









HD28

.M414

no. 409-  
69

WORKING PAPER

ALFRED P. SLOAN SCHOOL OF MANAGEMENT

NOTES ON THE APPLICATION OF ATTITUDE MEASUREMENT  
AND SCALING TECHNIQUES IN MARKETING RESEARCH

Alvin J. Silk

Sloan School of Management

Massachusetts Institute of Technology

Working Paper 409-69

MASSACHUSETTS  
INSTITUTE OF TECHNOLOGY  
50 MEMORIAL DRIVE  
CAMBRIDGE, MASSACHUSETTS 02139



NOTES ON THE APPLICATION OF ATTITUDE MEASUREMENT  
AND SCALING TECHNIQUES IN MARKETING RESEARCH\*

Alvin J. Silk

Sloan School of Management

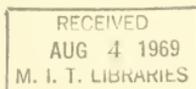
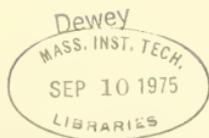
Massachusetts Institute of Technology

Working Paper 409-69

\*

Prepared for use in the "Management Science in Marketing"  
Summer Session, Sloan School of Management, M.I.T., July 14-25, 1969.

HL-8  
11-8  
no. 407-w



Scaling methods and attitude measures have been used by marketing researchers for many years. Use of these techniques for certain measurement problems is becoming fairly common and the list of areas of application is growing. The demand for more and better measurement plus the promise of future developments require that the management science/marketing research practitioner have some familiarity with methods found in a large and scattered literature that he is likely to find foreign if not complex. The scaling/attitude measurement literature draws on the work of three different research traditions. Much of scaling theory is rooted in psychophysics where the principal concern is sensory measurement. The second relevant field is psychometrics where attention is focused on mental testing. Finally there is the work on attitude measurement in social psychology and sociology. The present note is intended to serve as a (very) brief introduction to these materials with primary emphasis given to attitude research. The latter orientation is adopted in a great many types of marketing research studies and the subject of much misunderstanding and controversy. Some examples of scales used in marketing studies will be presented to illustrate methods and problems. A short list of useful references is also attached. The treatment of attitude scaling and measurement in the marketing literature currently available is very limited. Green and Tull<sup>1</sup> devote a chapter to the subject but their revision which is expected to appear shortly will contain a much more extensive coverage of this material. George Day and Michael Ray of Stanford are developing a volume on attitude research in marketing. Although it deals largely with a different set of techniques from those discussed here, mention should be made of an interim technical monograph on nonmetric scaling and related techniques prepared by Paul Green, and his associates and available from the Marketing Science Institute.<sup>2</sup> Presently however, anyone interested in utilizing attitude research methodology must seek answers to his questions in the behavioral science literature. A list of basic sources is attached.

---

<sup>1</sup> Paul E. Green and Donald S. Tull, Research for Marketing Decisions (Prentice-Hall, 1966), Chapter 7.

<sup>2</sup> Paul E. Green, Frank J. Carmone, and Patrick J. Robinson, Analysis of Marketing Behavior Using Nonmetric Scaling and Related Techniques (Marketing Science Institute, March, 1968).



## The Concept of Attitude

Before discussing procedures for measuring attitudes, we should briefly examine the concept of attitude itself. The number of definitions of "attitude" that have been proposed is vast--in 1935 Allport was able to find more than one hundred different definitions that had appeared in the literature! Consider a few examples:

"...a mental and neural state of readiness exerting a directive influence upon individual's response to all objects and situations with which it is related.

...the probability of occurrence of a defined behavior in a defined situation.

...an enduring system of three components centering about a single object--positive or negative evaluations or beliefs (cognitive component), emotional feelings (affective component), and disposition to take action (action tendency component).

Terms such as brand loyalty, brand preference, brand image, purchasing propensity etc. refer to concepts that could be encompassed by one or more of the above definitions. At first glance, it appears that these definitions indicate different conceptualizations of attitude. The first definition utilizes the concept of set or readiness to respond reflecting the mentalist tradition in psychology. Psychologists of a behavioristic orientation would be more sympathetic with the second definition which depicts attitude as a response rather than a set to respond. The third definition cited above is concerned with the composition of attitudes. Despite the apparent differences among the definitions, there are some common elements that run through most if not all of them. What is perhaps even more important, when it comes down to actually measuring attitudes the points of conceptual controversy seem to be forgotten or ignored. The following are some important properties of attitudes:

### 1. Hypothetical or latent variable.

An attitude is not directly observable but rather is inferred. An attitude cannot be diagnosed from any one particular act or response but rather is abstracted from a large number of related acts or responses.

