that are not a part of the universe. The U.S. law which requires all eighteen-year-old males to register with the Selective Service is an effort to eliminate this inaccuracy. However, this inaccuracy is *not* frame error.

Frame error occurs if the inaccuracy in the frame causes a researcher to make an estimate of a parameter of the universe which is incorrect. If a researcher wished to conduct a census (interview all of the people in a population), then an inaccurate frame would prevent him from doing so. He might contact everyone in the frame, but reach only 90% of the population. If this population is fairly homogeneous, then this would not be a major problem and would he a random form of error similar to sampling error (which will be discussed shortly).

However, the real danger of frame error comes when: 1) the frame is inaccurate and, 2) there are systematic differences between those accurately included in the frame, on the one hand, and those who sire inaccurately excluded or included, on the other hand This leads to a systematic form of error which is very dangerous. A study conducted by the author several years ago can provide a good example of this danger. The study involved 950 telephone interviews of voters in a medium-sized Southwestern city concerning their evaluation of the city council. Analysis for the total sample revealed that the people were impressed with the job that was being done and that upper-income home-owners were particularly impressed.

The fact that upper-income homeowners were particularly impressed with the job the city council was doing suggests that this study might have been especially vulnerable to frame error. The frame that was used was telephone directory. This frame is vulnerable because of those who were inaccurately included in the frame as well as those who were inaccurately excluded from the frame. On the one hand, the people who are listed in the phone book but who are not within the city limits might be inclined to respond favorably to the city council because they are in a position to enjoy the city's services without having to pay the taxes. On the other hand, many people who are not homeowners or who are not in the upper-income segments are often not well represented in the telephone book: the young, the old, students, Chicanos, and blacks. Both of these forms of inaccuracy would probably lead to frame error by exaggerating the level of voter satisfaction.

Thus an important task of the researcher is the selection or development of the most accurate frame possible. Once the survey has been conducted, the manager should ask their questions. First, how substantial is the inaccuracy of the frame? Second, are there any systematic differences between those who are inaccurately excluded or included and those who are accurately included? Third, given the answers to the first two questions, in what manner and to what degree should we modify the results we have attained?

Non-response Error

Even if the researcher were able to design the perfect measurement instrument and conduct a census using a frame which was completely accurate, he would still have to contend with error of several kinds. *One* of these forms of error is non-response error. The fact that every potential respondent is contacted initially in a survey does not guarantee that everyone will respond. In a mail survey, a researcher is most fortunate if half of those contacted complete the questionnaire. If those who did respond were representative of those who did not, then the researcher would be faced with another random form of error which is similar to sampling error. However, those who do respond are often quite unlike those who do not respond. If a survey investigates some controversial issue, then those who are strongly in favor of the issue or those who are strongly against it are more inclined to respond in general than are those with little feeling about the issue. A higher percentage of the middle class responds to surveys than of the upper and lower classes. Thus non-response error exists when: 1) less than 100 percent of those contacted respond to a survey, and 2) there is a systematic difference between those who have and those who have not responded with regard to their answers to questions on the survey.

The researcher can do a number of things to increase the response rate. In any kind of survey research this can be done by convincing the respondent of the legitimacy of the research, as well as guaranteeing anonymity. Extensive research has been conducted which has been concerned with increasing response rates in main surveys. Some of the devices that have been used to increase the response rate are: personalizing the cover letter, using first-class rather than third-class mail, sending follow-up reminders, and offering financial incentives.

Selection Error

Selection error occurs "if certain elements in the frame have a greater chance of falling into the sample than others, and if this difference is not corrected by [a] subsequently weighting".[6] That is, due to the sampling procedure used by the researcher, some elements of the population may have a greater chance of being included in the sample than others, thus causing a biased result.

