DOCUMENT RESUME

ED 050 167 TM 000 562

AUTHOR	Evans, G. T.
TITLE	Standardization of Selected Semantic Differential Scales with Secondary School Children.
INSTITUTION	Ontario Inst. for Studies in Education, Toronto.
PUB DATE	Feb 71
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Adjectives, *Attitudes, *Concept Formation, Factor Analysis, Factor Structure, High School Students, Junior High School Students, *Models, *Rating Scales, *Standardized Tests
IDENTIFIERS	*Semantic Differential Scales

ABSTRACT

A basic assumption of this study is that the meaning continuum registered by an adjective pair remains relatively constant over a large universe of concepts and over subjects within a relatively homogeneous population. An attempt was made to validate this assumption by showing the invariance of the factor structure across different types of subjects, and across different classes of concepts. The existence of systematic response tendencies independent of the meaning of the adjectives was demonstrated. Although the substantive results agreed generally with the Evaluation, Activity and Potency dimensions of earlier research, notable differences were found; specifically, that a clear separation between a hedonic response and a judgment of values seems to exist. The data from the 60 developed scales suggests seven useful composites which may be used fairly confidently (with similar samples) to measure attitudes without the need to factor analyze the scales for the new data. (TA)

STANDARDIZATION OF SELECTED SEMANTIC DIFFERENTIAL SCALES WITH SECONDARY SCHOOL CHILDREN.*†

G. T. Evans

The Ontario Institute for Studies in Education

The Semantic Differential Model

A semantic differential scale consists of a pair of adjectives of opposite polarity. Its use with a particular person requires that he react to a given "concept" in terms of these adjectives by placing a mark at an appropriate point on a line drawn between the two bipolar terms.

In the usual 7-point form of the scale the value 4 is regarded as neutral, while values of 1 and 7 are regarded as extremes.

The semantic differential scale has been elaborated following the work of Osgood, Suci and Tannenbaum (1957) into an instrument for recording affective responses to stimuli, for measuring "meaning" in a generalized sense. It is assumed that the verbal judgment responses can be characterized in terms of a few idealised responses known as semantic dimensions, which, when regarded as linearly independent vectors, are said to span the semantic space. A review of the methodology of the semantic differential has been recently made by Heise (1969). The purpose of this section is to extend and formalize some of Heise's observations.

---

*Paper presented at the AERA Annual Meeting, 1971.

†The assistance of Miss B. Kalchman and Mr. J. McLaren, The Ontario Institute for Studies in Education, is gratefully acknowledged.

The model considered below refers to some defined population of persons, and is thus a statistical model in that its parameters refer only to constructs defined over such a population, and are not defined by the responses of a single person. One version of the model is given by

$$(1) \qquad x_{kj} = \alpha_j \left(\sum_{p=1}^{m} a_{jp} f_{pk} + e_{kj} \right) + \beta_j , \qquad \begin{cases} k = 1, \ldots, \nu, \\ j = 1, \ldots, n, \end{cases}$$

where x_{kj} is a random variate interpreted as the response made by a randomly drawn subject on the jth scale to the kth concept, and f_{pk} is the hypothetical response made by the same subject to the kth concept on the pth idealized semantic dimension. The parameter a_{jp} is a regression weight, indicating the regression of the jth scale on the pth semantic dimension, e_{kj} is a random variate indicating the extent to which the semantic dimensions do not account for the random subject's response on the jth scale for the kth concept. The terms α_j and β_j are scaling constants associated with the jth scale which may be used to adjust its mean and dispersion.

In order for the model to be useful, it is necessary to place certain restrictions on the distributions of the quantities x_{kj}, f_{pk}, and e_{kj}. The last two are assumed to have expected values of zero and variances of unity, and β_j may be chosen so that $E(x_{kj})$ is zero. Further, it is supposed that error terms for different concept-scale combinations are uncorrelated with one another and with the "factor" variables, f_{pk}, $p = 1, \ldots, m$; $k = 1, \ldots, \nu$, which refer to the various combinations of concepts and semantic dimensions. The strict assumptions of the statistical model of factor analysis also require that the three sets of variates each have a multivariate normal distribution. In utilizing the model, a two stage strategy may be used. The first stage consists in estimating parameters, a_{jp}, $j = 1, \ldots, n$; $p = 1, \ldots, m$, for the population. The second stage involves measuring the responses, f, on the idealized semantic dimensions for groups of persons and concepts. The second stage depends in the standardization involved in the first.

To estimate parameters like a_{jp} it is necessary to draw a sample of, say, N persons from the population of interest and to select ν appropriate concepts. If the model were treated analogously to that for factor analysis the n data for each person-concept combination would constitute a single multivariate observation, and there would result an n x n correlation matrix based on νN observations. This is Heise's (1969) Design 2 for an ad hoc factor analysis of the scales. Heise discusses the considerations underlying the choice of both persons and concepts.