### 2. Measurement based on Response Consistency or Covariation

As both Campbell(6) and Green(10) as well as others have pointed out, while the many definitions of attitude that have been proposed differ in various ways, they all imply that the concept of attitude involves a consistency or predictability of responses. Furthermore, covariation among responses is basic to all the methods used to measure



attitudes. Campbell(6) proposed the following as an operational definition of attitude:

...a social attitude is (or is evidenced by) by consistency in response to social objects.

Response may be measured by self-reports of past behavior toward the object of the attitude or by written or verbal enunciations of statements about beliefs, feelings, or intentions involving the object presented in an interview or self-administered questionnaire. Regardless of the procedure used, the existence of a pattern of interrelationships among responses is typically the evidence used to diagnose an attitude. As will be seen shortly, the various attitude measurement and scaling techniques have approached the matter of "response consistency" in somewhat different ways.

This notion of diagnosing attitudes on the basis of response consistency is essentially the process that we use daily when we characterize a politician as "liberal," a family as "child-oriented," a friend as a "gourmet," or a neighbor as a "big Scotch drinker." Clearly, in making such designations we have in mind some pattern of response across a set of behaviors, stimuli, occasions, etc. Hence, when we ask a person to respond to an attitude scale consisting of a series of statements or items, we are taking a sample of an attitude universe.

### 3. Unidimensional Concept

The bipolarity implicit in most definitions of attitude suggests a simple unidimensional concept--like-dislike, favorable-unfavorable evaluation, pro-con action tendency. This is not to say that attitudes may not be multidimensional. Rather, researchers typically begin with a unidimensional conception of attitude and in developing measuring instruments they aspire to a one dimensional scale, and treat the scale score as a measure of a single variable. If a scale is unidimensional, then people with the same score will have about the same attitude. However, if the scale measures say two components, then the same score can be obtained in several different ways. Whether or not a given attitude domain or scale is unidimensional is an empirical question that the researcher must consider. Methods of investigating this issue will be discussed later.

### 4. Attitudes are learned or "residues of experience."

#### Properties of Scales

Stevens(1) defines measurement as "the assignment of numerals to objects, events, or person, according to rules." The result of a measurement is a scale. The rules employed in making the measurement determine the properties of the scale. The following is a widely used scheme for classifying scales:



1. Nominal Scales  
Objects are assigned to mutually exclusive categories but there is no necessary relation among the categories. All that is involved is classification and labelling. Assignment of numbers to categories is arbitrary. Torgerson(2) does not regard this as measurement in the sense of assigning numbers so as to signify the relative amount or degree of a property possessed by an object.
2. Ordinal Scale  
This level of measurement is achieved when objects can be arranged in rank order according to a variable. Numbers are assigned so that they are in the same rank order as the objects. Any monotonic transformation is permissible.
3. Interval Scale  
Here numerically equal differences on the scale represent equal differences in the property being measured. If such measurement were achieved in the case of a "willing-to-buy" scale, we could say that two people with respective scores of 2 and 4 differed by the same amount regarding this disposition as two other persons who scored 8 and 10 respectively. Transformation of interval scales is restricted to those involving positive linear relationships.
4. Ratio Scale  
This type of scale results when there is some way of showing how many times greater one object is than another. An example of this type of measurement is weight. A ratio scale implies a fixed zero point so that the only admissible transformation is multiplication by a constant.

This classification is not complete since there are some intermediate cases. Aside from the nominal variety, ordinal measures are most common in attitude research although there are procedures for constructing interval scales if one is willing to live with certain assumptions. An oft-debated issue is the applicability of various statistical operations and tests to these different types of scales.

#### Approaches to Scaling

Torgerson(2) distinguishes between the following:

1. Judgement Approach  
The systematic variation in the reactions of the subjects to stimuli is attributed to differences in the stimuli with respect to a designated attribute.

This approach would correspond to a marketing research problem where we wanted the consumers to judge the flavor of say coffees according to strength. The property, strength, is specified in advance and the task of the respondent is to order various brands of coffee along a "very strong" to "very weak" continuum.



## 2. Response Approach

Variability of reactions to stimuli is ascribed to both variation in the subjects and in the stimuli. No simple judgemental continuum is presupposed.

Such is the situation which obtains in a typical brand image study where consumers are asked to rate a brand on a series of scales referring to different product qualities, performance characteristics, etc.