There are two basic classifications of sampling procedures: probability sampling and non-probability sampling.[7] In *probability sampling* every element in the universe has a known chance of being selected for inclusion in the sample. The simple random sample is an instance of the probability sample for which the probability of choosing a sampling unit is the same for all units, and is known. In *non-probability sampling* the probability that an element is *included* in the sample is neither equal for all units nor known. The techniques which are traditionally taught for measuring sampling error require a simple random sample. These techniques can be applied in modified form to other types of probability

5 E. P. Cox, IU, W. T. Anderson, Jr. and D. G. Fulcher. "Reappraising Mail Survey Response Rates." ***Journal of Marketing Research*** **11 (November 1974).**

6 C. S. Mayer, "Application of Bayesian Statistics to Research Design," in Albaum, G., and Venkatesan (eds.), ***Scientific Marketing Research*****, New York: The Free Press, p. 86.**

7 H. W. Boyd, jr., and R. Westfall, ***Marketing Research*****, Homewood, Illinois: R. D. Irwin, Inc., 1972, p. 352.**

sampling methods where the probability of selecting a sampling unit is known, but not the same for all sampling units. However, sampling error can not be measured if a non-probability sampling procedure is involved. This is not to say that the sampling error is greater when a non-probability sampling procedure is involved. Rather, there is no basis for measuring sampling error because of the confounding nature of the selection error which is present. Thus when sampling error poses a threat to the accuracy of a study, a probability sampling procedure should be used in order to allow the researcher to estimate this error.

Sampling Error

As the name implies, sampling error is the error incurred because the researcher has chosen to conduct a survey (contacting less than one hundred percent of the population) rather than conducting a census (contacting one hundred percent of the population). The researcher can calculate some statistic such as the mean to describe some population or two statistics to compare two populations, or fit some descriptive statistical model such as regression in order to describe the relationship among variables which describe a single population. This can be done whether a sample or the whole population is used. If such descriptive statistics are based upon a sample and if certain assumptions are met, then estimate of sampling error can be made. This estimate of sampling error consists of a probabilistic statement regarding the difference between the calculated statistic and the true population parameter.

The assumptions of parametric statistical tests are: 1) the units included in the sample or samples have been selected on a random basis; 2) the variable(s) have been selected from normally distributed populations of infinite size; 3) if, more than one population is involved, the populations have the same variance; and 4) the variables are measured in at least interval scale.[5]

If these assumptions have been satisfactorily met, then the question can be asked: is the information which has been uncovered by means of the descriptive statistics meaningful, or is it simply a product of the random variation associated with sampling? Due to sampling theory, this question can be answered precisely. This is excellent from the point of view of the researcher as well as the user of that research. However, this preciseness can be deceptive and misleading.

One reason the measurement of sampling error is deceptive and misleading is that its very preciseness can be used as methodological slight-of-hand. Five types of errors of estimation have been discussed; measurement error, frame error, non-response error, selection error, and sampling error. The only error of estimation that can be measured with any degree of preciseness is sampling error. Therefore, much of the discussion in a research project focuses upon this measurement. The other four types of error cannot be measured with any degree of preciseness and are often submerged in the data which is being analyzed. The implication is that they do not exist. However, they do exist and in combination are certainly more of a potential threat to the meaningfulness of the research than is sampling theory. In other words, a researcher may be comparing the opinions of two groups by means of a t-test and find a difference which is significant at the .000,001 level (i.e. there is one chance in a million that the result is due to sampling), and still have worthless results because he has not been able to obtain meaningful expressions of opinion (measurement error) or because he has not reached a representative cross-section of the populations with which he was concerned (frame error, non-response error, or selection error).

There is a second reason why the measurement of sampling error can be deceptive *and* misleading. When a relationship has been found to be highly significant statistically, there is a tendency to feel greater certainty in discussing why such a relationship exists. For example, even if the attitudes of the member of one group have been found to be different from the attitude of a second group and there is little concern with any of the five errors of estimation, we are still not necessarily in a position to state with certainty why such a difference exists. In other words, even if a researcher has dealt successfully with the errors of estimation, he still has to avoid the errors of explanation if he is interested in explaining why certain things have occurred. The following section will concern itself with these errors of explanation.

Errors of Explanation

Errors of explanation exist when a researcher makes an inappropriate inference regarding cause-effect relationship, A study has been conducted by a researcher at a major university which suggests that students performed a manual task better when they did not have direct supervision than they did when there was direct supervision. The production manager in an electronics firm who is considering a.change in the type of supervision of his employees sees this study and appraises the study results by asking two questions: (1) Was the difference in the performance of the two groups of students really due to the method of supervision or is there some other explanation for this difference? (2) Even if the differences between the two groups *can* be attributable to the method of supervision, can the same results be successfully applied to my situation?

