THE EFFECTS OF SO-CALLED "IRRELEVANT" CONCEPTS ON THE RESPONSES MADE TO INTEREST TEST ITEMS

by

George Howard Walther

Captain, United States Air Force

A Thesis Submitted to the Faculty of the
DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

1964
STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Head of the Psychology Department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

[Signature]  [Date]

DOROTHY L. MARQUART
Assoc. Professor of Psychology
ACKNOWLEDGEMENTS

Among the many who have contributed to this thesis project through their ideas, suggestions, and technical advice, I am most especially grateful to Dr. Dorothy I. Marquart for her unselfish expense of numerous hours from her personal leisure time in the reading and re-reading of the manuscript, her patient and competent instruction and counsel, and her countless helpful suggestions.

Without the complete cooperation of the administration, faculty, and students of Sunnyside High School in Tucson, Arizona, this experiment could not have been conducted as it was. Mr. Jack J. Patton, School Psychologist at the high school, was particularly helpful in the scheduling and administering of the tests to the student subjects.

I am indebted to the University of Arizona's Numerical Analysis Laboratory for donation of the use of their electronic computers and peripheral punched card equipment which was used in the processing and analysis of the data. Special thanks go to Mr. Jack Gaines for his valuable and generous assistance in helping me "debug" the computer programs I wrote to handle the experimental data.

To the United States Air Force goes my deepest gratitude for providing me with this year of graduate study through the auspices of the Air Force Institute of Technology's Civilian Institution Division.

August 15, 1964
Tucson, Arizona

George H. Walther, 56 321A
Captain, USAF
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td>7</td>
</tr>
<tr>
<td>Meaning, Attitudes, and Measurement</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 3</td>
<td>10</td>
</tr>
<tr>
<td>The Semantic Differential</td>
<td>10</td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td>17</td>
</tr>
<tr>
<td>Objectives of This Study</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td>23</td>
</tr>
<tr>
<td>Methods of Study</td>
<td>23</td>
</tr>
<tr>
<td>Apparatus</td>
<td>23</td>
</tr>
<tr>
<td>The Interest Inventory</td>
<td>23</td>
</tr>
<tr>
<td>The Semantic Differential</td>
<td>25</td>
</tr>
<tr>
<td>Procedure</td>
<td>28</td>
</tr>
<tr>
<td>Directions to Subjects</td>
<td>31</td>
</tr>
<tr>
<td>Interest Test</td>
<td>31</td>
</tr>
<tr>
<td>The Semantic Differential</td>
<td>33</td>
</tr>
<tr>
<td>CHAPTER 6</td>
<td>38</td>
</tr>
<tr>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>Intercorrelations among Semantic Differential Scales</td>
<td>38</td>
</tr>
<tr>
<td>Relationship of Semantic Differential Scores and Interest Test Item Scores</td>
<td>38</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS--Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Effects of Order of Presentation and of Sex upon Semantic Differential Responses and Interest Test Scores</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER 7</td>
<td>47</td>
</tr>
<tr>
<td>Discussion and Conclusions</td>
<td>47</td>
</tr>
<tr>
<td>CHAPTER 8</td>
<td>57</td>
</tr>
<tr>
<td>Summary</td>
<td>57</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>61</td>
</tr>
<tr>
<td>Semantic Differential Profiles for the Experimental Concepts — Ungrouped Data</td>
<td>61</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>63</td>
</tr>
<tr>
<td>The Interest Test</td>
<td>64</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>68</td>
</tr>
<tr>
<td>Mean Scores and Standard Deviations of Experimental Concepts</td>
<td>68</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>71</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Frequency of Inconsistent Responses — Responses on Repeated Items Differing by More Than One Scale Step</td>
<td>37</td>
</tr>
<tr>
<td>2.</td>
<td>Intercorrelations among the Ten Scales</td>
<td>39</td>
</tr>
<tr>
<td>3.</td>
<td>Relationships between Semantic Differential Attitude “Scores” for the Concepts and Interest Test Scores for the Same Concepts — Ungrouped Data</td>
<td>41</td>
</tr>
<tr>
<td>4.</td>
<td>Relationships between Semantic Differential Attitude “Scores” for the Concepts and Interest Test Scores for the Same Concepts — By Sexes</td>
<td>42</td>
</tr>
<tr>
<td>5.</td>
<td>Relationships between Semantic Differential Attitude “Scores” for the Concepts and Interest Test Scores for the Same Concepts — For the Group as a Whole and For the Two Testing Sequences</td>
<td>44</td>
</tr>
<tr>
<td>6.</td>
<td>Significances of Differences between Scores Assigned by Sexes and by Groups (t)</td>
<td>46</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mark Sense Response Card and Semantic Differential Test Sheet</td>
<td>29</td>
</tr>
</tbody>
</table>
ABSTRACT

To determine the effect of so-called "irrelevant" concepts contained in many interest inventory items on the rating given that item by a test subject, an interest test, consisting of ten "irrelevant" concepts, each paired with an activity stem from ten broad occupational areas, was administered to 284 eleventh and twelfth grade high-school students. Semantic differential ratings based on eight evaluative scales indicated subjects' attitudes toward these concepts. Sixty-two data sets were excluded because of possible lack of validity.

Numerical values of semantic differential ratings of each concept were correlated with the ratings given the interest test items containing the concept as an "irrelevant" component. All correlations were positive and were significant at $P < .01$. They ranged between $+.577$ and $+.318$ when data from all eight descriptive scales on the semantic differential were included in analyses. Correlations differed only slightly when data from two scales, found to be only moderately related to the other six, were excluded.

Neither sex of subjects nor order of test administration affected the significance of these relationships.
The experimenter concludes that so-called "irrelevant" concepts in interest inventory items are actually relevant in that they influence responses made to the items.
CHAPTER I

Immediately after the Armistice of World War I the problem of how to determine the strength and direction of an individual's interests began receiving considerable attention. The apparent reason for this sudden impetus was the publication of several studies which showed that the traditional method of interest measurement, directly asking individuals in an interview what type of activities they enjoy or dislike, frequently yields responses which are unreliable, unrealistic, and practically useless (cf. Fryer, 1931). A standardized, verifiable method of measuring interests by some more subtle procedure was essential.

During the academic year 1919-1920 a graduate seminar on interests was conducted at the Carnegie Institute of Technology. Perhaps one of the most important conclusions reached by these scholars was that individuals engaged in different occupations share common interests which differentiate them from persons in other occupational fields. These interests are not necessarily referable to vocational activity, but include a vast array of objects, personality types, sports, and leisure time projects from everyday experiences about which individuals in similar vocations may be expected to
share similar attitudes (Anastasi, 1961). Development of one of the currently most popular interest inventories, the Strong Vocational Interest Blank, was a major outgrowth of the Carnegie seminar (Fryer, 1931). The Strong Vocational Interest Blank was one of the first attempts at criterion keying of items, each item being empirically keyed for different vocations.

In the next decade, the other most currently popular interest inventory, the Kuder Preference Record, came into being. This inventory is now titled the Kuder Preference Record — Vocational, to differentiate it from the Kuder Preference Record — Occupational and the Kuder Preference Record — Personal. The original Kuder Preference Record had greater face validity than did the Strong Vocational Interest Blank in that the items on the Kuder were not quite so seemingly unrelated to vocational activity. The preference-type item was added by the Kuder Preference Record to the rapidly growing repertory of interest-testing techniques.

Both the Strong VIB and the Kuder Preference Record — Vocational have received considerable research analysis while being widely used in counseling and vocational guidance. Many writers, chiefly in The Mental Measurement Yearbooks, feel that both Strong and Kuder deserve much credit for their efforts to improve the validity of their inventories. Nevertheless,
Rothney, et al. (1959) have observed that after thirty years of research only meager evidence of the inventories' validity is existent.

Fowler (1953, p. 742) is of the opinion that the Kuder Preference Record — Vocational scores obtained for "certain wide areas of interest are reasonably objective." (Italics added.) He warns, however, that the validity of these scores has not yet been completely demonstrated. "Personality measurement has not yet reached the point of producing completely objective and trustworthy scores..."

The Strong Vocational Interest Blank is generally described as having been fairly carefully constructed; it has been continually subjected to research and revision (Anastasi, 1961). However, the Kuder Preference Record — Vocational has been criticized by many authors (cf. Berdie, 1949) as having been constructed and revised according to much less rigorous methods. Rothney (1954, 1959) said that the items for some interest inventories were simply "christened" a measure of a certain type of interest because the inventory author thought that was what it might be. Kuder's Administrator's Manual (1956) for the Kuder Preference Record — Vocational explains that the items for that inventory were originally written and grouped on an a priori, or content validity basis. The inventory then underwent item analysis. Item
groups were developed from items showing high internal consistency and low correlations with other groups. However, Rothney (1959) has cautioned that such internal consistency indices only serve to show that the test authors have been "fairly consistent in the christening process."

