THE SEMANTIC DIFFERENTIAL: EXPLORING THE EQUAL-INTERVAL ASSUMPTION

BY

STEVEN MERRILL SCHNITZER

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1970

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
May, 1979

Say not "this is the truth" but "so it seems to me,

to be, as I now see the thing I think I see."

--Anon

THE SEMANTIC DIFFERENTIAL: EXPLORING THE

EQUAL-INTERVAL ASSUMPTION

Thesis Approved:

W. J. Ward
Thesis Adviser

William R. Berg

Marshall Alle

Norman N Durham
Dean of the Graduate College

PREFACE

This is a study of a judgmental rating device termed the semantic differential. The semantic differential (or S.D.) was conceived some thirty years ago as a behavioral tool to measure human affective and emotive feeling. This study's principal purpose was to analyze some of the semantic differential's primary constructs with a view toward increasing its validity.

The seeds for this work probably were sown, unbeknownst at that time to the author, in a general semantics class during the spring semester of 1977. It was always mysterious to me how small, black, inconspicuous books can wreak such profound effects on people. In undergraduate days I used to feel genuine sympathy for some of my (more devoted) fellow students in the physical sciences whom I regularly would see live, eat and sleep with such objects.

Then, at last, I was inspired. The little black book that did it, in my case, was Wendell Johnson's People in Quandaries. As inconspicuous things may go, the book and its attendant university course surely stand at the summit with high school typing as the two most invaluable courses I ever enrolled in, applied to practical everyday living.

I would like to thank all those who had a part in this project. Especial among those is Dr. Walter J. Ward, director of graduate studies in Mass Communication at Oklahoma State University and my major adviser. He was a constant source of inspiration and guidance from beginning to completion.

Likewise, I would like to acknowledge the attention and assistance of other members of my committee, Dr. William R. Steng, associate professor of Journalism and Broadcasting, and Marshall E. Allen, associate professor of Journalism and Broadcasting.

To my mother and father, for no small measure of encouragement and support, I am indebted.

Finally I must express my appreciation to Dr. Ray Harrell, longtime high school educator, dean of men, counselor, adviser, and most of all, friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

From the beginning of recorded philosophical thought with the ancient Greeks through the much younger discipline of psychology in the 19th Century, the pursuit of meaning has been a common calling.

Characteristic of these two disciplines, their respective addressings of meaning typically have had diverging purposes. In philosophy, meaning is focused upon as a second-order physical and conceptual phenomenon. It becomes an element of the philosopher's two-part problem of meaning and verification: First, how does one find out what is the meaning of a statement? Second, how does one find out whether the statement is true? To answer both, he must draw upon and make statements about the meaning and truth of other statements.[1]

Psychology addresses meaning in the associational orb of stimulus and response.

> In an attempt to explain semantic generalization it has been hypothesized that meaning is actually a response which serves, in turn, as a cue . . . The generalization of a response from the word won to beat is based upon their common meaning. This makes the generalization possible . . . The galvanic skin response is first conditioned to the written word won. The subject is then exposed to the word beat, which already has the tendency to evoke the response won because of previous learning which established them as synonyms. The response won serves as the stimulus cue to evoke the galvanic skin response. In other words, beat is not directly linked to the

> galvanic skin response; its capacity to evoke such a response is mediated by the ability of beat to elicit won.[2]

The only commonality (perhaps unfortunately) between the two disciplines in this regard was that invariably the quest for meaning veered toward virtually every object besides semantic meaning itself. The edification which was to provide an exclusive attention to meaning through words as symbols has only recently enjoyed considerable popularity--though its roots are nearly as remote in time as ancient philosophy.[3]

The study of general semantics, a term originated by Alfred Korzybski in 1933 as a name for a general theory of evaluation, is an empirical science which espouses a comprehensive method for human adjustment in our private, public, and professional lives.[4] It provides a system of evaluation and predictability in solving human thought-process maladies.

Modern-day application of general semantic thought began with Korzybski, a Polish-American scientist.[5] Korzybski, after the debacle of World War I, became convinced of ideational voids in the philo-psychological sciences; the precipitating factors of such human disasters which required addressing.