When a person is asking questions similar to the first one the manager asked, he is raising questions as to the internal validity of a study, a study has *internal validity* if it has been successful in isolating the true causal relationship. When a person is asking questions similar to the second one asked by the manager, he is concerned with an external validity. A study has *external validity* if the relationship which has been discovered to exist between variables also holds in other, similar circumstances.

There are three situations in which we can determine a causal relationship with varying degrees of certainty.[9] The first situation is one in which we infer causality from a "concomitant variation" between variables. For example, from a survey we find that people who are high in economic status are also high in educational status, and a regression analysis reveals that eighty percent of the variation in educational status is "explained" by corresponding variation in economic status. This strong association between the two variables suggests that there may be a causal relationship, but it does not suggest the nature of that relationship. It may in fact be that education determines income, or that income determines education, or that causality is bilateral, or that there is no causality between them but that high intelligence is the cause of both high education and high income. All of these possibilities are consistent with the association which was found between the two variables. [ll] It should be clearly understood, then, that association does not imply causation. However, we can say that causation does imply association.[10]

The second situation suggesting the possibility of an underlying causal relationship is a sequence of events which may occur in a changing relationship between variables. For example, an analysis of longitudinal data may suggest that persons with high-income backgrounds tend to reach higher levels of educational achievement. This certainly implies more about the relationship between the variables than does "concomitant variation." However, it does not ascertain a causal relationship because the true underlying variable may be the level of intelligence in the family.

Thus, with "concomitant variation[1]' and with "sequence of events" it is possible to isolate associations between variables, but it is very dangerous to infer causality because there may be a number of alternative explanations which are equally consistent with the data. There is only one way in which causality can be determined with certainty, and that is by systematically eliminating all possible explanations but one. A researcher can attempt to deal with rival explanations by incorporating additional variables into the analysis of data collected from an observational study, by way of regression analysis, for example. However, it is never possible to complete this process.

The only way in which all rival explanations of the data can be systematically eliminated is through experimentation. If we can create conditions where the only difference between the two groups is the nature of their contact with a predictor variable, then we can safely assume that any resultant differences in the criterion variable for those groups are attributable to the difference that existed in the predictor variable. That is, if we are successful in creating proper experimental conditions from which we can infer causality, then we can be assured of the internal validity of our research.

9 Green, op. cit., pp.79-82
10 Green op. cit., p. 329.

After reading the preceding discussion one might be inclined to ask the following question: if experimentation is the only way to guarantee the internal validity of research, then why is not all research of an experimental nature? One explanation for this is inherent in the subject matter. Research problems which are essentially of a micro nature, such as those involving psychology and social psychology lend themselves to research of an experimental nature. However, research problems of a macro nature such as are dealt with in economies, sociology, and political science are often too large and complex to be dealt with in an experimental manner. This explains why the empirical research in some disciplines of the social sciences has been more rigorous than all others.

This same dichotomy exists in applied business research- Problems dealing with the individual as consumer, employee, etc., lend themselves more readily to solution by the expert mental method. Problems dealing with large groups of individuals, such as organizations of markets, pose great problems for those who wish to apply the experimental method to their solution.

There is a second reason why all research is not of an experimental nature. Not all researchers are willing to make the sacrifices required by exp-
erimental research. As was suggested earlier, the bias of experimentation is the ability to control the environment in which the research is conducted to such a degree that all possible explanations for a situation but one have been systematically eliminated. The sacrifice then is that the realism of the situation must be restricted in order to obtain this degree of control. To the extent that the environment in which the research takes place is an artificial one, we may well have difficulty in transferring our result to natural circumstances, even if there is some assurance that the experiment has produced an internally valid result. This lack of generalizability is due to external invalidity.