Bordin (1953) also questioned the true worth of the validity data on the Kuder Preference Record — Vocational since much of its validation data was contributed by test customers. Bordin said that he doubted whether anyone would send to the test author any data which failed to support the author’s hypotheses about his instrument.

Berdie (1949) does not think that inadequate inventory construction methods are a significant factor if the inventory predicts successfully those occupations from which the subject will derive the most satisfaction. However, Rothney (1959) judges that the validity of even the more carefully constructed Strong Vocational Interest Blank is, to say the least, questionable. Even a cursory reading of The Mental Measurement Yearbooks, to name one source, will cast serious doubts as to how accurately the inventories are predicting those occupations from which the participants will derive the most satisfaction.

During the earlier years of the interest inventory, Kuder (1939) noticed that certain types of personality test
items are apparently susceptible to changes in form and context. He believed that this phenomenon was due to the dependence of the test subject upon a "vague background of experience in which the more recent situations encountered are generally most prominent" (p. 49). He conducted a study which he believed demonstrated that the preference-type item, where the subject chooses between two or more items, is less sensitive to form and contextual changes than are the individually rated statements. In the construction of his inventory Kuder initially utilized forced-choice pairs of items (paired preferences) and later changed to forced-choice triads of items. The present writer was unable to locate any subsequent studies on the comparative merits of the two types of items. It should be noted that the Strong Vocational Interest Blank, which has undergone the more intensive research, still utilizes the single item concept; and the Strong Blank’s reliability coefficients are no less impressive than are those of the Kuder Preference Record.

Why, after decades of research and revision, do interest inventories continue to lack the degree of validity which most critics require? Is it conceivable that the individual’s rating of an item is a function of the words in that item other than those which most obviously seem to relate to a broad interest area? Is it possible that certain so-called
"irrelevant" or "neutral" concepts in an item influence the responses made to the item because of some special connotative meaning the subject may attach to the "irrelevant" word or words? Or is the basic meaning of an item dependent only upon the "relevant" concept?

Although the objectives of this study are treated in far greater detail in Chapter 4, the reader will find the intervening material much more meaningful if he considers the questions posed below.

The item from the Kuder Preference Record — Vocational (1948) "Raise white mice for a scientist" may illustrate our point. A MOST LIKE response to this activity is keyed to the Scientific and Outdoor scales. How would a person who truly has a strong scientific interest be likely to rate the item if he happens to have a most unfavorable attitude toward mice? Or, how would someone with virtually no interest in science be likely to answer the item if he had a true interest in observing and handling animals, or for some other reason had a strongly favorable attitude toward mice? Before these and similar questions can be discussed further, an inquiry into the nature of meaning and attitudes is necessary.
CHAPTER 2

Meaning, Attitudes, and Measurement

A general psychological definition of meaning is that meaning is "the entire set of cognitions, feelings, and action tendencies evoked by a symbol" (Krech, 1962, p. 307). Interest inventory authors have tried to list some of the more common denotative meanings of words used in their questionnaires (cf. Kuder, 1948). However, to define all possible meanings of terms used in an inventory would be almost, if not entirely, unthinkable, even if the participants could be motivated to avail themselves of the listing of definitions to guide their cognitions.

Osgood (1957) defines meaning as simply one's emotive reaction to words. It was to measure an individual's emotive response to various linguistic signs that he developed the semantic differential technique. (A full discussion of this device follows in Chapter 3.)

An attitude is also a predisposition to respond, but is often differentiated from meaning in that an attitude incorporates chiefly positive or negative evaluations or emotional feelings. An attitude, then, is a type of approach—
avoidance mechanism which predisposes the individual to an evaluative-type response.

In the interest inventory the individual is asked to make an evaluative-type response to various items. The subjects are typically told to record their preference for each item by encircling one of the letters L I D, or for each set of activities to indicate which they would like most and which they would like least. It appears quite clear that the typical interest questionnaire measures the subject's attitude toward an activity. The present writer maintains that it measures the subject's attitude toward the whole activity, toward all concepts appearing in the item. In no interest inventory known to this writer have the "relevant" words or sets of words been italicized, underlined, or printed in capital or boldface letters. The subject is presented with the item in a completely unstructured manner. He must decide for himself when rating an item like "Train dogs to lead blind people" whether he is being asked primarily to express his feelings about training, dogs, or blind people. The interest inventory authors seem to have forgotten that "there are a great many different components in the meaning evoked by a word..." (Krech, 1962, p. 279). An individual's perception of his world is frequently altered by the introduction of carefully or carelessly chosen words into his momentary
environment. "If two persons have had different experiences with words, or if they perceive the communication context differently, the meaning of an utterance will differ for them" (Krech, 1962, p. 291).

When answering an interest inventory item the participant's momentary environment is the activity in which he is asked to imaginatively place himself!
CHAPTER 3

The Semantic Differential Technique

The semantic differential technique was developed by Charles E. Osgood (1952) as an instrument for measuring and quantifying the connotative meaning of concepts. Although it is a relatively recent method, early analyses show it to be highly reliable (Anastasi, 1961; Krech, 1962). Following a thorough statistical study of the technique, Messick (1957, p. 206) concluded that "the scaling properties implied by the semantic differential procedures have some basis other than mere assumption." Despite its inception as a meaning-measuring device, considerable research has indicated the differential's valuable applicability to psychotherapy and attitude scaling (Osgood and Luria, 1954; Lazowick, 1955; Anastasi, 1961). Of chief interest to this present study is the fact that the technique's validity as an attitude test appears quite satisfactory. Krech (1962) reported that substantial correlations have been found between semantic differential ratings and scores on Thurstone and Guttman scales. He further reported that semantic differential ratings on ethnic objects differentiated between high-scoring and low-scoring individuals on the California F scale.
Since a major portion of this study is based on the semantic differential technique, the instrument will be discussed in detail. The main assumption behind the semantic differential is that the process of description or judgment can be conceived as the allocation of a concept to pairs of experimental continua defined by pairs of bipolar terms (Osgood, 1954). A limited number of such continua can be used to define a semantic space within which the meaning of any concept can be specified. Osgood further explained this rationale by saying that the feeling, "My father has always been a rather submissive person," can be at least partially represented by the subject's rating of "My father" in terms of two bipolar adjectives:

"MY FATHER active_________________________________X____________________________________passive
soft________________________________________X________________________________________hard"

Osgood maintains that the greater the strength of the association, the further toward the extremes will be the subject's rating.

Several sets of bipolar descriptive terms, called scales by Osgood, are used; each concept under investigation is paired with every scale. In an attempt to tease out the major factors which comprise the total semantic space of a concept, Osgood (1957) conducted highly complex factor analyses and intercorrelational studies of a wide variety of
scales. He concluded that three major factors account for nearly all of the semantic space: evaluative, with high loadings in such scales as good-bad, clean-dirty, and pleasant-unpleasant; potency, consisting of scales like large-small, heavy-light, and strong-weak; and activity, with loadings in scales like fast-slow, active-passive, and sharp-dull. Of these three major factors, the evaluative factor, which appears to measure the valence of attitudes, has been found by Osgood (1957) to account for nearly 70 per cent of the extractable variance. Two types of factor analysis have shown it to be first in both magnitude and order of appearance. The evaluative factor seems to pervade the realm of human judgment, which leads Osgood to conclude that "the attitudinal variable in human thinking, based as it is on the bedrock of rewards and punishment both achieved and anticipated, appears to be primary" (Osgood, 1957, p. 72). Osgood feels it is reasonable to index attitudes by using sets of scales which have high loadings on the evaluative factor and negligible loadings on other factors. He believes that the evaluative dimension of the total semantic space is and should be identified with attitudes. To measure attitudes he advocates designing a semantic differential which utilizes only evaluative scales. An attitude "score" is thus obtainable by merely summing the points a subject assigns to the various scales.
Osgood (1957) considers the semantic differential to be a "highly generalizable technique of measurement which must be adapted to the requirement of each research problem to which it is applied. There are no standard concepts and no standard scales" (p. 77). The following construction criteria are recommended by Osgood:

(1) In attitude measurement the factorial composition of the scales should be limited to those with high loadings on the evaluative dimension and minimal loadings on all other dimensions.

(2) The scales should be relevant to the concepts being judged, either literally or metaphorically. Metaphorical representations make the purpose of the experiment less obvious.