More frequently, a different interpretation is placed on the model, and the ν concepts, instead of being regarded as fixed effects, are regarded as random observations from a population, or sub-population, of concepts. In this case the model may be written.

$$(2) \qquad x_j = \alpha_j \left(\sum_{p=1}^{m} a_{jp} f_p + e_j \right) + \beta_j, \quad j = 1, \ldots, n,$$

where x_j now refers to the response on the jth scale made by a randomly drawn subject to a randomly drawn concept.

To estimate the parameters like a_{jp}, it has frequently been the practice to find the average rating over all N subjects for each scale on each of the ν concepts in a sample of concepts. In this way the variability of the ratings reflects only concept variability, and the replication across persons merely serves to stabilize the observations. This is Heise's (1969) Design 1 resulting in an $n \times n$ matrix of correlations between scales based on ν observations. The use of this design seems difficult to justify in terms of the model, and its success probably depends on the fact that individual differences between subjects using the same scales for the same concept are generally much less than differences between ratings on concepts judged by a single subject using the same scale. Use of statistical factor analysis models are even harder to justify in this design, since the sample of concepts may rarely be considered as random, but is chosen to span the semantic differential space (See Heise,1969).

By analogy with factor analysis, the most direct way of estimating the pattern coefficients, a_{jp}, would be to draw a sample of person-concept combinations from all possible combinations in the population of persons and universe of concepts. This would be a very inefficient procedure in practice, but it can be approximated by choosing a large number of concepts (say over 100) and allocating a small number of concepts (between 1 and 8) at random from the list to each subject. If 100 subjects were used, each rating a sample of 4 concepts with say 50 scales, there would result a 50×50 matrix of scale intercorrelations, based on 400 observations. This design enables estimation of parameters for given populations of persons and concepts. If the samples from either population were chosen carefully, it would seem reasonable to use these estimates of parameters, without further factor analyses, in any further work using subsets of either the person or concept population. We will call this Design 3.

The factor variables suggested by the factor analysis are hypothetical constructs and the factor score, f_{pki}, for any particular person and any particular concept cannot be precisely determined. It is, however, possible to construct operational variables, linear weighted composites of the observed ratings, whose statistical distributions approximate, in various ways, the distributions of the factor scores. One method of doing this is to utilize the various formulae for "estimating" factor scores, using the estimates of the pattern coefficients as if they were known parameters. (See Harman, 1967, p. 345, Harris, 1967, and McDonald and Burr, 1967). These methods may only be used, of course, when all scales of the original standardization are used.

When the matrix of pattern coefficients contains only one substantial entry per row, the various scales form well defined, distinct clusters. This type of solution, most likely achieved by the use of an appropriate oblique transformation of the initial factor solution (Harman, 1967, p.273, Harris and Kaiser, 1964), indicates a direct link between the factor variables and the clusters of observed variables. Under these circumstances, a simple addition of ratings on scales in the same cluster may produce sets of scores which are highly correlated with appropriate factor variables. This is, in any case, the method advocated by Osgood et al (1957). The task of the factor analysis becomes one of clustering scales according to their degree of correlation, and the pattern matrix indicates which scales can reasonably be combined to provide measures of general semantic response tendencies.

How many scales should be combined to produce a variable which reflects the common factor of the scales depends on the size of the pattern coefficients. The best guide to the number needed would be a measure of the internal consistency of the resulting composite, as measured by coefficient alpha (Lord and Novick, 1968, p. 87).

An important question in the interpretation of the model is that concerned with concept-scale interaction. In terms of equation (1), it may be that two investigations using different sets of concepts but the same set of scales, give rise to very different factor constructs. Or it may be that the same factor constructs emerge but the loadings of some scales change their allegiances. Heise (1969) has given an excellent discussion of this problem and shows that such concept scale interactions may arise because of three artifactual reasons and two true conditions. An additional artifactual reason may be that, factor solutions being essentially indeterminate, the pattern matrices from the two studies have not been optimally matched (Evans, In Press).

In terms of the models above, concept-scale interaction may be characterized as within a given population of concepts or as between different populations of concepts. In the first case, use of Design 3 simply eliminates the possibility of noticing a concept-scale interaction, since only a single estimate of the pattern coefficients is made. This result gives rise to one difficulty. Adjective pairs like sharp-dull, which change meaning according to the concept rated, are a particular source of trouble if the population of concepts is very heterogeneous. Concept-scale interactions observable between different classes of concepts may be dealt with according to Heise's suggestions.

Response Tendencies

Variations in the extent to which subjects are prepared to utilize the full range of the scale in responding are well known (Peabody, 1962, Arthur, 1966). Heise (1969), in his review, and Kahneman (1963) describe a mechanism by which this kind of response bias, different within-subject variances of ratings, can affect the structure of the correlations between scales, particularly when only one concept is involved. Where many concepts are involved, it is not clear what will be its effect. It is convenient to designate this variable as response dispersion and to define it formally for a given subject, i, as d_i, where

$$(3) \qquad d_i^2 = \frac{1}{SC} \sum_{j=1}^{S} \sum_{k=1}^{C} \left(x_{kji} - \bar{x}_{..i}\right)^2$$

where x_{kji} is the response of the ith subject to the kth concept using the jth scale, and $\bar{x}_{..i}$ is the mean for that subject over all S scales and C concepts.