Thus, ironically, as a researcher takes steps to increase the internal validity of his work, he is generally reducing its external validity.[11] Conversely, by attempting to increase the external validity of his research, he is likely to be compromising its internal validity. It is certainly possible to have a study which is neither internally nor externally valid. However, this trade-off generally exists, as is illustrated in Figure 5 (page 31). Macro problems lend themselves to observational and quasi-experimental research strategies and tend to be subject to internal invalidity. On the other hand, micro problems lend themselves to an experimental approach and are susceptible to external invalidity. It should be pointed out that these generalizations are by no means absolute, for numerous studies of individual behavior have been of an observational nature, and problems of a macro nature have been simplified in order to conduct laboratory experiments.

Internal Invalidity

As has been suggested, internal invalidity is the form of error of explanation which involves a misinterpretation of a cause-effect relationship. Although observational studies which do

11 The following discussion of internal and external validity relies heavily on D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, New York: Rand McNally and Company, 1963.

THE NATURE OF BUSINESS PROBLEMS AND RESEARCH STRATEGIES

Nature of Business Problem	Macro Problems: Organizations, Markets.	Micro Problems: Customers, Employees.

Nature of Research Strategy	Observational	Quasi-Experimental Field Experiments	Experimental Laboratory Experiments

----------------------------- Increasing Experimental Control -----------------------------------

Nature of Experimental Error	Internal Invalidity 1. History 2. Maturation 3. Testing 4. Instrument Decay 5. Selection 6. Morality	External Invalidity 1. Testing/ x 2. Selection/ x 3. Reactive Arrangements 4. Interaction among x’s

not attempt to control the environment of the research are particularly vulnerable to internal invalidity, the fact that a researcher has attempted to conduct an experimental study is by no means a guarantee that internal invalidity does not exist. Thus it is always necessary to examine the results of a study in order to determine whether they might be explained by one of the several causes of internal invalidity.

History. The dimension through which any change is observed is time, and this is certainly the case when attempting to observe causal relationships. Evidence of the effectiveness of a medicine is demonstrated by giving it to patients and then observing that they get well. The benefits of additional advertising are demonstrated by observing an increase in sales. Managers are encouraged to have their shop foremen participate in sensitivity training because the productivity increased in one plant after such training,

It is quite possible that the sequences of events described in these examples demonstrate the suggested causal relationships. However, there may be alternative explanations which are consistent with the data. It may well be that the patients were confined to bed and that rest alone would have produced the improvement in health. The company's increase in sales might have reflected an increase in industry sales, and in actuality its market share might have decreased, suggesting that advertising might actually have had a negative effect. The increase in productivity in the manufacturing plant might have been due to changes which had taken place in the production process rather than the fact that the foremen had received training.

As is the case with these three examples: any time a *change takes place in the environment external* to the "experiment" which has an effect which is confounded with the effect of the predictor variable, then this is a case involving internal invalidity due to the *history effect.*

Maturation. A firm which specializes in sales training guarantees that salesmen who participate in their thousand-dollar program will be able to increase their sales substantially. A manufacturer has been having trouble with three of his young salesmen and sends them to participate in the training program. Within six months after the program, the salesmen have shown substantial improvement. Management considers the possibility of sending all of its salesmen to participate until a senior salesman points that all salesmen get better with time, particularly young ones, and that the same improvement in sales would have been experienced even if they had not participated in the program. Any time there is a *change which takes place naturally in the people* which is mistakenly considered to be the result of the "experiment, this is labeled as *maturation.*

Testing. Managers are interested in evaluating a training program. They compare the scores of a test group on a standardized industrial placement exam which was taken before the training program with the scores received on the examination which was given again after the training program. The results appear to be quite encouraging. However, it is pointed out to the managers that the scores might have been improved when the test was given the second time simply because the individuals has been forewarned by the first test that they were to be evaluated.

The testing effect occurs when the process of measurement has an influence on what is being measured and the change is mistakenly attributed to a change in the experimental variable. In the previous example, the participants performed better on the second exam because they were sensitized by the first test. When the tests involve the measurement of attitudes, it is possible for the second test to reflect changes in attitude that were brought about simply by taking the test the first time, which called attention to these attitudes. Or the first test may have provided a practice or learning experience which was reflected in improved performance on the second test.