(3) Scales should be defined by familiar and common opposites — true psychological opposites. Osgood found it necessary to merely assume that such scales are actually linear between their polar opposites and that they do in reality pass through the "origin," or center of the continuum which exists between the bipolar terms. Ideally, the scales should be strictly linear with respect to evaluation so that only one of the terms, if used in isolation, tends to be "favorable" in meaning. Although the scales employed in attitude measurement should be identified with the evaluative
factor, Osgood believes that scales of unknown factorial composition may be used in the semantic differential if such scales are considered highly relevant to a particular problem.

(4) It has been found that subjects most easily identify a 7-point graphic scale with the linguistic quantifiers "extremely," "quite," and "slightly," extending in both directions from a neutral "origin" where no attitude is assumed to exist. Osgood reports studies which indicate that each alternative on a 7-point graphic scale tends to be used with equal frequencies by subjects. This does not hold true when more or fewer categories are used, according to his findings.

(5) Although Osgood and his associates selected their bipolar terms from the adjective class, they believe that scales composed of polar nouns or verbs will yield the same dimensionality. The noun form of the concepts has been most frequently used because of the structure of the English language. Adjectives have also been satisfactorily used as concepts.

(6) Osgood reports having tried two types of scale differentials. In the first, items are presented in random order by concepts and scales to minimize the possibility of fostering a "halo" tendency:

LADY rough____:____:____:____:____:____:____smooth
ME fair____:____:____:____:____:____:____unfair, etc.
He noted that this constant changing of the concept actually changes the meaning of the concept from time to time. The chief advantage of such a format is that the subject is kept shifting from concept to concept and thus cannot readily compare his judgments on one scale with those on another.

The second type differential tried by Osgood presented all scales successively after the initial appearance of a concept. All judgments pertaining to one concept are elicited before the next concept is presented:

LADY

rough___:___:___:___:___:___:___smooth
fair___:___:___:___:___:___:___unfair
active___:___:___:___:___:___:___passive, etc.

A study conducted by Osgood to determine whether any differences in results should be expected from the use of one or the other of these two graphic forms failed to show any differences.

(7) Although scales constructed to measure the entire semantic space, as opposed to those measuring only the evaluative dimension, most frequently have the range minus three through zero to plus three assigned to the categories (the continuum increments are never numbered until after the technique is administered), Osgood assigned the scores one through seven to the points on the evaluative continuum in attitudinal measurements. (It should be noted that, mathematically, the
-3 to +3 method is identically comparable to the +1 to +7 convention. For scoring consistency the unfavorable poles (e.g., bad, sad, dirty, etc.) are scored +1 and the favorable poles (e.g., good, happy, clean, etc.) are scored +7.

(8) Osgood alternated the directions of the polarity among the items, randomly placing the antonyms at different ends of the continua. This, also, was an attempt to minimize the possibility of a "halo" or position tendency effect.

A later section is devoted to the specific ways in which the semantic differential technique employed in this experiment differs from Osgood's suggested specifications.
CHAPTER 4

Objectives of This Study

Despite the fact that many objections have been raised to the widespread use of interest questionnaires and inventories in various counseling situations, this study has been confined to one possible "mechanical" problem inherent in interest inventory design.

The majority of interest inventory items contains at least two concepts, only one of which is considered to be relevant to the interest being measured. The "relevant" concept is the one which most obviously relates the item to the interest area. For example, fixing some broken mechanical item is correlated with a general mechanical interest, serving as a bookkeeper relates to computational interests, etc. Other concepts included in the same item are considered to be irrelevant. An "irrelevant" concept, according to our definition, is any concept which most test authors most generally consider to be neither directly nor indirectly related to the major interest areas being measured. When the examples just cited are modified slightly, they more closely resemble items one would expect to find in an interest inventory, e.g., "Fix a broken bicycle," or "Serve as a bookkeeper of a loan
company." The italicized concepts are probably added by the inventory author to add variety to the items and realism to the activities. For whatever reason they are included, they interpose a new concept. In some forced-choice triad inventory designs, such as the one employed by the Kuder Preference Record - Vocational, rarely the "irrelevant" concept is held constant, and the subject may be asked to choose between selling sewing machines, fixing broken sewing machines, and convincing financiers to back development of sewing machines. In such exceptional cases the writer concedes that the effect of the "irrelevant" concept is minimized. However, where the individual is forced into choosing between teaching classes in English to applicants for citizenship, selling stocks and bonds, or being the chief cook in a fine restaurant, he necessarily has to consider teaching, selling, and cooking in particular situational settings, and in connection with entirely new concepts such as "stocks," "bonds," "applicants for citizenship," "English," "fine restaurant," and "chief" cook.

The same problem is believed to exist both in preference type item presentations and where single items are considered and judged independently by the test subject. An example from the latter class will serve to illustrate the problem which this study investigates:
"Draw the plans for a dance studio." There probably would be general agreement on the basis of this item's content validity that it measures a type of mechanical interest because of the architectural inference. Kuder includes architecture in the mechanical occupational grouping. For the sake of argument, speculate as to the cognitive meaning the word "dance" may have for the subject completing this hypothetical inventory. We may imagine that most people would immediately think of some very socially acceptable form of dancing. In such a case, no strong emotive response is elicited and the participant proceeds to weigh his feelings about drawing plans.

What if the subject has strong attitudinal feelings about "dance"? Drawing the plans for a dance studio would certainly necessitate the architect's close association with professional dance instructors who would advise him of the technical requirements for such a studio. Eventually his plans would have to be completely acceptable to the dance instructors. The architect would have to visit the dancers' present establishment, and the dancers may be coming to his office. He may be taking his customers out to lunch to discuss the plans.

Assuming that the subject's attitude toward "dance" runs in the strongly positive (favorable) direction, the
individual may welcome the opportunity to satisfy imaginatively his own individual needs and wants. If he thinks of dance instructors as very pleasant, interesting people with whom he would very much enjoy close association, the writer suggests that this attitude may be strong enough to influence the test subject to answer the inventory item in a positive direction, even though he may not have favorable feelings about becoming an architect or drawing plans for anything.

The speculation may be carried one step farther, into the realm of the self-concept. The literature is filled with studies on the fakability of interest inventories. As Rothney and Schmidt (1954) point out, forecasting efficiency of interest inventories is severely limited by the fact that the answers can be faked. Faking need not be caused by intentional malingering (faking for the sake of faking). Invalid responses may be made by the individual in an attempt to justify his acceptance or rejection of a stated choice (Rothney, 1954). Generalizing from findings about personality inventories reported by Krech (1962), it is noted that the clusters which emerge from an inventory may merely reflect the individual's need to be self-consistent, answering different items in a way which fits his "implicit personality theory." If an individual sees himself as a crusader against the "evils of dancing," for him to consider the above inventory
item favorably would be internally inconsistent. Similarly, if his self-concept includes the idea of his being very fond of children, it is highly doubtful that he would "vote against" such items as "Take children on a fishing trip" or "Fix a child's broken toy," regardless of his sentiments about angling or toy repairing.

In a counseling case, the writer discussed his client's interest inventory responses with him at great length. Examination of individual responses to the items revealed that the client had marked that he liked all items which contained any reference to art, drama, music, dancing, theater, ceramics, crafts, sewing, cooking, and "different" or unusual things. Needless to say, the profiles were most atypical! During the interview the young male high-school student stated that he aspired to become a great actor, that he believed great actors to have a profound interest in all the arts, a mild interest in sewing and cooking "since most great actors are bachelors," and a fondness for nonconformity of all descriptions. In many instances his selections had actually been made on the basis of so-called "irrelevant" concepts — the wrong ones. This case is not presented to suggest that such a radical approach is an everyday occurrence, but to advocate that the "irrelevant" concept in an item is no less immune to the effects of the "implicit personality theory" or self-concept
than is the "relevant" concept.

For nearly two decades psychometrists and educators have been concerned about the semantic problem in interest inventories (cf. Roeber, 1948; Rothney, Danielson, and Heimann, 1959). They have noticed that common, everyday words on an interest inventory call out a myriad of impressions. Cognitions of the word "actress" might range from burlesque queens to moving picture stars, television bit performers, or Broadway musical stars.

In the present reported study the question was asked whether words like "actress," when appearing as the so-called "irrelevant" concept in an item, mean something pleasant or unpleasant, favorable or unfavorable, good or bad, or interesting or uninteresting to the inventory-taker, and whether the words evoke a response strong enough to either positively or negatively influence the answer to an item.

At the conclusion of the study appear figures which graphically illustrate the meaning a sample of high school students attached to several "irrelevant" words appearing frequently in currently used interest inventories (see Appendix A, pp. 61f).
CHAPTER 5

Methods of Study

Apparatus

The present study consisted of the administration of a 107-item interest inventory (see Appendix B, p. 63) and a semantic differential to 284 eleventh and twelfth grade high school students. The two instruments were designed to test the hypothesis that a strong attitude about the so-called "irrelevant" concept in an interest item would influence the judgment made about that entire item. The semantic differential measured the subjects' attitudes toward those concepts.