It would be desirable to assess response dispersion independently of the data used to estimate the pattern coefficients of a particular set of scales. For this reason the set of S scales and C concepts referred to in formula (3) should be selected for just this purpose.

Another type of systematic bias on which there may be individual differences is the response centering, the mean response for a given subject referred to in formula (3), given by

$$\bar{x}_{..i} = \frac{1}{SC} \sum_{j=1}^{S} \sum_{k=1}^{C} x_{kji}.$$

Messick (1957) demonstrated a median tendency for subjects to displace the centre point of the scale.

<u>The Standardization Study</u>

The aims of the study may be described in terms of the above model as follows:

(i) To recheck the feasibility of using semantic differential ratings with secondary school students.

(ii) To study the reliability of semantic differential judgments for the above population of students.

(iii) To provide standardized estimates of the pattern coefficients of the model by using a large sample of students and a large sample of concepts for a particular school population. The population of concepts used was English nouns, and about 1,000 of the most frequently used nouns were included in the list.

(iv) To study the effects of individual differences in methods of responding on the factorial structure.

(v) To study the nature of the semantic dimensions for the above person-concept population, and to suggest operational measures.

(vi) To study the stability of the factor pattern across different sub-groups of the sample; boys, girls; grade 8, grade 10, grade 12.

(vii) To study the stability of the factor pattern across a random subdivision of concepts and persons.

(viii) To study the stability of the factor pattern across <u>a-priori</u> divisions of the concepts into different categories.

METHOD

Ten high schools in Ontario and their companion junior high schools were selected from electoral districts chosen at random to be representative of metropolitan, urban, and rural districts, providing a sample of ten sets of grade 8, grade 10, and grade 12 students, a total of 843 students. A total of 1,080 nouns of highest frequency were chosen from the Thorndike and Lorge Teachers Word Book (1944) and split at random into two groups. From each list, sets of 4 random words were generated without replacement, the lists re-randomized, and the process repeated. The sixty scales were chosen to represent a larger list of 340 scales compiled from Roget's Thesaurus and other sources. The sixty scales were listed in random order and the polarity was randomly alloted. For further details of the method of choosing the scales see Evans (1970). The actual adjective pairs are listed in Table 3.

To measure response dispersion, response centering, and scale reliability, two sets of six concepts were chosen from Heise's (1965) list in such a way as to have large variance and zero mean on their semantic coordinates for evaluation (E), activity (A), and potency (P), according to Heise's estimates. In the same way, two sets of adjectives were chosen, each set containing two adjectives as measures of each EPA dimension. These 72 combinations were presented, with random polarity, along with the main task, and 24 of these combinations were repeated in scrambled order after the main task. Details of the response bias measures are presented in Table 2.

The main task for each subject was to make 240 ratings, 60 for each of 4 randomly chosen concepts. The complete set of 336 combinations were presented in booklet form. Because of the way in which the lists of 4 concepts were generated, it was possible to designate subjects as having concepts from one random group or the other. Even-numbered subjects were given concepts from one list, odd-numbered subjects, concepts from the other.

The booklets used consisted of 12 "digitek" sheets, specially prepared for this study and later machine processed. All except the second page contained 30 adjective pairs; the second page contained 6. The first 36 items comprised Set 1 of the response bias combinations. Next followed four sets of sixty scales which referred to four different concepts, one concept for two pages. In this case the concept was typed at the top of the page in the following manner:-

GRANDFATHER (He is their <u>Grandfather</u>).

Heise's (1965) method of using an affectively neutral sentence to elaborate the meaning of the concept was adopted. Each subject received a different set of concepts which were numbered for later identification. Next followed Set 2 of the response bias measures and the 24 repeated items. The students worked in groups of either about 30 or about 60, under the supervision of a trained research assistant. Before beginning the task, each group was instructed in the use of the semantic differential, by means of separate practice sheets, and in the use of digitek responses. No time limit was made, but an hour usually proved more than sufficient.

All response sheets were machine scored, and the data stored on magnetic tape. Because of omissions and other errors, only 800 of the 843 sets of data were usable. The ratings were initially scored between 1 and 7, the pole corresponding to a "7" rating in each adjective pair is placed first in Tables 2 and 3. These poles were initially intended to represent a more positive evaluation, more activity, or more potency.

ANALYSIS OF RESULTS

The 24 repeated concept-scale combinations provide data on short term test-retest reliability, the correlations being shown in Table 1. These range in value between .37 and .73. For combinations of, say, six scales, use of the Spearman Brown formula indicates the corresponding reliabilities would range from .80 to .94, and are thus very promising. Other reliability data of the same kind have been reported by Evans (1970).