Instrumentation. As a means of evaluating a sensitivity training program the interpersonal skills of a group of line managers are observed and recorded both before and after participating in the program. While the unproved scores may have reflected the benefits of the program, they also may have reflected changes that may have taken place in the observers. For example, the observers might have become more skillful in recognizing the demonstration of interpersonal skills, and what they failed to observe at the beginning of the "experiment" was recorded toward the end. Thus, the *instrumentation effect or instrumentation decay* involves changes of inconsistencies in the measurement process over time which may be confounded with the result of changes in the experimental variable. This instrument decay can also occur with physical measurement instruments. For example, if we are weighing a series of heavy objects with a spring scale, the spring tends to stretch over time and objects are increasingly weighed heavier than they actually are.

Selection. A manufacturer is interested in determining the effectiveness of a sales training program and wishes to do so by comparing the average sales figure for those who have participated in the program with the average sales figure for those who did not participate. On three consecutive Saturday mornings a group of fifteen volunteers participate in the training program. Sales data are collected for the six months after the program and used as a basis for comparison of the two groups. The average sales figure was substantially higher for those who had been in the training program. Was the higher sales figure attributable to the effectiveness of the training program? Or was it due to the fact that the sort of people who are willing to devote Saturday mornings to education are going to work harder at selling than the other group and do better even without the training program? If this latter explanation is at least partially true, then we have an example of the selection effect. The *selection effect* exists when the process of selecting individuals to participate in two groups is such that the groups are not equivalent and thus differences cannot later be interpreted as an indication of the influence of some experimental variable.

Mortality. The research chemists of a toothpaste manufacturer have come up with a brilliant new discovery of a toothpaste which is more effective in cleaning teeth than anything else on the market. Unfortunately it tastes like [expletive deleted]. The company's marketing researchers have been commissioned to determine whether the toothpaste's strong quality offsets its weak quality sufficiently to cause consumers to buy it. To make this determination the researchers decide to conduct a scientific experiment. They select two thousand people at random and randomly assign them to two groups. They give a six months supply of the experimental toothpaste to one group and a corresponding amount of the most popular toothpaste on the market to the second group. After the six months was over the researchers had everyone rate the toothpaste they used on an attitude scale. The analysis of this data revealed that the average rating of the experimental toothpaste was substantially higher than the other brand.

Because of the results of this analysis management was about to conclude that the new toothpaste should be introduced, until the highly-trained M.B.A. from the University of Texas spoke up. He suggested that they look at the numbers of people who completed the experiment in the two groups. They found that 957 of the people that had been given the popular toothpaste had completed the experiment, while only 173 people who had been given the experimental toothpaste completed the experiment. This suggested that those few individuals who had continued using the experimental toothpaste were such Spartans that they could put up with anything in order to have cleaner teeth and thus rated the toothpaste very highly. Most of the people in the experimental group considered the taste of a toothpaste to be an important consideration and chose to drop out rather than suffer for six months.

Mortality occurs, reducing the internal validity of a study, when there is some factor which causes the attrition rate in one group to be greater than for another, and thus causes differences between the two groups which can be incorrectly attributed to the experimental variable.

The six types of internal invalidity fall into two categories. History, maturation, testing, and instrumentation all occur when a researcher compares the score on tests taken before a trial or "experiment" with scores taken afterward. History and maturation are due to confounding influences which are either internal {maturation) or external (history) to the experimental environment. Testing and instrumentation are both instances of instrument unreliability as it was discussed in an earlier section. Instrument unreliability can be an important cause of internal invalidity if it leads a researcher to misinterpret the cause of changes which have taken place in a group.

While research is susceptible to the first four types of internal invalidity when comparisons are made of a single group over time, selection and mortality can plague research which attempts to compare two groups at a point in time.

It is certainly possible to develop a research design which is quite successful in dealing with these and other forms of internal validity. It is beyond the scope of this paper to discuss the requirements of an internally valid experimental design, which is a lengthy and technical discussion in itself. However, the user of marketing research should be in a position to appraise the internal validity of a study by asking if there might not be several explanations which are consistent with the data. An understanding of the six types of internal invalidity should help to provide direction to this process of considering these alternative explanations.