The Interest Inventory

The interest inventory was made up of ninety-eight items, ten to measure each of eight concepts, and nine to measure each of two concepts. This inequality in the number of items to measure the various concepts was unintentional. Two items, one containing selling and the other containing vegetables as the "irrelevant" concepts, were inadvertently not printed in the test booklets. To determine which subjects understood and followed the directions carefully and
marked their responses sincerely and accurately, an "inconsistency" scale was constructed. Eight items appeared twice in the inventory to measure consistency of response.

Each item was worded as a description of an activity and contained one "relevant" and one "irrelevant" concept. Ten "irrelevant" concepts were arbitrarily chosen from currently popular interest inventories. These ten were considered to be representative of the various types of concepts generally found in interest test items, and included selling, dance, drama, history, vegetable, music, teaching, children, art, and dogs.

Ten "relevant" activity stems were written for each of the ten broad occupational areas reported by the Kuder Preference Record - Vocational, viz., Outdoor, Mechanical, Computational, Scientific, Persuasive, Artistic, Literary, Musical, Social Service, and Clerical.

Each of the ten "irrelevant" concepts was paired with an activity stem from each occupational area, resulting in a ten-by-ten design of "relevant" and "irrelevant" concepts (with the exception of the two erroneously omitted items).

A five-point rating scale was used, the verbal quantifiers "Very Much" and "Somewhat" appearing with ENJOY and DISLIKE on either side of an "INDIFFERENT" origin. The
negative extreme was assigned the scale value "1" while the positive extreme was scored "5." Items were presented in a random order, the same order for all subjects.

The Semantic Differential

Subjects were asked to rate each of twenty concepts — the ten experimental words listed in the preceding section and ten similar control words — in terms of eight sets of bipolar descriptive scales. As Osgood (1957) suggested for attitudinal indexing, the scales were chosen from the evaluative dimension of meaning. The eight scales used were those isolated through factorization by Osgood and his associates as being evaluative, or were thought to be close in meaning to those used by Osgood. The scales of unknown factorial composition were selected because of their relevance to the problem under investigation. Scales used, in their order of appearance to subjects, were unpleasant-pleasant, foggy-clear, dangerous-safe, dirty-clean, uninteresting-interesting, bad-good, sad-happy, and dislike-like. Post-test intercorrelations among all the scales (see Table 2, p. 39) indicate that six of the eight scales used in this study are closely interrelated, and that they are apparently measuring the evaluative dimension. The other two scales, dangerous-safe and dirty-clean, are highly correlated with each other but are only slightly related to the other scales. Therefore,
results are reported in terms of all eight descriptive scales, and in terms of scales remaining after the exclusion of dangerous-safe and dirty-clean.

Instead of using Osgood's 7-step scale, extending in the positive and negative directions from a neutral "origin" representing "NO ATTITUDE," a 10-step scale, pictorially represented by the numbers zero through nine, was employed. The chief purpose in using ten steps was to provide an even number of categories which would necessarily preclude provision of a neutral position. The subjects were high school students, and no incentive beyond a promised report of results was offered for their participation. It was assumed and later confirmed that none of them had ever been asked to use a semantic differential before and that they would likely find the logic behind it quite unfamiliar. Thus, if given the choice, they might report "INDIFFERENT" or "NO ATTITUDE" without attempting to make judgments. A secondary reason for the use of a 10-step scale was the fact that the response medium, IBM Mark Sense cards, were available only with the numbers zero through nine preprinted over the response blanks. To decrease this number would have required obliteration of the blanks not being used. It was hoped that the subjects would perceive of the range between each of the bipolar terms as a continuum, and the experimenter believed that eliminating part of the
available blanks on the cards might make perception of a continuum more difficult.

For purposes of scoring consistency the method advocated by Osgood (1957), of uniformly assigning the numerically lowest score, zero, to the unfavorable (negative) poles and the numerically highest score, nine, to the favorable (positive) poles, was utilized. A subject's attitude "score" for a concept was obtained by summing the scores he selected on each of the eight scales for a given concept. Also reported are values based on his attitude "scores" obtained with the scales dangerous-safe and dirty-clean excluded.

Although Osgood (1957) believes that the direction of the polarity of each comparison should be randomly alternated to minimize the possibility of the operation of a "halo" or position tendency, the arrangement of scales and concepts used in this present study was not compatible with such reversal. The subjects would have experienced difficulty marking cards on which such shifting of the importance of the numerically defined blanks was employed. Also, randomizing the direction of the polarity of each item would have grossly complicated analysis of the data by an electronic computer.

The form of differential used was a graphic, or scale differential which in one sense was the exact reversal of the second type form Osgood used (see Osgood, 1957, and Chapter 3.
of this thesis). Osgood asked subjects to rate one concept in terms of each of the descriptive scales before considering another concept. In the present study subjects were asked to rate all twenty concepts in terms of one of the descriptive scales before judging the concepts in terms of the next scale. (Figure I shows a Mark Sense card positioned in the card outline on a test sheet.) This form of presentation was used in preference to either of those Osgood recommends because this experimenter believes that subjects having any difficulty perceiving the metaphorical relevance of a particular scale would find it easier to consider all concepts in terms of that scale before having to perceive the relevance of a new scale.

Procedure

Students were given both tests during a single fifty-minute class period. No prior announcement of the scheduled experiment was made. Six groups of subjects completed the semantic differential first, the other half took the interest test first. The number of students in the former category is somewhat smaller because, unknown in advance to the writer, school officials had scheduled considerably fewer students in the sixth group than in any of the other groups.
1. This is not a test of intelligence, personality, or aptitude. It is a survey to help us determine how high school students feel about certain everyday terms.

2. You are asked to "rate" each of 20 words in terms of 8 sets of descriptive adjectives which are printed on cards. 0 and 9 are the extremes of each set of opposite adjectives. Numbers in between represent varying degrees between the extremes. There is no neutral number! A 4 is the weakest rating of the upper adjective, and a 5 is the weakest rating of the lower adjective. Therefore, in cases where you feel neither adjective properly describes a word, please show what your choice would be if you had to choose, and decide in favor of one or the other adjective.

For example, consider the word PICKLE. If you were asked to rate PICKLE in terms of the adjectives SWEET-SOUR, where would you say it belonged? If you thought of it as SWEET and not SOUR at all, you would rate it 9. SOUR and not SWEET at all would be rated 0. Let's assume that you consider PICKLE both SWEET and SOUR. In this case, you would need to decide whether you consider PICKLE to be more often SWEET, or more often SOUR. Then you would probably select some number near the center of the range.

3. Blacken the space of the number which you feel best describes each word. MARK HARD and COMPLETELY FILL THE SPACE OF YOUR CHOICE. Erase incorrect markings completely.

4. Please do not spend a lot of time on one group. Put down your first reaction and go on to the next.

---

**Fig. I. Mark Sense Response Card and Semantic Differential Test Sheet.**
Students were informed only that they were being given a two-part interest test.

Subjects were under pressure to work rapidly. Although they were not specifically told that they had to finish by the end of the school period, they were not told otherwise. Further, several times during each administration of the interest test the experimenter noted the progress of the apparent majority of the group and warned, "If you are working at the right speed, you should be answering page ___. If you are not yet on page ___, you will need to work faster to finish in time." During each session there was a small number of subjects who did not finish during the fifty-minute period. The school principal allowed them to complete the tests without being penalized for tardiness to the next class.

The fact that the students were speeded through the tests is considered by the writer to have been advantageous. Though neither author suggests imposing time limits on subjects, both Kuder (1948) and Osgood (1957), in administering interest tests and semantic differentials respectively, encouraged their subjects to work fairly rapidly and not to deliberate over individual items, but to record their first impressions or reactions.
Since the two parts of the experiment required very
dissimilar directions, the second part was never begun until
all subjects had completed the first part.

Directions to Subjects

**Interest Test**

1. This is an interest inventory designed to tabu­
late your preferences for certain types of activities. It is
NOT a test. Therefore, there are no right or wrong answers.
An answer is "right" only if it is a true expression of your
likes or dislikes.

2. Frequently, some students answer an item a cer­
tain way because they would like to score either high or low
in a particular field, and they feel this particular activity
will affect their rating in that interest field. However,
research has shown that most often, students guess wrong;
that is, some of the areas they thought they wanted to be
high in turned out to be quite low, primarily because they
didn’t rate each activity completely honestly, but instead
rated the activity in terms of the field of interest they
thought the activity was related to.