In the judgement approach, respondents are assumed to be homogeneous and the stimuli are scaled along some assumed continuum for a specified property. In the response approach, variability or differences with respect to both respondents and stimuli is investigated.

Scaling methods employing these two approaches and examples of their application in marketing research are listed below. A detailed discussion (with examples) of the various techniques of attitude scale construction may be found in (12).

### Types of Attitude Scales

#### 1. Differential Scales

A differential scale consists of a number of items whose position on the scale has been determined by some ranking or rating procedure carried out by the judges.

The pioneering work on this type of scale was done by Thurstone. His procedures are examples of the judgement approach and represent attempts to develop interval scales. Thurstone developed various methods for securing judgements of scale position from judges.

##### a) Paired comparisons

Thurstone proposed a "law of comparative judgement" which, under certain assumptions, provides a means for developing an interval scale from comparative proportions.

Green and Tull give an example of the use of this method to determine the position of 5 different varieties of tomato juice on a "flavor" scale.<sup>4</sup>

A sample of respondents were given samples of the 5 brands--two at a time for all possible pairings (a total of 10 pairs). For each pair, respondents were asked to taste the sample and indicate which of the two they preferred. Suppose that 90% of respondents prefer A over B but only 55% prefer B over C. The intuitive idea underlying Thurstone's method is that the difference between the scale positions or values of A and B is larger than the difference in scale positions between B and C. In other words, it is assumed that the magnitude of

---

<sup>4</sup> Green and Tull, *op. cit.*, p.194 ff.



the perceived difference between A and B with respect to "flavor" is some function of the proportion of respondents who prefer A over that of B. Using this and some additional assumptions, Thurstone formulated a mathematical model which can be used to obtain estimates of the position of the five tomato juices along a psychological continuum of flavor from data representing the frequency of preference for one brand over another. Working backwards, one can test how well the estimated scale values predict the original preference proportions. Significance tests have been developed for this purpose.

Of course, there is no assurance that a method such as this will yield a psychological continuum for any particular property. A concept like "flavor" may be multidimensional making unidimensional scales inappropriate. The population may be heterogeneous with respect to their judgements and ability to discriminate among stimuli. However, applications of this methodology has met with some success. Kuehn and Day report that the sweetness of cola drinks, the sudsiness of detergents and the saltiness of margarine are scalable.<sup>5</sup> They have applied preference scaling methods to the problem of determining what attributes consumer products ought to have. A scale is first developed for the product quality of interest and then the distribution of consumer preferences over various values of the scale is estimated. The latter step is accomplished by conducting paired comparison tests with product samples.

#### b) Equal-Appearing Intervals

The method of paired comparisons becomes unwieldy when the number of objects or items one wishes to scale becomes large since all possible pairs must be judged. The method of equal-appearing intervals may be used as a short-cut in such situations. Respondents are required to make only one judgement per item or object. For this reason, the method has been frequently used in constructing attitude scales where one starts out with a large number of items (statements, adjectives, etc.) and hones down the list so as to end up with a multiple-item scale consisting of perhaps 5-10 items. Judges are presented with a set of items (sometimes a hundred or more) and asked to sort them into fixed number of categories (usually eleven) along some continuum like favorable-unfavorable. Typically the judges are instructed to place the items in categories so that the intervals between the categories are subjectively equal. In the first category a judge places the items he considers most favorable to the object; in the second, those he considers next most favorable; in the last category, those he regards as most unfavorable. The sixth category is defined as the "neutral" position. The scale value of an item is computed as the median value (or category) to which it was assigned by the judges. Items whose ratings have a large variance are discarded. A series of items are chosen to form a scale. The position of each item on a scale of favorable-unfavorable attitude toward the object studied has been determined by the judges' classifications. The resulting multiple-item scale then becomes an instrument to measure the attitude in question.



The set of items are presented to subjects in a questionnaire(or interview) and they are asked either to check all the statements with which they agree or to check the two or three that are closest to their position. The mean or median of the scale values (determined by the previous judging) of the items checked is taken as a measure of the strength of his attitude.