Insert Table 1 about here

Two measures of response bias were calculated. For each subject, the response centering and response dispersion for each set of 36 bias scales were found. The correlation between the two measures of response centering was .47, and that between the two measures of response dispersion was .59. Other statistics are provided in Table 2. These bias measures thus appear to represent fairly consistent individual differences in subject response tendencies.

Insert Table 2 about here

Before proceding to the main analysis, it is useful to study the response characteristics for the 60 scales being standardized. Each scale was used to rate four concepts by 800 subjects, giving rise to a distribution of 3,200 ratings on the scale. The mean and standard deviations of each of these 60 sets of ratings may be used to detect any abnormalities in the use of any particular scale. The means of the 60 ratings ranged from 3.2 for "humble-proud" and 3.8 for "simple-complex" to 5.2 for "good-bad" and "valuable-worthless" and 5.0 for "meaningful-meaningless". The grand mean was 4.54 and the standard deviation of the mean was 0.37. The average standard deviation for the 60 scales was 1.76, and standard deviation, 0.13, the range being from 1.5 to 2.1. It thus appears that, while there is a systematic bias towards the more pleasant, valuable, active, and potent pole of the scales, no set of ratings was very skewed or leptokurtic.

The main analyses were concerned with the factorial structure of the 60 scales. The first analysis utilized data from the 60 scales and all 800 subjects. The 3,200 observations were used to compute a .60 x 60 matrix of correlation coefficients between scales. This was factor analysed using Thomson's least squares iterative technique (See Harman, 1967, p.135). The number of factors was set at 10, the number of principal components extracted in the first iteration with latent roots greater than unity. The factor matrix was transformed by varimax (Kaiser, 1958), Promax (Hendrickson and White, 1964), and Schmid and Leiman (1957) programs, and each of the alternative solutions compared. The varimax solution was chosen as being most useful, since it provides the same information as the other two, and was in this instance more easily understood. It is presented in Table 3.

- - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - -

The adjective pairs in Table 3 have been rearranged from their original random order of presentation to bring out more clearly the nature of the factors, and lines have been drawn to group the scales as far as possible into clusters, although some overlapping is apparent. The first seven of the clusters correspond very closely with the main oblique factors of the promax analysis.

This table is central to the paper, since all other factor analytical results were compared with it, including those derived from the ratings after they had been adjusted for person response bias. Before we proceed to discuss the factors, the effect of response bias on the factor structure may be thus fairly quickly dismissed. The second factor analysis was performed in exactly the same way as that above, except that the data for each subject were adjusted by subtracting the constant $\bar{x}_{..i}$ from each rating and dividing the result by the response dispersion d_i. In this way the data for each person were adjusted for independently estimated response biases. The results of the factor analysis of the adjusted ratings were then compared with the coefficients in Table 3.

Before such a comparison is made an attempt should be made to maximize the similarity between the two matrices, so as to eliminate artifactual dissimilarities due to using different bases. A convenient method, in this case, is to regard the matrix based on the unadjusted ratings as a standard and to rotate the second matrix to maximum congruence with it. (See Cliff, 1966, Schoneman, 1966, Evans, <u>In Press</u>).

When this has been done the two matrices may be compared factor by factor. Two measures of agreement between the factor coefficients may be used- Tucker's (1951) coefficient of congruence (τ), which has a maximum of 1.0, and the root mean square difference between corresponding coefficients (δ). Each of these indices is shown for each factor in Table 4. The first two rows

Insert Table 4 about here

of numbers refer to the match between the factors for the adjusted and unadjusted ratings. Congruence coefficients (τ) greater than 0.9 indicate good agreement, those better than 0.8 usually indicate sufficiently good agreement to accept the factor variables as referring to similar dimensions of attitude. Low values of the root mean square difference (δ), less than 0.05, say, indicate not only approximate proportionality between the two columns of factor coefficients, but, in addition, a high level of absolute agreement. Of the two indices, the congruence coefficient is to be preferred because of its relative standardization.

In terms of both criteria, there is evidently very close agreement between the factorial structures derived from the unadjusted and adjusted ratings. For this reason, in all other comparisons of structures, response bias corrections have not been made.

Comparisons of sub-samples of persons with the total sample were also made. The measures of factorial similarity with the total sample (matrix A_0) for boys (matrix A_2), girls (A_3), Grade 8 (A_4), Grade 10 (A_5), and Grade 12 (A_6) are shown in Table 4. In each case the extent of agreement is high, even taking into account the fact that the sub-sample is included in the total sample. This finding is in line with the many reported findings of invariance across different samples of persons in semantic differential dimensional structure. It does, in this case, give evidence of the generality of the results throughout the population of persons sampled.