3. Your response cards will be scored by...an elec­
tronic computer.... However, even this most modern equipment
can record your choices accurately only if you follow
instructions very carefully. It is extremely important that you make VERY HEAVY marks.... Second, you will have to keep the numbers along the left-hand edge of each card lined up with the numbers printed at the right-hand edge of each inventory sheet. And third, you need to make sure that the "CARD NO." and "SHEET NO." agree in all cases.

4. ...the results of your interest inventory...are for your use. They will not go on your school record. Therefore, it is especially to your advantage to do your best in being as truthful with yourself as you can.

5. Some activities listed would require formal training or higher education. In these cases, assume that you HAVE the necessary training.

6. Even though some activities may seem foolish or unimportant, please rate EACH and EVERY activity. If any activity is not scored by you, your entire inventory will be invalid and will not be scored or processed by the computer.

7. Please do not spend a lot of time on one item. Put down your first reaction and go on. Again, research has shown that your first answer on this type of questionnaire is your best answer in most cases. However, if you feel you have made a mistake in indicating your true choice, you may erase your mark completely, then mark the response of your choice.
The Semantic Differential

1. This is not a test of intelligence, personality, or aptitude. It is a survey to help us determine how high school students feel about certain everyday terms.

2. You are asked to "rate" each of twenty words in terms of eight sets of descriptive adjectives which are printed on cards. Zero and nine are the extremes of each set of opposite adjectives. Numbers in between represent varying degrees between the extremes. There is no neutral number! A 4 is the weakest rating for the upper adjective, and a 5 is the weakest rating of the lower adjective. Therefore, in cases where you feel that neither adjective properly describes a word, please show what your choice would be if you had to choose, and decide in favor of one or the other adjective.

   For example, consider the word PICKLE. If you were asked to rate PICKLE in terms of the adjectives SWEET-SOUR, where would you say it belonged? If you thought of it as SWEET and not SOUR at all, you would rate it 9. SOUR and not SWEET at all would be rated 0. Let's assume that you consider PICKLE to be more often SWEET, or more often SOUR. Then you would probably select some number near the center of the range.

3. Blacken the space of the number which you feel best describes each word....
4. Please do not spend a lot of time on one group. Put down your first reaction and go on to the next.

Scoring

Subjects' marks on the cards were electro-mechanically converted into numerical Hollerith codes. Several response sets rejected by the equipment because of faint markings were re-marked and reprocessed satisfactorily. Other sets rejected because of stray pencil markings in the scored fields of one or more cards were manually keypunched. Name, and sex-group codes, which had been written by subjects on a first, information card, were manually keypunched into the first card; codes were duplicated into all cards of a set. Each card had been prepunched with a subject and card number before the experiment. Twenty-one subjects either did not indicate their sex and group or failed to return their information cards. Their data have been included in analyses made without regard to sex or group, but were necessarily excluded from sex and group analyses.

All other processing was done by IBM 7072 and IBM 1401 electronic computers at the University of Arizona. Before any response set was computer tabulated, every used card column was inspected for blanks and for more than one punch (mark) in a column. Any discrepancy resulted in
exclusion of the entire set of data. Absence of any card other than the first (information) card also resulted in rejection of that subject's responses. Every interest item was scored in terms of the "irrelevant" concept it contained, and on the basis of the occupational groupings to which the "relevant" concept pertained. The semantic differential was graded in terms of each concept, and totals were accumulated for each descriptive scale. For both tests, a "score" was obtained by mathematical summation of the values assigned by the subject. Repetitious items in the interest test were scored only the first time they appeared; the second time they appeared, they were merely compared with the response made to them initially. One item was repeated three times instead of twice. Its third appearance was completely ignored by the computer.

An "inconsistency" score for each data set was found by comparison of the repetitious items on the interest test. Where discrepancies between responses on the identical items on the two occasions were no greater than a single step, such as "ENJOY very much" and "ENJOY somewhat," or "INDIFFERENT" and "DISLIKE somewhat," the item was considered consistently answered and no record of a discrepancy was made. One point was added to the "inconsistency" score when responses on repeated items differed by more than one scale step. Data
were retained for those subjects having an "inconsistency" score of "2," but a score of "3" or more was considered to suggest that the subject had responded carelessly or without a full understanding of the instructions. Table I shows the number of inconsistencies per subject.
TABLE I
FREQUENCY OF INCONSISTENT RESPONSES—RESPONSES ON REPEATED ITEMS DIFFERING BY MORE THAN ONE SCALE STEP

<table>
<thead>
<tr>
<th>Number of Inconsistencies</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>58</td>
</tr>
<tr>
<td>Data Analyzed:</td>
<td>222</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Data Disregarded:</td>
<td>30</td>
</tr>
</tbody>
</table>
CHAPTER 6

Results

Intercorrelations among Semantic Differential Scales

The first question answered by statistical methods was whether all eight descriptive scales of the semantic differential were sufficiently highly intercorrelated to measure the same thing. Table 2 gives the eight-by-eight matrix of the intercorrelations among each of the scales. The coefficients range from .831 through .321 with most values above .50. Two scales, dangerous-safe and dirty-clean, tended to correlate at low levels with all scales except one another. Because of this fact, results are reported for all eight scales, and for six scales with dangerous-safe and dirty-clean excluded.

Relationship of Semantic Differential Scores and Interest Test Item Scores

The relationships between subjects' ratings of the items containing each of the so-called "irrelevant" concepts and their semantic differential attitude scores on those concepts were determined by the use of the Pearson Product-Moment correlation coefficient formula.
**TABLE 2**

**INTERCORRELATIONS AMONG THE TEN SCALES (r)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpleasant-Pleasant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foggy-Clear</td>
<td>.704</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>.321</td>
<td>.495</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dangerous-Safe</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>.324</td>
<td>.496</td>
<td>.707</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dirty-Clean</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>.707</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>.768</td>
<td>.697</td>
<td>.364</td>
<td>.423</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninteresting-Interesting</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>.364</td>
<td>.423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>.646</td>
<td>.648</td>
<td>.588</td>
<td>.541</td>
<td>.695</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad-Good</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>.588</td>
<td>.541</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>.656</td>
<td>.679</td>
<td>.531</td>
<td>.542</td>
<td>.717</td>
<td>.812</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Sad-Happy</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>.531</td>
<td>.542</td>
<td>.717</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>.743</td>
<td>.726</td>
<td>.464</td>
<td>.462</td>
<td>.831</td>
<td>.810</td>
<td>.815</td>
<td>1.000</td>
</tr>
<tr>
<td>Dislike-Like</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>.464</td>
<td>.462</td>
<td>.831</td>
<td>.810</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Numbers along the top of the matrix correspond to those beside the scale names.
All ten relationships between scores on interest test items and semantic differential "scores" employing all eight scales are statistically significant beyond the .01 level (see Table 3). Taking into account both ends of the sampling distribution, when there are 200 degrees of freedom, an \( r \) of .18 is significant at \( P = .01 \). The lowest two correlations found were an \( r \) of .318 for vegetables and an \( r \) of .404 for selling. These two values may have been lowered slightly by the test construction error mentioned on page 23. Two items, one containing vegetables and the other using selling, were inadvertently not printed in the test booklets. Therefore, the number of times each of these concepts appeared in an interest inventory item unintentionally was reduced from ten to nine.

The largest coefficients found with all eight scales' values included in the analysis were \( r \)'s of .577 for art, .572 for drama, and .565 for history. The median correlation coefficient was found to be .488.

Table 3 also shows the correlation coefficients based on only six scales. The median correlation coefficient in this case is .498. No differences between 6-scale and 8-scale correlations exceed 0.036. Therefore, for the purposes of this study, the 6- and 8-scale values are comparable.
TABLE 3

RELATIONSHIPS BETWEEN SEMANTIC DIFFERENTIAL ATTITUDE "SCORES" FOR THE CONCEPTS AND INTEREST TEST SCORES FOR THE SAME CONCEPTS UNGROUPED DATA (r)*

<table>
<thead>
<tr>
<th>Concept</th>
<th>Eight Scales</th>
<th>Six Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selling</td>
<td>.404</td>
<td>.421</td>
</tr>
<tr>
<td>Dance</td>
<td>.423</td>
<td>.446</td>
</tr>
<tr>
<td>Drama</td>
<td>.572</td>
<td>.576</td>
</tr>
<tr>
<td>History</td>
<td>.565</td>
<td>.580</td>
</tr>
<tr>
<td>Vegetables</td>
<td>.318</td>
<td>.327</td>
</tr>
<tr>
<td>Music</td>
<td>.415</td>
<td>.425</td>
</tr>
<tr>
<td>Teaching</td>
<td>.555</td>
<td>.581</td>
</tr>
<tr>
<td>Children</td>
<td>.448</td>
<td>.475</td>
</tr>
<tr>
<td>Art</td>
<td>.577</td>
<td>.613</td>
</tr>
<tr>
<td>Dogs</td>
<td>.529</td>
<td>.522</td>
</tr>
</tbody>
</table>