Myers and Warner have used this method to develop rating scales for evaluating products and advertisements. A common marketing research procedure is give consumers a product to use and ask them to rate its performance on a five or seven point evaluative scale consisting of descriptive words or phrases such as "Very Good"(7), "Good"(6), etc. to "Very Poor" (1). They suggest that the descriptive terms included in such scales may mean different things to different people and undertook a study to develop rating scales of this general type that had interval properties. Using procedures similar to those discussed above, samples of housewives, students, and business executives were asked to rate 50 commonly used statements applicable to products or advertisements on a 21 point scale. The means and standard deviations were calculated from the ratings for each of the 50 statements. Myers and Warner suggest that these data can be used to select statements that will form interval scales. Suppose one required a 5 point evaluative scale. From their list, the following descriptive terms would be selected to constitute an approximation of an interval scale:

|                 | Mean | Standard Deviation |
|-----------------|------|--------------------|
| Remarkably Good | 17   | 2.2                |
| Good            | 14   | 2.4                |
| Neutral         | 10   | 1.9                |
| Reasonably Poor | 6    | 2.0                |
| Extremely Poor  | 3    | 1.7                |

While the kind of information Myers and Warner provide is clearly useful, there is really no way of determining whether in fact, the intervals are subjectively equal.

## 2. Cumulative Scales

The response approach to attitude measurement is illustrated by Guttman's scalogram analysis. A Guttman scale is an ordinal scale. Respondents are asked to agree or disagree with a series of items or statements--frequently, but not necessarily, only a dichotomous judgement is required. An individual's score is obtained by summing the number of items with which he agrees. If items form a Guttman scale, they have a special cumulative property. In particular, the items are related to one another such that if an individual agrees with the second item he also agrees with the first; someone who agrees with the third also accepts the first two, and so on. Thus, all those who indicate agreement with a given item should have a higher total score on the entire scale than individuals who disagree with that particular item.

6

James H. Myers and W. Gregory Warner, "Semantic Properties of Selective Evaluative Adjectives," Journal of Marketing Research, (Vo. 5, No. 4 (Nov., 1968), pp. 409-413.



If these relationships hold, then given only knowledge of an individual's total score on the entire scale, one could reproduce exactly the pattern of his responses to the individual items.

These ideas may be illustrated by reference to some data from a study conducted by Paroush involving an unusual application of Guttman scaling. Paroush was interested in determining whether the order in which consumers acquired durable goods followed any kind of predictable order or pattern. Data from a survey of ownership of four appliances were used to develop the following table:

| Scale Score       | Cooker | Vacuum | Washer | Refrigerator | % of Families Owning |
|-------------------|--------|--------|--------|--------------|----------------------|
| 4                 | x      | x      | x      | x            | 6.4                  |
| 3                 | x      | x      | x      |              | 17.7                 |
| 2                 | x      | x      |        |              | 34.7                 |
| 1                 | x      |        |        |              | 30.6                 |
| 0                 |        |        |        |              | <u>1.1</u>           |
|                   |        |        |        |              | 90.5                 |
| <b>Deviations</b> |        |        |        |              |                      |
| 3                 | x      | x      |        | x            | 5.5                  |
| 2                 | x      |        | x      |              | 2.6                  |
| 2                 | x      |        |        | x            | 1.0                  |
| 1                 |        | x      |        |              | 0.2                  |
| 1                 |        |        | x      |              | <u>0.1</u>           |
|                   |        |        |        |              | 9.4                  |

Suppose we ask whether the ownership of these four appliances form a Guttman scale. Owning or not owning a particular appliance can be taken as analogous to agreeing or disagreeing with an item in an attitude scale. Owning an appliance is scored "1", not owning "0". Each family receives a total score equal to the number of appliances they own. Scores can range from 0 to 4. The data in the upper part of the table indicate a perfect Guttman scale. Those families owning only one appliance all have cookers; those owning two appliances have a cooker and a vacuum; the three appliance owners have a cooker, vacuum, and a washer. As well there are those who have all 4. If the five patterns in the upper part of the table accounted for ownership patterns of the entire sample we would have a perfectly reproducible scale. It would be perfectly reproducible in the sense if we know how many appliances a family has in total then we can say exactly which particular appliances that family does or does not own.

Note however, there are "deviations"--departures from the pattern shown in the top part of the table. There were, for example, instances of families owning a combination of three appliances different from that represented in the top half of the table. In fact, the pattern of the



top half of the table was established so as to make the number of deviations as small as possible. The presence of these deviations means we could not reproduce which appliances a family owned with complete accuracy knowing only how many appliances they owned in total--e.g., not all families having three appliances own the exact same ones. The proportion of cases which are deviations (referred to as "errors" by Guttman) is one of the criteria used to evaluate how closely a set of items approximates a perfect unidimensional scale. As a measure of the extent to which this criteria is met, Guttman defines a quantity termed the "coefficient of reproducibility" ( $1 - \text{no. of errors}/\text{no. of item responses}$ ). The coefficient takes on a value of 1 for a perfect scale. The lower limit depends on the marginal distributions of responses to the individual for items. Guttman suggests that a coefficient value of .90 is needed for items to constitute a satisfactory approximation of a scale. The coefficient of reproducibility for the data shown in the table is .976. Guttman mentions other criteria for judging scalability in addition to this coefficient.