Invariance of the semantic differential dimensions within different classes of concepts has not always been demonstrated -- rather, the reverse has been the case, giving rise to the notion of concept-scale interaction. To study this type of invariance or interaction, the list of 1,080 concepts was divided in two ways. First a random division of concepts was made, as explained in the procedure, and the two resulting lists used for a random division of subjects (odd and even). The similarity of the factors (A_7 and A_8) resulting from this division with those for the total sample are shown in Table 4. Again the agreement is high, as might be anticipated. The second division was based on an a priori classification of the concepts. Each of the 1,089 concepts was allocated, by three independent judges, to one of six concept classes. Disagreements between judges amounted initially to 51 nouns. The disagreements were discussed and a final consensus allocation made, except for two concepts. The classes of concepts used were as follows:

(1) Persons: the referent of the noun is a person, class of persons, or group of persons. For example, nurse, stranger, crowd. (436 observations).

(2) Concepts pertaining to the human body. For example, birth, shoulder, wound. (124 observations).

(3) Substances or things grown, mined, made, collected, hunted, by people in this culture -- food, clothing, buildings, transport, communications, works of art, etc. For example, jewel, shop, rug, highway, butter. (844 observations).

(4) Abstract concepts associated with human endeavour. For example, news, title, fun, mystery, campaign, scorn. (1133 observations).

(5) Naturally occurring things and animals. Nature at large. For example, plant, sky, lake, mouse. (315 observations).

(6) Relationships: The referent of the noun is an abstract relationship. For example, corner, form, connection. (340 observations).

Most disagreements between the judges were in categories 3 and 5. The correlation matrices for ratings of each concept type were compared with the correlation matrix for the total person-concept sample. For concept types 1, 3, 4, and 6 the agreement was very close, indicating a close similarity of factorial structure. For two types, however, there were large enough disagreements to warrant a separate factor analysis. These were concept types 2 and 5. Measures of similarity between the factor matrices corresponding to these two concept groups (A_9 and A_{10}), and the target matrix, A_0, are presented in Table 4.

The largest disagreements are with concept type 2, for which the only acceptable coefficients of congruence refer to factors I, V, and VI, factors of general affect, activity, and potency. A visual inspection of the coefficients also indicates some degree of similarity for factors II, III, IV, and VII. Concept type 5 shows a greater measure of agreement, the congruence coefficients reaching acceptable values for all but factor VII, Ease, and Factor X. The discrepant factors are also minor in terms of the amount of variance accounted for.

Thus, as far as concept classes are concerned, there is evidence of general similarity particularly within dimensions which are similar to Osgood's EPA factors, but it is clear that some classes of words lead to noticeable discrepancies. With more narrowly defined classes, the discrepancies could no doubt be greatly increased. The interaction between concept class and factorial structure appears, in this case, to be due more to lack of scale relevance than any other condition. Judgments of competence, success, ease, meekness, or value would probably be considered irrelevant by most of the subjects for parts of the body. The data are currently being analyzed more intensively to study the factorial shifts associated with various types of concept.

The Factor Variables

The coefficients in Table 3 may be used in the usual way to provide a tentative description of the semantic response dimensions which may, on the basis of these data, be hypothesized to underly the semantic differential scales. First, it should be noticed that many scales have quite low communalities--the average is .40--indicating a fair amount of specificity. If factor scores are estimated, or scales in other ways combined, it is essential to check the internal consistency of the composite. Second, in the varimax solution, there is a high degree of overlap among the scales which define the factors. For this reason if scales are simply added, it will be difficult or impossible to produce even approximately uncorrelated composites.

The first factor corresponds closely to what has been termed evaluation (Osgood et al, 1957), but the scales most closely related to the judgment of value ("right-wrong","good-bad", "valuable-worthless"), also load on the second factor along with "wise-foolish", and "meaningful-meaningless". In the oblique analysis these five scales are separated from the rest of the scales in factor I. So, too, are the scales which load on both Factor I and Factor III ("contented-discontented", "successful-unsuccessful", "sure-uncertain", "rich-poor"). For this reason, Factor I has been named "General Affect," suggesting that it represents a general affective response of liking vs. disliking. The variable is defined principally by the scales "pleasant-unpleasant", "friendly-hostile", "pleased-angry", "kind-cruel", "beautiful-ugly", "comforting-frightening", "gentle-violent", "delightful-dreadful", "happy-sad", "loving-hating", and "sweet-sour", all of which have coefficients greater than .60, and all of which are defining scales for the corresponding oblique factor. The scale "good-bad" has a coefficient of .61 but is not a defining scale for the oblique factor, and it also has a high loading on varimax Factor II. The hedonistic interpretation of Factor I differs from that made of similar factors by many other workers, but for the data presented here, there is a clear need to differentiate between a pleasure-pain response and evaluation. However, Factor I is moderately correlated with scales which apparently measure value and also with scales which are measures of activity. The name "General Affect" has therefore been used.

The judgment of value characterizes Factor II, while Factor III correlates highly with scales suggesting care and competence. It is noteworthy that "anxious", as opposed to "carefree", is perceived to be at the positive end of this factor. The first four factors appear to span a cluster of judgments which may be broadly described as evaluative in the sense used by Osgood et al. Heise (1969) reviews arguments and evidence that the use of ratings of individual subjects does tend to result in a splintering of the evaluative dimension. This has certainly occurred here. The split between the hedonic and value factors also agrees in part with the results of Komrita and Bass (1967).