* r = +.16, P < .01 when df = 200.
TABLE 4

RELATIONSHIPS BETWEEN SEMANTIC DIFFERENTIAL ATTITUDE "SCORES" FOR THE CONCEPTS AND INTEREST TEST SCORES FOR THE SAME CONCEPTS—BY SEXES*

<table>
<thead>
<tr>
<th>Concept</th>
<th>Correlation Coefficients</th>
<th>Critical Ratio**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ungrouped</td>
<td>Males</td>
</tr>
<tr>
<td>1. Selling</td>
<td>.421</td>
<td>.467</td>
</tr>
<tr>
<td>2. Dance</td>
<td>.446</td>
<td>.443</td>
</tr>
<tr>
<td>3. Drama</td>
<td>.576</td>
<td>.584</td>
</tr>
<tr>
<td>4. History</td>
<td>.580</td>
<td>.517</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>.327</td>
<td>.357</td>
</tr>
<tr>
<td>6. Music</td>
<td>.425</td>
<td>.418</td>
</tr>
<tr>
<td>7. Teaching</td>
<td>.581</td>
<td>.596</td>
</tr>
<tr>
<td>8. Children</td>
<td>.475</td>
<td>.432</td>
</tr>
<tr>
<td>9. Art</td>
<td>.613</td>
<td>.563</td>
</tr>
<tr>
<td>10. Dogs</td>
<td>.522</td>
<td>.423</td>
</tr>
</tbody>
</table>

(N=112) (N=90)

* The results shown in this table are based on six semantic differential scales.
** CR = 1.645, P = .10; CR = 1.960, P = .05.
Correlation coefficients for responses made by males and by females are listed in Table 4. For comparison purposes the correlation coefficient obtained for each concept on six scales for ungrouped data is also shown. To determine whether significantly different relationships between attitudes and interest test responses exist for males and for females, critical ratios were computed on the standard error of the difference between each of the correlation coefficients. Only one of the ten ratios is significant at the .10 level.

The order of administration of the semantic differential and the interest inventory did not influence the correlation coefficients obtained (see Table 5). Only one difference is significant at even the 10 per cent level.

The Effects of Order of Presentation and of Sex upon Semantic Differential Responses and Interest Test Scores

A close inspection of the test scores, which included computing the mean for each stimulus word for each of the four groups, viz., males and females — interest test first, and males and females — semantic differential first, uncovered some differences. Females almost always assigned higher scores to items on both experimental instruments than did the males. This tendency was particularly noticeable for the concept children on the semantic differential where a ten-point sex difference in averages exists. This difference yields a t
### TABLE 5

**RELATIONSHIPS BETWEEN SEMANTIC DIFFERENTIAL ATTITUDE “SCORES” FOR THE CONCEPTS AND INTEREST TEST SCORES FOR THE SAME CONCEPTS—FOR THE GROUP AS A WHOLE AND FOR THE TWO TESTING SEQUENCES**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Correlation Coefficients</th>
<th>Critical Ratio**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ungrouped</td>
<td>Interest Test First</td>
</tr>
<tr>
<td>1. Selling</td>
<td>.421</td>
<td>.442</td>
</tr>
<tr>
<td>2. Dance</td>
<td>.446</td>
<td>.426</td>
</tr>
<tr>
<td>3. Drama</td>
<td>.576</td>
<td>.602</td>
</tr>
<tr>
<td>4. History</td>
<td>.580</td>
<td>.606</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>.327</td>
<td>.356</td>
</tr>
<tr>
<td>7. Teaching</td>
<td>.581</td>
<td>.627</td>
</tr>
<tr>
<td>8. Children</td>
<td>.475</td>
<td>.521</td>
</tr>
<tr>
<td>9. Art</td>
<td>.613</td>
<td>.589</td>
</tr>
<tr>
<td>10. Dogs</td>
<td>.522</td>
<td>.513</td>
</tr>
</tbody>
</table>

(N=115) (N=87)

*The results shown in this table are based on six semantic differential scales.

**CR = 1.645, P = .10; CR = 1.960, P = .05.*
of 6.257 which is significant considerably beyond the one per cent level. The female participants gave interest items containing children as the "irrelevant" concept significantly higher ratings also ($t = 5.346$, which is significant at the .01 level). As shown in Table 6, females' semantic differential scores are significantly higher on seven of the ten concepts, while their ratings of interest items containing five of the "irrelevant" concepts are significantly higher than those of the males.

Subjects who took the semantic differential first also assigned higher values to both tests than did those who took the interest test first. Eight out of ten semantic differential mean scores were higher when the semantic differential was administered first; however, of the eight, only values for four concepts are significantly higher ($P < .05$). Although seven of the ten concepts were rated higher on the interest test by subjects taking the semantic differential first, none of the differences yields $t$'s significant at even the five per cent level.

The tables in Appendix C give the mean scores and standard deviations for each of the experimental concepts appearing in the semantic differential and in the interest test. Results are shown for ungrouped data, data of males, and data of females.
TABLE 6

SIGNIFICANCES OF DIFFERENCES BETWEEN SCORES
ASSIGNED BY SEXES AND BY GROUPS (t)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Semantic Differential Males, Females*</th>
<th>Testing Sequences*</th>
<th>Interest Test Males, Females*</th>
<th>Testing Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Selling</td>
<td>-1.854**</td>
<td>1.667</td>
<td>0.990</td>
<td>0.678</td>
</tr>
<tr>
<td>2. Dance</td>
<td>-2.187***</td>
<td>-0.524</td>
<td>-5.063***</td>
<td>-0.489</td>
</tr>
<tr>
<td>3. Drama</td>
<td>-3.957****</td>
<td>-3.104***</td>
<td>-2.423**</td>
<td>-1.628</td>
</tr>
<tr>
<td>4. History</td>
<td>0.549**</td>
<td>-2.949***</td>
<td>0.320</td>
<td>-1.757</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>-2.722***</td>
<td>-2.063**</td>
<td>-0.703</td>
<td>-0.449</td>
</tr>
<tr>
<td>6. Music</td>
<td>-1.967**</td>
<td>-1.973**</td>
<td>-2.836***</td>
<td>-0.148</td>
</tr>
<tr>
<td>7. Teaching</td>
<td>-2.089**</td>
<td>-1.278</td>
<td>0.492</td>
<td>-0.787</td>
</tr>
<tr>
<td>8. Children</td>
<td>-6.257****</td>
<td>1.843</td>
<td>-5.346***</td>
<td>0.090</td>
</tr>
<tr>
<td>9. Art</td>
<td>-2.279**</td>
<td>-1.043</td>
<td>-3.555***</td>
<td>-1.592</td>
</tr>
<tr>
<td>10. Dogs</td>
<td>-0.754</td>
<td>-0.772</td>
<td>-0.652</td>
<td>0.278</td>
</tr>
</tbody>
</table>

* Negative t's refer to significantly higher female or semantic differential first values.
**  $P < .05$.
*** $P < .01$. 
CHAPTER 7

Discussion and Conclusions

The results of this experiment leave little doubt that the so-called "irrelevant" concepts contained in many interest inventory items are not irrelevant at all. They, by being an integral part of the item, are relevant and influence responses made to the item. If the correlation for the sampling population between the semantic differential responses to these words and the responses to the interest items were actually 0.00, only once in 100 trials would an \( r \) as large as \( \pm 0.16 \) arise from fluctuations of sampling alone. The smallest \( r \) obtained between semantic differential ratings and interest inventory scores for combined male and female data in this experiment was +0.318. Most of the obtained \( r \)'s were considerably larger, and all were positive. There would be no choice but to reject a null hypothesis, had one been made, that no significant relationship exists between subjects' reported feelings about each secondary concept and their rating of the interest item in which each concept appeared.

It would seem that this phenomenon may be chiefly attributed to the subjects' failures to discriminate between the so-called "relevant" and "irrelevant" concepts in an
The data show quite clearly that the subjects responded to more than one concept in the multiconcept items constructed for this study. It appears that the subjects, or some of the subjects, responded to one concept while others responded to another. They answered at least some items in terms of the immaterial (wrong) concept rather than on the basis of the concept keyed to an interest area. Most subjects answering the Kuder Preference Record — Vocational item "Write articles about wild animals" would be extremely surprised to learn that MOST LIKE selection of this activity raises only the Outdoor score and in no way affects the Literary profile. Should the subject respond to write articles or to wild animals?