A more conventional application of Guttman scaling in marketing research may be found in the work done by Wells in developing scales<sup>8</sup> to measure several dimensions of consumers reactions to advertisements.

### 3. Summated Scales

An attitude scale of this variety again consists of a series of items. However, unlike the Thurstone or Guttman scales, no attempt is made to find items that distribute evenly over the attitudinal continuum. The type of summated scale most often used in attitude research is that associated with the name of Likert. Respondents are presented with a series of statements either definitely favorable or unfavorable and the object under study. Neutral items are not used. The respondent indicates a favorable disposition toward the object or viewpoint being investigated. An individual's scale score is the sum of his score on the items. The scaling model implied by this procedure is that all items measure a single common factor. Were this not the case, it would not make sense to sum the individual item scores to arrive at a person's scale score. Items are selected so that the scale is internally consistent or homogeneous. Items are evaluated in various ways such as examining the correlation between responses to an item and the total score for the entire scale. Items that do not discriminate between high and low total scores are discarded. Another (imperfect) indication of internal consistency is the magnitude of the inter-item correlations. The procedures used to select items are very similar to those used in developing mental tests. Various indices and tests of homogeneity exist; Note that operationally, homogeneity, internal consistency, and single-factor-ness or unidimensionality all have essentially the same meaning here.

Another way of verifying the presence of a common factor is to apply factor analysis (or a related technique like cluster analysis or latent structure analysis) to the inter-item correlations. This is frequently done in marketing studies where scales are developed on an ad hoc (or post hoc) basis and there is little or no basis for developing prior expectations about the structure of the attitude being studied. Davis' investigation of



the relative influence of husbands and wives in family purchases of automobiles and furniture provides an illustration of the use of the Likert-type scale in a consumer study. Davis asked a sample of married couples questions of the following kind (husbands and wives separately):

Who decided what make of car to buy?

Who decided where the car would be purchased?

Who decided what color to buy?

Responses were made to each on the following five-point scale:

|         |           |         |           |      |
|---------|-----------|---------|-----------|------|
|         | Husband   | Husband | Wife      |      |
| Husband | more than | & wife  | More than | Wife |
|         | wife      | Equally | Husband   |      |
|         | <hr/>     |         |           |      |
|         | 1         | 2       | 3         | 4    |
|         |           |         |           | 5    |

Application of a simple clustering technique to the matrix of inter-item correlations revealed the presence of two fairly distinct groups of decisions. Davis labelled one set "product selection decisions" (make, model, color, dealer) and the second "allocation or scheduling" of the purchase (when to purchase and how much to spend). Similar clusters were found in both husbands and wives responses and for furniture purchases as well as automobiles.<sup>9</sup>

#### 4. Semantic Differential

This technique is often used in marketing research studies, especially where the problem is to assess the content of corporate or brand images or the effects of advertising on them. Originally developed by Osgood et al. as a method for measuring the "meaning" of an object, to an individual, it can also be viewed as an attitude scaling device.<sup>10</sup> Respondents are asked to rate some concept or object on a series of bipolar adjectives presented as a 7 point scale. The manifest content of the adjectives does not unequivocally indicate the nature of the underlying attitudinal dimension being measured. Instead it is inferred from the manner in which the adjectives are interrelated. Hence, there is the problem of determining and interpreting the number of factors represented by a given set of adjectives similar to that discussed above in connection with Davis' use of a Likert scale. Osgood et al. had subjects rate a great many diverse concepts and objects ("India," Chinese People," Truman, "Christianity,") on a great many different sets of adjectives. On the basis of results obtained from factor-analytic procedures, they identified three dominant dimensions along which judgements were made and labelled them "evaluative", "potency", and "activity".

The technique has had great appeal to marketing researchers because it lends itself to making comparisons of various kinds.

<sup>9</sup> Harry L. Davis, "Dimensions of Marital Roles in Consumer Decision Making," Journal of Marketing Research, forthcoming.

<sup>10</sup> Charles Osgood et al., The Measurement of Meaning (Urbana, Illinois: University of Illinois Press, 1957).