The activity and potency factors reported in many semantic differential studies are presented here also in Factor V and Factor VI. Seven of the scales which define the Activity Factor appear to refer more to personal energy or excitement than to physical activity as such. There is also some overlap with the general affective factor with five of these seven scales. The scale "hot-cold" does not load on this factor. The potency factor differs from that frequently obtained in that the scales "rough-smooth" and "hard-soft" are not even moderately highly correlated with other scales which define the factor. Factor VII clearly discriminates the easy from the difficult. It is of interest that "familiar-strange" and "simple-complex" evoke a similar response to "easy-difficult," justifying, from the students' point of view, traditional maxims of teaching. Factor VIII appears to discriminate humility and unselfishness from their judgmental opposites, and is named "Meekness". The two remaining factors involve only three adjective pairs and are difficult or impossible to interpret.

Summary and Discussion

A central assumption in this study was that the meaning continuum registered by an adjective pair remains relatively constant over a large universe of concepts and over subjects within the relatively homogeneous population tested. Real shifts in meaning of adjectives of the kind discussed by Heise (1969) were regarded as rare. Variability in ratings was assumed to arise, not from changes in meaning of the adjectival pair, but from different responses to different concepts by the same person, and from different responses to the same concept by different persons. This assumption gives the adjectival scales a constancy of interpretation in research with the particular population of subjects, without which their use would be logically difficult. It also enables the model described at the beginning of of the paper to be used in the search for useful clusterings of adjectival meanings, which might reflect some underlying semantic generalities within the population studied.

An attempt was made to validate this main assumption by showing the invariance of the factor structure both across different types of subjects and across different classes of concepts. It was suggested that the small departures of two concept types from the general structure could be explained by the fact that some of the scales were irrelevant for these kinds of concepts. This would not invalidate the general use of the scales, although it might obviously result in a loss in efficiency in some instances.

The existence of systematic response tendencies independent of the meaning of the adjectives were demonstrated. Such tendencies are important for interpreting individual results, and a set of 72 items which fairly reliably measures response dispersion was presented. Response bias appears to have little effect on the factorial structure of semantic differential scales.

The general substantive results agreed generally with the Evaluation, Activity, and Potency dimensions of Osgood et al (1957) and others, but notable differences were found. In particular, it is argued that there is a clear separation between a hedonic response and a judgment of values. The data from these 60 scales suggest seven useful composite scales which may be formed from the following adjectival pairs, by simple addition.

General Affect:	pleasant-unpleasant, kind-cruel, friendly-hostile, delightful-dreadful, pleased-angry, beautiful-ugly, loving-hating, comforting-frightening.
Value:	good-bad, meaningful-meaningless, valuable-worthless, right-wrong.
Success:	sure-uncertain, successful-unsuccessful, contented-discontented, unruffled-embarrassed.

Competence:	alert-dreamy, educated-ignorant, careful-careless, grown-up-childish, wise-foolish.
Activity:	Excited-bored, moving-still, fast-slow, fresh-stale, changing-permanent, young-old, up-to-date - out-of-date.
Potency:	strong-weak, large-small, deep-shallow, brave-cowardly.
Ease:	familiar-strange, easy-difficult, simple-complex.

Such composites could fairly confidently be used for samples similar to that described above, to measure attitudes and differences in attitudes to various concepts, without the need to factor-analyze the scales for the new data.

REFERENCES

Arthur, A. Z. Clinical Use of the Semantic Differential. Journal of Clinical Psychology, 1965, 21, 337-338.

Cliff, N. Orthogonal Rotation to Congruence. Psychometrika, 1966, 31, 33-42

Evans, G. T. Use of the Semantic Differential Technique to Study Attitudes During Classroom Lessons. Interchange, 1970, 1, 4.

Evans, G. T. Transformation of Factor Matrices to Achieve Congruence. British Journal of Mathematical and Statistical Psychology. In Press.

Harman, H. H. Modern Factor Analysis (Second Edition). Chicago: University of Chicago Press, 1967.

Harris, C. W. On Factors and Factor Scores. Psychometrika, 1967, 32, 363-380.

Harris, C. W., and Kaiser, H. F. Oblique Factor Analytic Solutions by Orthogonal Transformations. Psychometrika, 1964, 29, 347-362.

Heise, D. R. Semantic Differential Profiles for 1,000 Most Frequent English Words. Psychological Monographs: General and Applied. 1965, 79 (8, Whole No. 601), 31.

Heise, D. R. Some Methodological Issues in Semantic Differential Research. Psychological Bulletin, 1969, 72, 406-422.

Hendrickson, A. F., and White, P. O. Promax: A Quick Method for Rotation to Oblique Simple Structure. British Journal of Statistical Psychology, 1964, 17, 65-70.

Kahneman, D. The Semantic Differential and the Structure of Inferences Among Attributes. American Journal of Psychology, 1963, 76, 442-452.