External Invalidity

An electronics firm located in Los Angeles is interested in determining whether they should pay their assembly-line workers (nonunion) on a piece-rate basis or by the hour, they hire a research firm to help them decide. The researchers say that the only way to make a good decision is to base it on experimental evidence. They indicate that the company could pay half of the employees on a piece-rate basis and the other half by the hour and observe the result. However, this would be costly, methodologically unsound, and would mean automatically making a mistake with half of the employees.

The researcher suggests that a controlled laboratory experiment would be far superior to such a field experiment. With management's agreement the researchers select one hundred employees to participate in a month-long experiment conducted in a behavioral laboratory located in a suburb of Los Angles. After the experiment had been completed, the researchers recommended that the piece-rate method of compensation be adopted.

The research staff of the electronics firm evaluated the experiment. Their evaluation indicated that the experiment was well designed {i.e., it was high in terms of internal validity) and thus endorsed the recommendation of the research firm. Confidently management placed all of their employees on a piece rate. A week later, the personnel director was found wired for sound and management had a riot on its hands. What happened? Someone failed to mention the possibility that the experiment was externally invalid. A study is *externally invalid* when there is a lack of generalizability to the real world of what may be valid casual relationships found in the experimental environment. Poor bases for external invalidity will be discussed.

Interaction of Testing and the Experimental Variable. Earlier it was suggested that the testing effect exists when a questionnaire given before an experiment sensitizes a group of individuals so that they respond differently when the questionnaire is used again after the experiment. Thus, if we are evaluating a training program on the basis of improvements in performance on a placement examination, we cannot determine how much of that improvement was due to the training program and how much was due to retaking the examination. For this reason we face problems with internal validity. One way to deal with this problem would be to give the examination to two hundred people, randomly select one hundred of them to participate in the training program while using the remaining people as a comparison or control group, and then give the examination to all two hundred people afterward. If there was a ten- percent improvement in the scores for the people who had not been in the training program and a thirty-percent improvement in the scores of those who had, we might be able to make the following statement regarding the second group: their scores improved approximately ten percent due to the fact that they had taken the same examination twice, and the remaining twenty-percent improvement in their scores can be attributed to the effectiveness of me training program.

Can we then conclude that if a firm has all of its employees participate in the training program, their performance will on the average increase approximately twenty percent? Not necessarily. We may be able to say that people who take a placement exam and then participate in a training program improve in performance by approximately twenty percent, but that is not to say that all people participating in a training program will on the average experience the improvement of twenty percent. That is, it may be that taking the examination sensitizes people and causes them to be more susceptible to the influence of the training program, if this is the case, then we have an example of external invalidity due to the interaction of testing and me experimental variable.

Thus testing can cause both internal invalidity and external invalidity. If the result of testing is *confused with* the result brought about by the experimental variable, then we have an example of internal invalidity. On the other hand, if the process of testing *has an effect on* the result brought about by the experimental variable, then we have a case of external invalidity,

Interaction of Selection and the Experimental Variable. Selection can be involved in either internal invalidity or external invalidity, as was the case with testing. If the process of selecting individuals for membership in the experimental group and in the comparison or control group is different, then the differences between the two groups which exist after an experiment can be attributed to either the process of selection or to the experimental variable. When the influence of selection *is confused with* the influence of the experimental variable, then internal invalidity is involved.

When there is an interaction between the influence of the process of selection and the influence of the experimental variable, then external invalidity exists. For example, a. distinguished marketing professor conducts an experiment, using his students, which demonstrates that the most important factor in a new-car-purchase decision is economy. Even if the experiment is completely valid internally, Detroit should be very careful in generalizing the result. This study suggests that the most important decision factors for marketing students are economy. It does not suggest that the most important decision factor is economy for all automobile purchasers or even for all students. When the general liability of an experimental result is limited due to the fact that a unique set of individuals participated in the experiment, then this is an instance of external invalidity due to the interaction of selection and the experimental variable.