For example, the ratings on a set of bipolar adjectives given two or more competing brands by a sample of consumers are often used to develop brand "profiles" and identify competitive strengths and weaknesses. The manner in which different groups, say users and non-users, perceive the same brand may also be compared. In making such applications, marketing researchers have frequently borrowed the adjectives that Osgood et al. found to load on the evaluative, potency, and activity factors on a piecemeal basis without considering whether consumers find these terms meaningful when applied to products. What does a consumer have in mind when he rates a Ford on a "stale-fresh", "obvious-subtle" scale? He may well make a check mark on questionnaire when asked to do so but how does one interpret it? The fact that the adjectives used by Osgood et al. when applied to brands and products have not always been found to result in the same factor structure as these authors found in their original studies certainly casts doubt on the practice of simply borrowing their scales as a matter of convenience. A notable exception is Wells' work in developing a semantic differential-type instrument called the "Reaction-Profile" for measuring consumers' evaluations of print advertising.<sup>11</sup> There have been numerous other applications of the semantic differential reported in the marketing literature.<sup>12</sup>

### The Evaluation of Scales

In order to assess the merit of a scale, we require statistical evidence concerning its properties, the most important of which are reliability and validity.

#### A. Reliability

The reliability of an instrument is an index of the extent to which repeated measurements yield consistent results. There are two aspects to reliability: stability and equivalence.

##### 1. Stability

Stability is another name for test-retest reliability. The correlation between scores obtained from the same sample of respondents on two separate occasions is an indication of stability. Note that the length of time that lapses between successive administrations of the scale to respondents will affect the degree of stability a scale exhibits. Over extended periods of time, "real" changes may take place in the attitude being studied thereby reducing the reliability of the scale. If the interval is short, memory and familiarity with the scale will tend to inflate estimates of stability. It is only rarely that one sees evidence of test-retest reliability reported in marketing studies.<sup>13</sup>

##### 2. Equivalence

Equivalence corresponds to the notion of internal consistency, item

11

Wells, op. cit.

12

See for example, William Mindak, "Fitting the Semantic Differential to the Marketing Problem," Journal of Marketing, Vol. ,No.2(April, 1961, pp.28-33, and Al. E. Birdwell, "A Study of the Influence of Image Congruence on Consumer Choice," Journal of Business, Vol.41, No.1 ( January, 1968), pp.76-88.

13

However, see Joel Axelrod, "Attitudes That Predict Purchase," Journal of Advertising Research, Vol.8, No.1(March, 1968),pp.3-17 and Douglas J. Tigert, "Psychographics: A Test-Retest Reliability Analysis" Presented at the 1969 Fall Conference of the American Marketing Association, Cincinnati, Ohio.



homogeneity, and unidimensionality. That is, another way of looking at the consistency with which a scale measures a given attitude is to examine the extent to which the various items comprising the scale measure the same thing. As Green puts it, "If our purpose is to measure an attitude universe by means of a sample of items, then we must determine what score differences could be expected if a different sample of items were chosen." (10, p.339) One early approach to evaluating equivalence was to estimate "split-half" reliability. The items in a scale are randomly split into two subsets and a total score is computed for each. The correlation between the two scores obtained for each half is an index of equivalence. A related measure often reported is the Kuder-Richardson statistic which is essentially the average of all possible split-half reliabilities for a given set of items. For a Guttman scale, the coefficient of reproducibility mentioned earlier is a measure of unidimensionality. Other indices have been proposed as well (see 10).

## B. Validity

For a scale to be considered valid we need evidence that it does indeed measure whatever it is that it purports to measure. Operationally, reliability and validity can be viewed as falling along a continuum:

Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods.<sup>14</sup>

There are several possible approaches to establishing validity of a scale.

### 1. Pragmatic Validity

This involves determining whether a correlation exists between scores obtained on an attitude scale and some external criterion variable.

#### a. Concurrent Validity

A scale that can distinguish among individuals who differ according to some aspect of their current status is said to possess concurrent validity. As an example, we would expect that current users of a given brand would score higher on a scale designed to measure attitudes toward it than would current non-users. This has often been shown to be the case.<sup>15</sup>

#### b. Predictive Validity

We might also investigate the extent to which we could predict the future status of individuals on the basis of an attitude measure.

To illustrate, Wells administered an eleven point "willingness to buy" scale to a sample of 900 housewives with reference to

---

14

Donald T. Campbell and Donald W. Fiske, "Convergent and Discriminant Validation By the Multitrait-Multimethod Matrix," Psychological Bulletin, Vol. 56, No. 2 (March, 1959), p.83.