Kaiser, H. F. The Varimax Criterion for Analytical Rotation in Factor Analysis. Psychometrika, 1958, 23, 187-200.

Komorita, S. S., and Bass, A. R. Attitude Differentiation and Evaluative Scales of the Semantic Differential. Journal of Personality and Social Psychology, 1967, 6, 241-244.

Lord, F. M., and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

Messick, S. S. Metric Properties of the Semantic Differential. Educational and Psychological Measurement, 1957, 17, 200-206.

McDonald, R. P., and Burr, E. J. A Comparison of Four Methods of Constructing Factor Scores. Psychometrika, 1967, 32, 381-402.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. The Measurement of Meaning. Urbana: University of Illinois Press, 1967.

Peabody, D. Two Components in Bi-polar Scales: Direction and Extremeness. Psychological Review, 1962, 69, 65-73.

Schmid, J., and Leiman, J. M. The Development of Hierarchical Factor Solutions. Psychometrika, 1957, 22, 53-61.

Schonemann, P. H. A Generalized Solution of the Orthogonal Procrustes Problem. Psychometrika, 1966, 31, 1 - 110.

Thorndike, E. L., and Lorge, I. The Teachers Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Tucker, L. R. A Method for Synthesis of Factor Analysis Studies. Personnel Research Section Report, No. 984. Washington, D. C.: Department of the Army, 1951.

TABLE 1

RELIABILITY OF SEMANTIC DIFFERENTIAL SCALES

CORRELATIONS BETWEEN REPEATED ADMINISTRATIONS

Concept	Scale					
	1 Good	2 Beautiful	3 Fast	4 Hot	5 Strong	6 Thick
1. Mother	.67	.73	.55	.55	.65	.61
2. Enemy	.53	.56	.43	.44	.59	.50
3. Laugh	.54	.54	.42	.49	.38	.47
4. Iron	.40	.59	.39	.52	.37	.46

TABLE 2

RESPONSE BIAS MEASURES, MEANS, STANDARD DEVIATIONS, AND CORRELATIONS

SET 1. (36 scale-concept combinations)

Scales: Good-bad, beautiful-ugly, fast-slow, hot-cold, strong-weak, thick-thin.

Concepts:		Mother	Enemy	Fire	Sleep	Laugh	Iron	Mean
Heise Profile Values	E	1.7	-3.3	-3.5	1.7	0.7	0.6	-0.3
	A	1.4	0.3	2.7	-3.2	1.9	-2.3	0.1
	P	-2.8	0.3	0.1	-0.7	-1.6	4.5	0.0
Polarization		3.6	3.4	4.4	3.7	2.6	5.1	

M_1, Response Centering Measures: Mean = 4.80, S.D. = 0.38

S_1, Response Dispersion Measures: Mean = 1.83, S.D. = 0.35

SET 2. (36 scale-concept combinations)

Scales: Nice-awful, fair-unfair, moving-still, sharp-dull, hard-soft, heavy-light.

Concepts:		Baby	Argument	Danger	Rest	Play	Rock	Mean
Heise Profile Values	E	1.4	-2.4	-2.8	1.9	0.3	-0.3	-0.3
	A	1.4	1.4	1.5	-2.6	1.9	-3.2	0.1
	P	-3.2	0.2	0.5	0.0	-0.4	3.9	0.2
Polarization		3.8	2.8	3.2	3.2	1.9	5.0	

M_2, Response Centering Measures: Mean = 4.51, S.D. = 0.35

S_2, Response Dispersion Measures: Mean = 2.03, S.D. = 0.42

Correlations		M_1	M_2	S_1
	M_2	.47		
	S_1	.11	.24	
	S_2	.36	.32	.59

Estimated Reliability of 72 item Scale	
Centering:	.65
Dispersion:	.74

TABLE 3

VARIMAX FACTOR MATRIX FOR UNADJUSTED SEMANTIC DIFFERENTIAL SCALES[a,b]

(Based on 1,089 concepts and 800 persons)

Scale[c]	I General Affect	II Value	III Success	IV Compe-tence	V Acti-vity	VI Potency	VII Ease	VIII Meekness	IX -	X -	h^2
4.[d] sweet-sour	60										.41
5. comforting-frightening	63										.46
7. healthy-sick	51				23						.43
8. clean-dirty	57		22						26	-23	.52
9. pleasant-unpleasant	74	20									.68
10. honest-dishonest	50			22				22			.40
11. rough-smooth	49								-27		.38
13. friendly-hostile	74										.58
14. pleased-angry	69										.52
15. willing-unwilling	51										.36
16. soft-hard	54					-26					.47
17. kind-cruel	76										.63
18. beautiful-ugly	69										.58
20. light-dark	41										.23
26. fair-unfair	57							22			.50
28. safe-dangerous	55						20				.44
33. gentle-violent	61										.52
56. optimistic-pessimistic	38										.23
35. delightful-dreadful	71	26									.64
37. happy-sad	62	20			21						.56
48. loving-hating	68	21									.61
53. unselfish-selfish	41	21						34			.38
47. right-wrong	45	41									.51
54. good-bad	61	47									.65
29. valuable-worthless	39	44		20		23					.37
52. meaningful-meaningless	29	50			28						.47
43. wise-foolish	34	44	29	39							.60
49. contented-discontented	49	24	32								.47
40. successful-unsuccessful	32	36	33	20							.46
44. sure-uncertain	24	20	39	28							.37