In examination of human existential problems, he found vital a wholesale revision of old notions about human nature set down by the Greek philosopher Aristotle 2,300 years ago and largely used as a pattern for living by mankind since.[6] Korzybski likened the premise of his non-aristotlian system to an analogy of a map and its territory:

> First. A map is not the territory (notion of non-identity).
> Second. A map does not represent all of a territory (notion of non-allness).
> Third. A map is self-reflexive in the sense that an 'idea'

> map should include a map of the map, etc., indefinitely (notion of non-contradiction).
>
> Applied to daily life and language:
>
> First. A word is not what it represents.
> Second. A word does not represent all of the 'facts', etc.
> Third. Language is self-reflexive in the sense that with language we can speak about language.[7]

Korzybski concluded, through his research and a view of world history, that the composition and structure of these forms of representation (i.e., language) were instrumental in shaping the historical interaction and behavior of human cultures.

Drawing contrast with an exact science he asked:

> Why is it that structures built by engineers do not as a rule collapse, or if they do, then the physico-mathematical or other evaluational errors are easily discovered; yet social, economic, political, etc., systems, also man-made, do sporadically collapse in the form of wars, revolutions, financial depressions, unemployment, etc?

This led to the question:

> What is it that engineers <u>do</u> neurologically when they build bridges, etc?[9]

The answer was:

> They use a special, narrow but 'perfect' language called mathematics, which is <u>similar in structure to the facts they deal with</u>, and which therefore yields predictable empirical results.[10]

Korzybski then investigated what builders of political, economic, social, and other inexact human behavioral sciences do and found that they utilize languages (forms of representation just as the engineer's mathematics) which are not similar to facts of exact sciences. Consequently, these results are unpredictable and breakdowns or collapse follow.

Works of a number of scholars including Johnson, Hayakawa, Lee and Thorndike followed, but the development of a standardized tool for

measuring human affective judgment of words and concepts was to be reserved for a Yale graduate student named Charles E. Osgood. Osgood had been interested in the way in which people draw inferences about what is happening at the semantic (ideational) level of a fellow communicant's mind. Indeed, we do not see and hear the mind of the other, but rather we observe his gestures and written or spoken words--otherwise termed the communicative product. Osgood reported:

> An extensive survey of the literature fails to uncover any generally accepted, standarized method for measuring meaning . . . This certainly seems simple enough, yet it has troubled philosophers for centuries . . . Perhaps it is because of the philosophical haziness of this concept, perhaps because of the general belief that "meanings" are infinitely and uniquely variable, or perhaps because the word 'meaning' as a construct in our language connotes mental stuff, more akin to 'thought' and 'soul' than to anything observable--for some combination of reasons there has been little attempt to devise methods here. Nevertheless, whether looked at from the viewpoints of philosophy or linguistics, from economic or sociological theory, or--interestingly enough--from within the core of psychological theories of individual behavior, the nature of meaning and change in meaning are found to be central issues.[11]

Osgood's search for a more precise quantitative assessment of meaning expressed through words led him to adopt a controlled association and scaling device based on philosophic-psychologic, semantic, and geometric principles.

> We provide the subject with a concept to be differentiated and a set of bipolar adjectival scales against which to do it, his only task being to indicate, for each item (pairing of a concept with a scale), the direction of his association and its intensity on a seven-step scale. The crux of the method, of course, lies in selecting the sample of descriptive polar terms. Ideally, the sample should be as representative as possible of all the ways in which meaningful judgments can vary, and yet be small enough in size to be efficient in practice. In other words, from the myriad linguistic and non-linguistic behaviors mediated by symbolic processes, we select a small but carefully devised sample, a sample which we shall try to demonstrate is chiefly indicative of the ways that meanings vary, and largely insensitive to other sources of variation.[12]

Since its inception some twenty-five years ago, the technique of semantic differentiation has increasingly captured the interest and imagination of social researchers, psychologists, behaviorists, and pollsters. The method has been utilized in a tremendous variety of studies, and in 1968 it was effectively used to help elect a president.[13] Hundreds of related articles have appeared in the interim in trade and professional publications.