15

See, for example, George H. Brown, "Measuring Consumer Attitudes Toward Products," Journal of Marketing, Vol. 14 (Apr., 1950), pp.691-698.



toilet goods and grocery items. He later reinterviewed them to determine whether those who indicated a strong disposition to buy on the scale had actually purchased the brand within four weeks after making the ratings. He reports finding some correlation between his willingness-to-buy measure and subsequent purchases but the strength of the relationship varied markedly among the different brands studied.<sup>6</sup>

The two examples of validity mentioned here both involved correlating attitude scores and criterion measures for the same sample of respondents. Another approach is to perform correlations with aggregate data from cross-sectional or time series studies.<sup>17</sup>

## 2. Construct Validity

Psychologists have frequently been faced with situations where they wish to develop a measure of some trait but the nature of the trait is such that it cannot be readily identified in some specific kind of behavior. Under these circumstances, it may be difficult or impossible to find a suitable criterion variable for use in establishing concurrent or predictive validity. Personality and intelligence are examples of constructs which have appeal as abstractions but which do not manifest themselves in any simple fashion. What has been done in such situations is to attempt to measure the same construct by a different, independent method and correlate the two measures. Confirmation by independent measurement procedures is called "convergent validation." However, such a correlation may be suspect when, for example, the two independent methods are both paper and pencil tests. A competing hypothesis might be that the two tests really do not measure the same construct but rather the observed correlation is merely the result of some extraneous methods bias such as "yeasaying" or a tendency to attribute socially desirable traits to oneself. We require a method for assessing whether the convergence between two independent measures of the same trait or construct is inflated by shared methods variance.

A technique for making such an assessment has been proposed by Campbell and Fiske and utilizes what they call a "multitrait-multimethod matrix." Their method requires multiple measures--each of several traits must be measured by at least two methods--maximally independent or different. All measures of all traits are intercorrelated. Significant positive correlations between different measures of the same trait constitutes evidence of convergent validity. However, tests can be invalidated by correlating too highly with other tests purporting to measure different traits. Hence, a variable should correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method. Meeting this condition is an indication of discriminant validity.

16

William D. Wells, "Measuring Readiness to Buy," Harvard Business Review, Vol. 39, No. 4(July-August, 1961), pp. 81-87.

17

See, for example, M. Bird and A.S.C. Ehrenberg, "Intentions-To-Buy and Claimed Brand Usage," Operational Research Quarterly, Vol. 17, No.1 (March, 1966), pp.27-46.



Evidence of the reliability and validity of measures used is only very rarely presented in marketing studies. Often it is argued that time and budget pressures do not permit much effort to be expended on these matters that are so common. With some forethought the Campbell-Fiske procedure could often be applied in marketing research studies. To date, there has been little utilization of it in the marketing field. Davis developed a multitrait-multimethod matrix to assess the validity of several different ways of measuring the relative influence of husbands and wives in decisions involving the purchase of durable goods.<sup>18</sup> The present writer has also recently made a partial application of the approach.

The recent controversy over the relationship between attitudes and behavior has forced marketing researchers to consider the issues of reliability and validity and some valuable studies are now beginning to appear. One that deserves the attention of all who use these types of measures is Axelrod's "Attitude Measures that Predict Purchase."<sup>20</sup> The available evidence concerning the relationship between buyers attitudes and their purchasing behavior has been insightfully reviewed recently by Day.<sup>21</sup>

### Unobtrusive Measures

The great bulk of behavioral research utilizes data obtained from questionnaires and interviews. Webb et al.(15) argue that while there are many well-known limitations to these data collection procedures, the most serious objection to them is that they are used alone. Given that virtually every method has a bias of one kind or another, they stress the need for multiple measures--utilization of various approaches that have different methodological weaknesses so as to obtain a number of measures of the same variable.<sup>22</sup> They combed the social science literature searching for studies that were based on data not obtained from interviews or questionnaires and summarized them in a volume that for a time had the working title, Oddball Research. It provides exciting reading that is also worthwhile to marketing researchers if only because it will serve to place them on the lookout for "unobtrusive measures." Below are the principal types of nonreactive methods they list along with some examples relevant to marketing measurement:

18

Harry L. Davis, "On the Measurement of Husband-Wife Influence in the Purchase of Consumer Durables," Paper presented at the Conference on Research Methodology held by the New York chapter of the American Marketing Association, April 10, 1969.

19

Alvin J. Silk, "Response Set and the Measurement of Self-Designated Opinion Leadership," unpublished working paper, Sloan School of Management, M.I.T., July, 1969

20

Axelrod, loc. cit.

21

George Day, Buyer Attitudes and Brand Choice(New York: Free Press, 1969).