The truism, that every word in a sentence occurs in a context of other words, is almost too elementary to be discussed in this thesis. It is obvious that the meaning of a word in even a phrase depends on the pattern and nature of other words in the unit. Unless external verbal contexts are supplied for the reader of a complex sentence, the reader will provide his own internal verbal context (Krech, 1962). When answering a multiconcept interest inventory item, the participant must supply his own verbal context.

It seems to this writer that there must be at least two stages through which the inventory-taker must guide his
cognitions if he is to answer intelligently most multiconcept items.

First, he must make sure that he is thinking of the relatively universal meaning of such concepts as shampoo, statistics, stenography, artificial lung, erosion, and mimeograph — to name only a few of the words which have been suspected by Kuder (1948) of having variant or unknown meanings for subjects.

Second, if content validity has been utilized in standardizing the test, the participant must ultimately focus his attention on one concept, and only one, on which he will base his final decision of interesting or uninteresting. Unless he happens to consider all concepts as equally pleasant or unpleasant, interesting or uninteresting (which this writer believes is highly unlikely), the participant must say in effect, "I like this concept, but that one is very distasteful to me. The latter one must be the really important thing here. Therefore, I will rate the item unfavorably." He has considered all concepts, but since they represent differing valences to him, his deliberations necessarily must end with judgment of one concept. This concept most probably is the one which evokes the strongest emotive response, either positively or negatively.
Only if the participant has unusual knowledge about the specialized and complex activity presented to him, could he validly rate the item *in toto* without somehow internally reducing the scope of his judgment. Such extensive acquaintance with the activity should permit the participant to respond to the entire item, regardless of its complexity or ambiguity.

As an example of this process one of the items from the Kuder Preference Record - Vocational can be utilized: "Train dogs to lead blind people." One might argue that the verb *train* is the dominant idea in the item, and that a response to the item actually represents a response to the idea of *training* animals. However, the concept of *training* is far from invariant in meaning. Training parakeets to talk might be familiar and fun for the subject while the prospect of exercising patience and ingenuity with a huge German Police dog could be boring or frightening. *Train* could evoke such a prominent response that the subject looks no further into the item.

Others might argue that *dogs* is the salient concept, and that it is representative of a variety of faunal objects. However, what happens if the subject is not aware of the type of dogs which are typically used for "seeing-eye" training? He may think of small dogs, short-haired dogs, lap dogs, big
dogs, friendly dogs, aggressive dogs, etc. An ardent canine devotee may find it difficult to consider train or to lead blind people, and would likely rate the item on the basis of dogs only. Conversely, an abhorrence to dogs, if radical enough, might well be responsible for an unfavorable rating on the entire item.

Does the training of dogs to lead blind people entail close association with the blind? If a benevolent subject thinks so, he might welcome the opportunity to satisfy his altruistic needs; uncharitableness could cause the opposite effect. Is it possible that a willingness to accept responsibility is involved in the answering of this item? If one of the "seeing-eye" dogs to be trained by the participant errs and leads a blind person into the path of an oncoming vehicle, is the trainer morally liable? An extensive acquaintance with this occupation could just as easily require the dog lover to reject this item on the basis that he could not comply with the regulation that animals which prove to be untrainable be destroyed to preclude their later fraudulent sale as competent "seeing-eye" dogs.

To summarize this point of view, it appears that an average or below average familiarity with an activity necessitates ultimate reaction to only one concept in a multi-concept item; it is on the basis of this one concept that the
conscientious inventory-taker will make his final judgment as to whether the activity is interesting or uninteresting for him. Only extensive and complex knowledge about such activities would permit his reaction to the sum of the parts of an interest item.

The inventory writer is faced with a dilemma. He must make the test valid enough to be acceptable to the critics, yet interesting, colorful, and realistic enough to command the participant's attention for forty-five minutes or longer. Limiting the items to only one concept is not a feasible solution to the problem. Even if sufficient items could be so constructed, such an inventory would be drab and meaningless in many respects.

If the attitudinal components of the meaning which a concept holds for a subject were known, items containing concepts which evoke extreme reactions could either be discarded or weighted according to how atypical the reaction is. This would result in lengthy testing procedures and would necessitate the use of an electronic computer for scoring, making the use of such an inventory prohibitively expensive in both time and financial cost.

Randomizing a large number of secondary and tertiary concepts throughout the inventory still would leave much to chance and in no way would increase the validity of single
items. Neither would the whole test validity be expected to increase.

There are two possible solutions which this writer believes are worthy of experimental validation. In the first, forced-choice preference-type items would be constructed so that all elements of an item, except the "relevant" concept, are held constant within any one comparison. The participant could then be asked to decide which he would MOST LIKE and which he would LEAST LIKE:

- a. Visit an art gallery.
- b. Draw the plans for an art gallery.
- c. Be the bookkeeper for an art gallery.

In this way, only one collective concept is being judged since there is no opportunity to choose an item with a more favorable secondary concept.

A second possible solution would involve the moving away from preference-type items in deference to the single item, individually rated as LIKE, INDIFFERENT, or DISLIKE with possible intermediate ratings. The inventory author would choose a predetermined number of secondary concepts on the basis of their relatively invariant meaning to facilitate mutual understanding. Since no group of words would evoke common, invariant attitudes in all subjects, a control of using every secondary concept with an activity stem from each major occupational area would be necessary. The relevant
concept, which in most cases would contain the predicate of the sentence, should be italicized or otherwise differentiated from the immaterial concepts. This might increase the inventory's susceptibility to deliberate, conscious faking since subjects would have no doubt as to the important concept in the item. But, more importantly, it should greatly increase the test's validity since this practice would minimize judgments made on the basis of the wrong concept; the use of each irrelevant concept with every relevant concept would eliminate the remaining effect upon total score.

This study has also shown that the effect of the non-primary concepts does not depend on whether the subject is male or female. Although sex may be related to the direction of the effect, there can be no question that both groups are significantly vulnerable to the spurious effects of the so-called "irrelevant" concepts.

At least six of the eight scales used in this study measured the same dimension of semantic space. An inspection of the names of the scales strongly suggests their evaluative, and possibly emotional, nature. It is not so clear why dangerous-safe and dirty-clean correlated highly with each other but less highly with the other scales. Osgood (1957, p. 71) suggests that scales "defined as evaluative do produce factors beyond the general evaluative factor." The number of
scales analyzed in this study was too small to indicate whether a non-evaluative dimension was emerging through these two scales. It is not so much the significance levels of the obtained correlations which cast doubts on their kinship to the evaluative factor ($P \ll .01$); rather, it is the fact that they both consistently correlated much lower with other scales than did any of the remaining six scales. The fact that results which included and those which excluded the questionable scales were practically the same lends support to Osgood's hypothesis that these two factors may merely describe a more specific evaluative factor.

Although statistical analyses have indicated that the relationships between interest test and semantic differential scores are not dependent upon the order of administration, comparison of the two scores for specific words suggests that some concepts may be affected by a transfer effect from the first-administered instrument. After a subject has expressed his feelings in either test situation, it is logical to expect that his need to be internally consistent would influence some of his ratings on a second test. For this reason, the experimenter suggests that any future research of this nature allow for the administration of the attitude measuring device and the interest inventory at least several days apart.
An analysis of the study in retrospect leads the experimenter to wish that there had been a measure of the subjects' attitudes toward the "relevant" concepts. This would have determined what percentage of the total variance is explainable by the effect of the "irrelevant" concept. Any replication study should definitely include a semantic differential score for each so-called "relevant" concept in the test items.
CHAPTER 8

Summary

This study investigated the vulnerability of interest test items to unwanted response effects arising from the presence of so-called "irrelevant" concepts written into the items to give them variety and realism. In the item "Determine the cost of shipping vegetables in different ways," the italicized words would be considered a "relevant" concept because of their relation to a general computational type interest; all other words in the item are generally regarded as "irrelevant" since they have no intentional bearing on the interest area of concern to the particular test. From among these "irrelevant" words in the item, the experimenter selected the concept vegetables to serve as an indicator of the "irrelevant" concepts, since vegetables seemed to be the least neutral one.

This rationale formed the basis for the construction of 107 singly-rated interest inventory items. Ten "irrelevant" concepts, chosen from currently used interest tests, were paired with an activity stem written to investigate the subjects' interests in ten broad occupational areas. Only ninety-eight items so constructed were actually administered since two were unintentionally not printed in the test.
booklets. The remaining items were duplicates of others and were included to check the accuracy of the subjects' test-taking behavior. Each item was rated from "1" to "5." Subjects were allowed to vary two or more scale points between their initial and subsequent ratings of only two repetitious items. Greater disparities resulted in rejection of that subject's entire data set. The assumption behind these exclusions was that the subject either did not understand and follow instructions or that he had not responded carefully. Sixty-two sets were rejected for inconsistencies, because of improper marking practices followed when completing response cards, or the failure of subjects to return all response cards. After these exclusions, 222 response sets remained. The interest test was scored on the basis of the "irrelevant" concept which an item contained.