Because of its extreme flexibility, and thus, proliferation (and subsequently its inevitable modification and contortion) in literature and studies of considerable variation, an understanding of the semantic differential and ramifications of utilization of varying models increasingly has become difficult to obtain.

In essence, the semantic differential is a scaling instrument designed to give quantitative representation to a continuum along which reactions, opinions, or judgments vary. In the early course of its use, the technique acquired its label through semantical orientation of the scale via polarized and/or graduated verbal-quantitative descriptions linked by geometric progression along the scale continuum. (See Figure 1.)

Good ___:___:___:___:___:___:___ Bad

Figure 1. The Semantic Differential Scale

The name for the process is not without its merits, since, as Osgood and Suci point out: ". . . this label points quite accurately

to the intended operation--a multivariate differentiation of concept meanings in terms of a limited number of semantic scales of known factor composition."[14]

The device has proved to be a precise instrument for the recording of affective stimuli associations. Through analysis of a wide variety of studies, it has been determined that affective judgments on bipolar semantic differential scales tend to resolve themselves into three categories or dimensions, namely: Evaluation, Potency and Activity (otherwise termed measures of semantic space.)[15] The Evaluative dimension is represented by such scales as good--bad, kind--cruel, beautiful--ugly, fair--unfair and wise--foolish. Scales labeled with hard--soft, strong--weak, large--small and masculine--feminine commonly connote Potency. The Activity dimension can be represented by fast--slow, hot--cold, sharp--dull and active--passive.[16] Factor analysis of large groupings of semantic differential scales has demonstrated that frequency of occurrence of the three dimensions falls in the order listed above.[17]

The principles of semantic differential methodology may be summarized as follows:

> 1. Ratings on bipolar adjective scales--whatever the number and variety of scales used--are largely a function of a few dimensions of judgment.
> 2. These dimensions or factors are meaningfully related to affect.
> 3. A few appropriate scales can be used to obtain reliable measurements on any one dimension.
> 4. Measurements made on a given dimension are comparable for stimuli of greatly different character (words, colors, sounds, etc.).[18]

The Problem

Utilizing Osgood's criteria for satisfactory measuring instruments, one can define academically what is desired of the ideal measurement

tool for a given situational study:

> (a) Objectivity. The method should yield quantitative and verifiable (reproducible) data. (b) Reliability. It should yield the same values within acceptable margins of error, when the same conditions are duplicated. (c) Validity. The data obtained should be demonstrably covariant with those obtained with some other, independent index of meaning. (d) Sensitivity. The method should yield differentiations commensurate with the natural units of the material studied, i.e., should be able to reflect as fine distinctions in meaning as are typically made in communicating. (e) Comparability. The method should be applicable to a wide range of phenomena in the field, making possible comparisons among different individuals and groups, among different concepts, and so on. (f) Utility. It should yield information relevant to contemporary theoretical and practical issues in an efficient manner, i.e., it should not be so cumbersome and laborious as to prohibit collection of data at a reasonable rate.[19]

The great body of research heretofore accomplished appears to have been concentrated in the areas of objectivity, reliability, validity and comparability. A review of the literature fails to reveal comprehensive study in the criterion of sensitivity.[20] States Osgood:

> There has been no explicit statement of the relation between the theoretical conception of meaning as a representational mediation process, and the operations of measurement which constitute the semantic differential technique.[21]

Specifically, there has been a minimum of analysis on the geometric linguistic makeup of the measurement device itself especially as related to the portrayal of weighted/gradated semantic space of that scale. In this study, the question is one of whether quantitative differentiation of a semantic differential (S.D.) scale is significantly affected by (1) the mandatory allocation of responses to linearly and equidistantly gradated space within the scale, and (2) the inclusion of verbal differentiators of intensity (termed "linguistic quantifiers" by Osgood) in pre-testing instructional material, on the evaluational scales themselves, or in analysis summaries to represent computed affective/emotive intensity.