22

For an excellent example of a study that utilizes both multiple and unobtrusive measures to assess political attitudes, see Michael L. Ray, "Neglected Problems (Opportunities) in Research: The Development of Multiple and Unobtrusive Measurement" in Robert L. King, ed., Marketing and the New Science of Planning(1968 Fall Conference Proceedings, Series No.28; Chicago:American Marketing Association, 1968, pp.334-340.



1. Physical Traces--"erosion and accretion," evidence of past behavior

A Chicago automobile dealer has his mechanic determine what radio stations cars being serviced are tuned to. He used these tallies to evaluate the potential of different radio stations as advertising media.

2. Archival Record

A time series based on liquor and life insurance sales at airports was used to develop an index of the anxiety aroused by plane crashes among air travelers.

3. Observation

College students were paid to spy on members of their families while the latter watched television as a means of obtaining data on audience behavior during the airing of commercials.



## REFERENCES

### Scaling and Measurement

1. S.S. Stevens, "Mathematics, Measurement, and Psychophysics," in S.S. Stevens, ed., Handbook of Experimental Psychology (New York: Wiley, 1951), pp. 1-49
2. Warren S. Torgerson, Theory and Methods of Scaling (New York: Wiley, 1958)
3. Clyde H. Coombs, "Theory and Methods of Social Measurement," in Leon Festinger and Daniel Katz, eds., Research Methods in the Behavioral Sciences (New York: Holt, Rinehart and Winston, 1953), pp. 471-535
4. \_\_\_\_\_, A Theory of Data (New York: Wiley, 1964)
5. Joseph L. Zinnes, "Scaling" in Paul H. Mussen and Mark R. Rosenweig, eds., Annual Review of Psychology, Vol. 20 (Palo Alto, Calif.: Annual Reviews, Inc., 1969), pp. 447-478

### Attitudes and Attitude Measurement

#### a) General

6. Donald T. Campbell, "Social Attitudes and Other Acquired Behavioral Dispositions," in Sigmund Koch ed., Psychology: A Study of Science, Vol. 6 (New York: McGraw-Hill, 1963), pp. 94-172
7. Martin Fishbein, ed., Readings in Attitude Theory and Measurement (New York: Wiley, 1967)

#### b) Non-Technical Discussions of Attitude Measurement

8. Claire Selltitz et. al., Research Methods in Social Relations, rev. ed. (New York: Holt, 1959), Chaps. 5-8, 10
9. Harry S. Upshaw, "Attitude Measurement," in Hubert M Blalock, Jr. and Ann B. Blalock, eds., Methodology in Social Research (New York: McGraw-Hill, 1968), pp. 60-111

#### c) Technical Reviews

10. Bert F. Green, "Attitude Measurement," in Gardner Lindzey, ed., Handbook of Social Psychology, Vol. 1 (Cambridge, Mass.: Addison-Wesley, 1954), pp. 335-369
11. William A. Scott, "Attitude Measurement," in Gardner Lindzey and Elliot Aronson, eds., Handbook of Social Psychology, Vol. 2 (Rev. Ed.,: Reading, Mass.: Addison-Wesley, 1968), pp. 204-273



d) Technical Volumes

12. Allen L. Edwards, Techniques of Attitude Scale Construction (New York: Appelton-Century-Crofts, 1957)
13. S.A. Stouffer, et al., Measurement and Prediction (New York: Science Edition, Wiley, 1950) Paperback

Psychometrics

14. J. Guilford, Psychometrics, 2nd ed. (New York: McGraw-Hill, 1954)

Other

15. Eugene J. Webb, et. al., Unobtrusive Measures: Non Reactive Research in the Social Sciences (Chicago: Rand-McNally, 1966)

FEB 23 '72

NOV 3 '71







RESERVE  
Date Due

MAR 17 77

MAY 11 79

JUN 8 78

DEC 4 1980  
OCT 9 81

MIT LIBRARIES  
3 9080 003 874 861

409-69

MIT LIBRARIES  
3 9080 003 874 853

411-69

MIT LIBRARIES  
3 9080 003 874 846

412-69

MIT LIBRARIES  
3 9080 003 874 838

413-69

MIT LIBRARIES  
3 9080 003 874 812

414-69

MIT LIBRARIES  
3 9080 003 874 820

415-69

MIT LIBRARIES  
3 9080 003 874 804

416-69

MIT LIBRARIES  
3 9080 003 869 879

417-69

MIT LIBRARIES  
3 9080 003 874 622

418-69

MIT LIBRARIES  
3 9080 003 901 664

418-69A