If a researcher is interested in avoiding the problems associated with the interaction of selection and the experimental variable, then those who participate in the experiment should be truly representative of those in the population to which the results of the study are to be generalized. This is, of course, virtually impossible to do this completely in practice, particularly when controlled experiments are involved. Alternatively, a researcher can reduce this problem by choosing to conduct an observational study, but then he has to contend with the problems of increased internal invalidity.

Reactive Arrangements. A market researcher is interested in determining the impact on sales of a new package design for a food product. One means of examining this relationship is to introduce the product into a few stores and then measure the resultant sales. However, the researcher recognizes the fact that there are all sorts of other variables over which he has no control but which might affect the product’s sales. The position of the product on the shelves may be a very important confounding factor if shoppers have a tendency to buy the first brand *they* see. If the product is placed at one end of the section of the shelf, sales may be very high, and if it is placed at the other end sales may be very poor. The researcher is also unable to control the marketing activities involving the other products while sales are being observed. If a competing product is offered as a special, then the sales during that period may be a poor indication of the effectiveness of the package design. In addition, there is the problem of inventories. What happens if his product sells like gang busters for a while and then stocks out, so that no additional products can be purchased, or if something similar happens to a competing product?

A natural reaction of a researcher would be to attempt to eliminate some of this chaos. A laboratory experiment is a logical device by which the researcher could eliminate some of these unwanted variables. He might place the relevant products on some shelves in a laboratory and maintain constant prices for all products, constantly maintain adequate inventory levels, and systematically vary the order of tile products on the shelves.

Unfortunately, this control which has been introduced often brings with it an undesirable degree of artificiality in the participants in the experiment. Under such circumstances we may well be able to say that a certain relationship exists in such an artificial environment; but this does not necessarily mean that it also holds in a natural environment. If it does not, then we have a case of external invalidity due to the reactive arrangements of the experiment

This artificial reaction to experimental situations is sometimes called the "Hawthorne effect." A field experiment was conducted at the Hawthorne Works of the General Electric Company in Chicago during the 1920's. In examining the impact of changes in physical working conditions upon production, researchers found that the mere presence of the experiment positively influenced production, regardless of the particular experimental conditions.

12 F. J. Roelhlisberger and W. J. Dickson, *Management and the Worker*, Cambridge, Mass.: Harvard University Press, 1939.

Multiple Treatment Interference. An advertising agency is interested in evaluating the effectiveness of five television commercials for a brand of television sets. Because of the costs involved in testing, they decide to use only one city. They show one commercial every night for a week and then observe television sales for the next month. This process continues until all five of the commercials have been examined. Based upon the analysis of the results, the agency decides that the commercial shown last was most effective because sales for the last month were higher than in the other four observational months.

The inference that the last commercial is best is fallacious for two reasons. First, the sales impact of the first commercial certainly did not end with the month which was set aside to measure its impact. Second, the first commercial might have been effective in shaping the opinions of its viewers and thus enabled the subsequent commercials to be more effective. If this second possibility was true, then it would be an example of multiple treatment interference. *Multiple treatment interaction* occurs when experiments are conducted in sequence and there is a particular compounding effect which would not have occurred if the sequence had been varied or if the experiments had not been conducted sequentially.

In summary, many of the *factors* that can cause internal invalidity can also cause external invalidity. What distinguishes them is the effect of these factors on the quality of the information obtained from a study. If one of these factors provides the basis for an alternative explanation of the inferred causal relation, then it is a case of internal invalidity. If on the other hand, the factor is capable of creating an artificial set of circumstances which restricts the possibilities of generalizing the results of the study, then it is an instance of external invalidity.

CONCLUSION

Errors in applied business research tend to fall within one of three categories: errors of definition, errors of estimation, and errors of explanation. Errors of definition *do not* occur if the correct decisions are made as to what information to obtain. Errors of estimation do not occur if that information is obtained accurately. Errors of explanation are avoided if only the proper inferences are drawn from that information,

The astute reader will recognize that the last three statements are incorrect in implication. That is, it is never possible for *any* study, no matter how well designed, to be error free. Certainly some studies can be excellent decision-making tools, while others are actually harmful. However, all studies contain some of the types of error just discussed, A manager who is successful in using research as a decision-making tool recognizes this fact. He is capable of rejecting the research outright if these problems are too severe. He is also capable of qualifying and revising the results of a study if the problems are not so severe as to make them meaningless.