Finally--and of most import, if either or both of the above factors are significant, is present measurement accuracy compromised?

Purpose and Objectives

Referring once again to item (d) of Osgood's criteria for satisfactory measuring instruments:

> Sensitivity. The method should yield differentiations commensurate with the natural units of material studied, i.e., should be able to reflect as fine distinctions in meaning as are typically made in communicating.

This study sought to determine if sensitivity of the S.D. scale could be increased. Could any enhanced specificity be wrought through design modifications in the scale itself and minor modifications in the method of analysis? Furthermore, would any such modifications increase the conformity of the S.D. scale to Osgood's criteria in the area identified above?

ENDNOTES

1. Elmer Sprague and Paul W. Taylor, Knowledge and Value (New York, 1959), p. 252.

2. Howard H. Kendler, Basic Psychology (New York, 1963), p. 346.

3. Wendell Johnson, People in Quandaries (New York, 1946), pp. 23-26. The Greek philosopher Heraclitus likely qualifies as the grandfather of semantic thought with his contention that "one cannot step in the same river twice."

4. Alfred Korzybski, Science and Sanity (Lancaster, Pa., 1933), pp. 1-110.

5. _____, Manhood of Humanity (New York, 1921), pp. 1-26.

6. Johnson, p. 6.

7. Korzybski, Science and Sanity, p. 6.

8. Ibid., p. 14.

9. Ibid., pp. 15-16.

10. Ibid.

11. Charles E. Osgood, "The Nature and Measurement of Meaning," Psychological Bulletin, Vol. 49, p. 199.

12. _____, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Ill., 1957), p. 20.

13. Joe McGinniss, The Selling of the President 1968 (New York, 1969), pp. 79-82.

14. Charles E. Osgood and George J. Suci, "Factor Analysis of Meaning," Journal of Experimental Psychology, Vol. 50, p. 325.

15. David R. Heise, "Semantic Differential Profiles for 1000 Most Frequent English Words," Psychological Monographs, Vol. 79, p. 1.

16. John B. Carroll, "Review of the Measurement of Meaning," Language, Vol. 35, p. 66.

17. Jum C. Nunnally, Psychometric Theory (New York, 1967), pp. 536-537.

18. Heise, pp. 1-5.

19. Osgood, p. 221.

20. Norman Cliff, "Adverbs as Multipliers," Psychological Review, Vol. 66, pp. 27-44. This work is apparently the most approximate related study.

21. Osgood, Suci, and Tannenbaum, p. 25.

CHAPTER II

REVIEW OF LITERATURE

Introduction

If the concept of "measurement" is important in research, it is doubly important in research on a measurement device itself. When the research forays upon a semantical measurement tool, the stage is really set. The researcher finds himself fighting two wars rather than one; not only is he struggling with the primary analysis for which purpose his work is justified, but he is warring semantically with every fundamentalist definition of measurement.

Probably the gravest error of a student of semantics would be to make the Aristotelean assumption that his weights and standards were an actual measure of an entity, rather than an attribute of the entity to a convention established by him or his peers.

The point was made succinctly by Kerlinger, who suggested the situation of a male judge who was asked to stand near an attractive young woman and rate her qualities of niceness, strength of character, personality, musical ability, and intelligence on a scale of one to five.[1] Deduced scientifically, the problem would be "How will he ever analyze the criteria by simply looking at her?" Deduced semantically, the problem becomes "How will we analyze the findings if he gets them?"

It is easy to understand why one analyst wrote:

In exasperation about the confusion in theories of measurement,

> it is tempting to wish that there were no yardsticks and no balances . . . Then all scientists might more readily see that measurement is a matter of convention rather than of discovering the 'real' measure.[2]

But every measurement is purposive. So long as the system is a method commonly understood for what it is by the encoder, transmitter/receiver and decoder, it is scientifically and semantically worthwhile. It may help move mountains or crack atoms.