Semantic differential ratings of the ten "irrelevant" concepts served as indicators of the subjects' attitudes toward those concepts. The subjects, eleventh and twelfth grade high school students, had ten increments on which to rate the concepts in terms of eight bipolar evaluative scales. A score of "1" represented the most unfavorable extreme, a "9" signified the most favorable, and the intervening numerical increments were indicative of ratings somewhere between the bipolar extremes. There was no neutral increment; it
was therefore impossible for a subject to indicate that neither descriptive scale applied to a concept, or that he had "NO ATTITUDE" about a concept in terms of a scale.

All responses were recorded on IBM Mark Sense cards which were electro-mechanically translated into Hollerith codes and were computer scored and analyzed.

Six of the eight descriptive scales were found to be highly intercorrelated with each other. Two tended to correlate at low levels with all scales except one another.

All ten relationships between subjects' ratings of interest test items, containing each of the so-called "irrelevant" concepts, and their semantic differential attitude scores for those concepts were found to be significantly correlated with each other ($P < .01$).

The results show that the non-primary concepts in multi-concept interest inventory items cannot be considered to be "irrelevant." Their presence in the item definitely affects the rating given that item by the subject. Design of the experiment did not provide for determination of the percentage of the total variance which may be explained by this phenomenon.

The effects of the "irrelevant" concepts existed with equal significance for male and female subjects. The relationship between interest test ratings and attitudes was
not dependent upon the order of administration of the two instruments.
APPENDIX A

SEMANTIC DIFFERENTIAL PROFILES FOR THE EXPERIMENTAL CONCEPTS — UNGROUPED DATA

Unpleasant - Pleasant
Foggy - Clear
Dangerous - Safe
Dirty - Clean
Uninteresting - Interesting
Bad - Good
Sad - Happy
Dislike - Like

LEGEND
Selling
Dance
Drama
History
Vegetables
LEGEND

Music
Teaching
Children
Art
Dogs
APPENDIX B

A Complete Copy of the Interest Test

Face Sheet
CODE

5 - I would ENJOY this very much.
4 - I would ENJOY this somewhat.
3 - I am INDIFFERENT about this.
2 - I would DISLIKE this somewhat.
1 - I would DISLIKE this very much.

MAKE SURE YOU USE ONLY CARD NO. ONE WITH THIS SHEET!

MARK HEAVY.

Determine the cost of shipping vegetables in different ways
Take an art course in scenery painting
Operate the lighting panel for a drama production
As a psychologist, interview children in an experiment
Teach a course in general business
Fix a broken music box
Ask people's opinions about a new frozen vegetable product
Organize results from opinion surveys about a new frozen vegetable product
Sell tickets to a leadership course
Conduct experiments on the effects of training dogs in different ways
Teach a course in agriculture
Write a poem which you would like have set to music
Develop an electronic musical instrument which anyone can play
Be a teacher of painting, modeling, and ceramics
Sell farm equipment in a rural area
Learn about the history of early musical forms
Sell goods at very reduced rates to people on relief
Be the bookkeeper for an art gallery
Read music appreciation lessons to a blind student
Make a very artistic jigsaw puzzle
Draw pictures of a stage set for a drama
Teach students who are slow learners
Write a musical comedy with the help of a famous dramatist
Chaperone a dance in a slums area of town
Take care of a well-known pianist's dogs while the pianist is on tour
Put together an electric vegetable peeler
Help plan the budget for a school musical production
5 - I would ENJOY this very much.
4 - I would ENJOY this somewhat.
3 - I am INDIFFERENT about this.
2 - I would DISLIKE this somewhat.
1 - I would DISLIKE this very much.

MAKE SURE YOU USE ONLY CARD NO. TW O WITH THIS SHEET!

M A R K H E A V Y.
CODE

5 - I would ENJOY this very much.
4 - I would ENJOY this somewhat.
3 - I am INDIFFERENT about this.
2 - I would DISLIKE this somewhat.
1 - I would DISLIKE this very much.

MAKE SURE YOU USE ONLY CARD NO. THREE WITH THIS SHEET!
MARK HEAVY.

CARD NO. THREE

persuade children to get their parents to buy a new cereal
Design costumes for a dance review
Raise vegetables
Teach a course in mechanical engineering if you had the ability
Draw the plans for a dance studio
Be a science teacher
Write articles for an art magazine
Convince financiers to back a historical excavation
Be well known as a famous musician
Help immigrants who know little English to learn their U.S. History lessons
Sell reproductions of famous paintings in an art gallery
Draw sketches of musical instruments for a yearbook
Mend a broken toy for a child
Lead a children's orchestra if you had the ability
Type original copy of a history book that is being published
Write a newspaper article about a visiting dance troupe
Demonstrate a modern cash register while a salesman is trying to sell it
Compose the music to accompany a dance routine
Determine the cost of putting on a dance review
Be a tractor salesman
Get the programs and tickets printed for a drama presentation
Take neighborhood children on a fishing trip
Raise fine dogs
Teach a course in arithmetic
Try to get a big crowd to attend a school dance
Teach a class a new theory which contradicts a traditionally held view
Select the best books on dogs for a school library
CAR D NO. FOUR

Through interviews learn people's opinions about a new kind of modern art

Be the treasurer for a historical society

Draw illustrations of vegetables for a magazine

Train "seeing-eye" dogs to lead blind people

Be a music teacher

Sell classical records in a music store

Write a play for a drama group to perform

Be a private secretary to a famous musician

Sell your services as a mathematician

Help children model and paint

Paint portraits of dogs for a hobby

Serve as advertising agent for a concert singer

Help crippled people learn art in a rehabilitation center

Sort fan mail being received by a professional dancer

Paint watercolor impressions of important events in history

Do research to determine the symbolism in contemporary dance forms

Select the best books on dogs for a school library

Teach a class a new theory that contradicts a traditionally held view

Assist with a drama group for an organized charity

Act as chairman of a publicity committee for a dog show

Help plan the budget for a school musical production

Put together an electric vegetable peeler

Sell chemical laboratory apparatus

Be the writer of children's stories

Determine the cost of shipping vegetables in different ways

Take an art course in scenery painting

***END OF THIS SECTION***
APPENDIX C, TABLE I

MEAN SCORES AND STANDARD DEVIATIONS
OF EXPERIMENTAL CONCEPTS—UNGROUPED DATA

<table>
<thead>
<tr>
<th>Concept</th>
<th>Semantic Differential</th>
<th>Interest Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>2. Dance</td>
<td>42.212</td>
<td>12.490</td>
</tr>
<tr>
<td>3. Drama</td>
<td>32.892</td>
<td>14.179</td>
</tr>
<tr>
<td>4. History</td>
<td>35.757</td>
<td>14.270</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>34.072</td>
<td>10.576</td>
</tr>
<tr>
<td>7. Teaching</td>
<td>33.495</td>
<td>13.005</td>
</tr>
</tbody>
</table>

* The results shown in this table are based on six scales on the Semantic Differential.
**APPENDIX C, TABLE 2**

**MEAN SCORES AND STANDARD DEVIATIONS OF EXPERIMENTAL CONCEPTS—BY SEXES**

**MALES**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Semantic Differential</th>
<th>Interest Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>1. Selling</td>
<td>27.054</td>
<td>13.474</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>32.179</td>
<td>10.031</td>
</tr>
<tr>
<td>8. Children</td>
<td>34.348</td>
<td>12.105</td>
</tr>
</tbody>
</table>

*The results shown in this table are based on six scales on the Semantic Differential.*
APPENDIX C, TABLE 3

MEAN SCORES AND STANDARD DEVIATIONS
OF EXPERIMENTAL CONCEPTS—BY SEXES*

<table>
<thead>
<tr>
<th>Concept</th>
<th>Semantic Differential</th>
<th>Interest Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>2. Dance</td>
<td>44.333</td>
<td>12.384</td>
</tr>
<tr>
<td>3. Drama</td>
<td>37.511</td>
<td>12.706</td>
</tr>
<tr>
<td>4. History</td>
<td>35.422</td>
<td>14.909</td>
</tr>
<tr>
<td>5. Vegetables</td>
<td>36.211</td>
<td>10.875</td>
</tr>
<tr>
<td>6. Music</td>
<td>46.100</td>
<td>8.803</td>
</tr>
<tr>
<td>8. Children</td>
<td>44.256</td>
<td>9.783</td>
</tr>
</tbody>
</table>

* The results shown in this table are based on six scales on the Semantic Differential.
LIST OF REFERENCES