It is not necessary for a manager to know the proper name of a particular type of error in order to be able to recognize it and deal with it. What he should always know is the types of error which might afflict the research he is dependent upon, as well as how to spot the types of errors if they do occur. To this end an appendix has been included. It consists of a series of questions the manager should ask about the research he hopes to use. The numbers after each of these questions refer to the numbering of the potential sources of error as they were presented in Figure 1. By answering these questions the manager should be in a better position to evaluate the results of a research project.

These questions can also be used by the manager in his evaluation of a proposal for research. The fact that satisfactory answers cannot be provided regarding a particular question may well indicate that a change should be made in the research design. The *exact* nature of these changes may involve technical considerations beyond the scope of this paper. However, the researcher should be able to translate such criticisms by the manager into substantive improvements in the research design. In this way the manager can improve the chances that the results will be accurate as well as useful.

Finally, it is the hope of the author that the reader will not reject the possibility of using research in order to make better decisions, on the basis of this discussion of the many pitfalls which occur while conducting and using research. Rather, it is hoped that this discussion will help the manager to *utilize* the potential of applied business research as a means of reducing the uncertainty involved in decision making.

BIBLIOGRAPHY

Blalock, H. M., Jr., and Blalock, A. B. (eds.) *Methodology in Social Research,* New York: McGraw-Hill Book Company, 1968.

Boyd, H. W., Jr., and Westfall, R., *Marketing Research,* Homewood, Illinois: R. D. Irwin, Inc., 1972.

Campbell, D.T., and Stanley, J.C., *Experimental and Quasi-Experimental Designs for Research,* New York: Rand McNally and Company, 1963.

Coombs, C. H., *A Theory of Data,* New York: John Wiley and Sons, 1964.

Cox, E. P., III., Anderson, W.T., Jr., and Fulcher, D. G., "Reappraising Mail Survey Response Rates," *Journal of Marketing Research* (November, 1974).

Crano, W. D., and Brewer,.M. B., *Principles of Research in Social Psycbotogy,* New York: McGraw-Hill Book Company, *1973.*

Ghiselli, E. E., *The Theory of Psychological Measurement,* New York; McGaw-Hill Book ; Company, 1964.

Green, P. E., and Tull, D. S., *Research for Marketing Decisions,* Englewood Cliffs; Prentice Hall, 1966.

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

Kinnear, T, C,, and Taylor, J_ R., "Multivariate Methods in Marketing Research: A Further Attempt at Classification," *Journal of Marketing,* 34 (October 1971), pp. 56-59.

Mayer, C, S., "Application of Bayesian Statistics to Research Design," in Albaum, G., and Venkatesan (cds,). *Scientific Marketing Research,* New York: The Free Press, pp. 86-100.

Roethlisberger, F, J., and Dickson, W. J., *Management and the Worker,* Cambridge, Mass.: Harvard University Press, 1939.

Sheth, J. M., "The Multivariate Revolution in Marketing Research," *Journal of Marketing* 34 (January 1971): 13-19.

Siegel, S., *Nonparametric Statistics,* New York: McGraw-Hill Book Company, 1956.

Simon, J, L., *Basic Research Methods in Social Science,* New York; Random House, 1969.

Torgerson, W. S., *Theory and Methods of Scaling,* New York: John Wiley and Sons, 1960, Tuckman, B. W-, *Conducting Educational Research*, New York: Harcourt, Brace, Jovanovich, Inc. 1972.

Uhl, K. P., and Schoner, *B*,, *Marketing Research*, New York: John Wiley and Sons**,** Inc,, 1969.

Uhr**,** E. B. and Wilcox, J. B., "Application of Muitivariatic Techniques in Marketing," *Proceedings, Southeastern Chapter*, The of Management Sciences, October 1973,pp, **332-35.**

Webb, E. J, et. al., *Unobtrusive Measures; Nonreactive Research in the Social Sciences*, Chicago; Rand McNdl^, 1966.