Table 3.--Continued

Scale[c]	FACTOR I General Affect	II Value	III Success	IV Compe-tence	V Acti-vity	VI Potency	VII Ease	VIII Meekness	IX -	X -	h^2
50. rich-poor	27		30								.30
60. unruffled-embarrassed			31								.12
1. educated-ignorant	29			41							.40
23. careful-careless	31			51							.40
24. anxious-carefree				47							.26
39. grown-up - childish	22	28		35		27					.36
31. alert-dreamy				55	21						.37
3. interesting-boring	34	21			36	20					.34
41. colorful-colorless	36	21			34						.35
45. fresh-stale	40	26			41						.47
51. excited-bored	32	26			51						.48
30. young-old	22				39						.23
21. up-to-date - out-of-date		25			37				36		.42
46. varied-repetitive		23			34						.21
27 fast-slow					47						.29
32. moving-still				25	51						.36
38 changing-permanent					39						.18
42. noisy-quiet					33			25			.30
12. large-small						47					.25
19. strong-weak						57					.40
25. brave-cowardly	29			22		41					.43
34. deep-shallow						32					.20
36. masculine-feminine	-24					37			-20		.27

Table 3.—Continued

Scale[c]	I General Affect	II Value	III Success	IV Compe-tence	V Acti-vity	VI Potency	VII Ease	VIII Meekness	IX -	X -	h^2
55. easy-difficult	26						49				.36
57. familiar-strange							53				.33
58. simple-complex							45				.32
2. humble-proud				-21		-20		27			.25
22. rising-falling					24	26			28		.29
59. hot-cold					21					35	.21
6. conventional-unconventional											
% Variance	18	04	02	03	04	03	02	01	01	01	40

a. Decimal points have been omitted.
b. Coefficients less than .20 are not shown.
c. Adjective pairs were not necessarily presented with the polarity shown.
d. Scale numbers indicate the position of the scale on the response sheet.

TABLE 4

MEASURES OF SIMILARITY BETWEEN THE UNADJUSTED FACTOR MATRIX (A_0) FOR THE COMPLETE PERSON-CONCEPT SAMPLE AND OTHER FACTOR MATRICES

Factors of A_0 Compared with:		FACTOR I General Affect	II Value	III Success	IV Compe-tence	V Acti-vity	VI Potency	VII Ease	VIII Meekness	IX -	X -
A_1: Complete, Adjusted	τ	1.00	1.00	.00	1.00	1.00	.99	.99	.98	1.00	.98
	δ	.01	.02	.02	.01	.01	.02	.02	.02	.01	.02
A_2: Males	τ	1.00	.99	.97	.99	.99	.99	.99	.96	.97	.92
N^a = 380	δ	.02	.03	.03	.02	.03	.02	.03	.03	.03	.04
A_3: Females	τ	1.00	.99	.99	.99	.98	.99	.98	.98	.97	.96
N = 420	δ	.02	.02	.02	.02	.02	.02	.03	.02	.03	.03
A_4: Grade 8	τ	1.00	.97	.93	.97	.97	.98	.96	.91	.91	.93
N = 287	δ	.04	.04	.06	.04	.05	.03	.04	.05	.05	.04
A_5: Grade 10	τ	1.00	.98	.96	.99	.99	.98	.97	.98	.94	.85
N = 258	δ	.02	.04	.04	.03	.04	.04	.04	.03	.04	.06
A_6: Grade 12	τ	1.00	.99	.96	.98	.99	.98	.97	.95	.93	.85
N = 255	δ	.04	.03	.04	.04	.04	.04	.04	.04	.04	.06
A_7: Odd Subjects	τ	1.00	.99	.98	.99	.99	.98	.98	.97	.95	.94
N = 397	δ	.02	.02	.03	.03	.03	.03	.02	.03	.03	.08
A_8: Even Subjects	τ	1.00	.99	.98	.99	.99	.99	.98	.95	.96	.93
N = 403	δ	.02	.03	.03	.02	.03	.02	.03	.03	.03	.04
A_9: Concept Type 2	τ	.98	.76	.72	.78	.90	.84	.75	.58	.52	.61
W^b = 124	δ	.10	.14	.14	.12	.09	.11	.12	.13	.13	.11
A_{10}: Concept Type 5	τ	.98	.89	.83	.92	.95	.92	.74	.87	.74	.61
W = 315	δ	.09	.09	.09	.07	.07	.08	.11	.07	.09	.11

a. In each case the number of persons in the sample is shown. Since each person rated four concepts, the number of observations is four times this number.

b. In this case the total number of observations (W) is given.