To be measurable, an attribute must fit the specifications of a quantitative variable.[3] And to the degree one is successful in defining attributes in a quantitative manner, a reason for devising a system to measure those attributes is established.[4] Notably, it makes no sense to attempt a measurement of some attributes. A person's name, place of birth, or race might be examples. The property observed, then, is both dependent on man's ability to conceive it and his efficacy in observing it.[5]

As science develops new analytic abilities in every facet of observational phenomena, there are fewer and fewer areas that justifiably can exclude it. Every triumph for the scientific approach is one more strike against the Old Man's prescientific civilization.[6] As Protagoras announced, "Men are the measure of all things."

Evolution of the Semantic Differential

Although the semantic differential is a relatively recent development, its ancestry is not. Attitudinal- and ability- studies in the psychological and sociological sciences have precipitated a variety of data-gathering tools. Evolution of the semantic differential can be traced through these works.

Guilford reports the first rating scale to be applied to a psychological problem was by Galton on a vividness-of-image test in 1883.[7] Pearson used a seven-point scale for estimating intelligence in 1906.[8]

L. L. Thurstone enhanced methodology with his incorporation of equal-appearing intervals into the rating scale in 1929.[9] He was intent on devising an attitude polling device which reflected a range of choices actually possessed by the public. Thurstone utilized a panel of judges drawn from the subject population to rank-order a group of social issues. He then took those statements which most nearly represented an ascending equidistant discriminal distribution, and used these to compose the attitude scale. Other statements were discarded.

In Thurstone's scale, the lower the scale value, the more positive the attitude, as is demonstrated in a 1929 questionnaire addressing social attitudes toward the church.

> (1) I believe the church is the greatest institution in America today. (scale value: .2)
> (2) I believe in religion, but I seldom go to church. (scale value: 5.4)
> (3) I think the church is a hindrance to religion for it still depends upon magic, superstition, and myth. (scale value: 9.6)[10]

The normal distribution of scale items then would approximate the curve in Figure 2.[11]

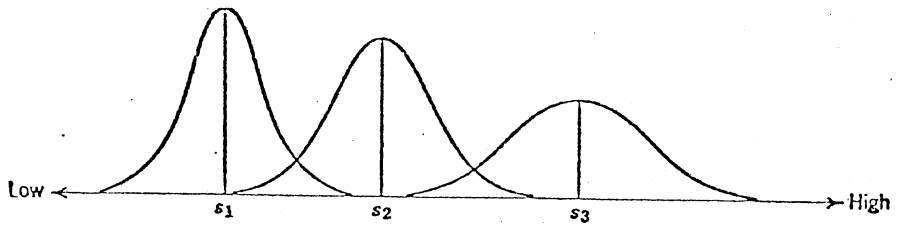


Figure 2. Distribution of Discriminal Differences

Items were subsequently assembled into a questionnaire in the format depicted in Table I.[12] As Goode and Hatt note, each item becomes a scale, and the total scale is in essence a battery of scales.[13]

TABLE I

THURSTONE EQUAL-APPEARING INTERVAL MEASUREMENT ITEM

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
I Believe the Church is the Greatest Institution in America Today					

The very factors which advanced Thurstone's equal-appearing interval research beyond previous works were also grounds in part for his most significant criticism--the scale's values were dependent upon the number and character of the judges selected.[14] However, his work was monumental for three reasons: (1) the development of the method of internal consistency (items were relevant only if they were endorsed in similar fashion to other items with equivalent scale values), (2) the introduction of a verbal measure of intensity on scales (strongly agree, agree, uncertain . . . as above), and (3) the use of equal-intervals, or equidistant gradations of intensity.

Scales have exhibited a proliferation of designs since, but one common feature of virtually all is the attempt to represent gradations between polar opposites in equal fractional increments. The scales in Figures 3 and 4 are examples of several techniques developed by Louis H. Guttman and colleagues.[15] All produce essentially the same results but differ somewhat in mechanics involved.[16]

1. *A Nation of Nations* does a good job of analyzing the ethnic groups in this country.

_____ Strongly agree 4 _____ Agree 3 _____ Undecided 2
_____ Disagree 1 _____ Strongly disagree 0

2. On the whole, *A Nation of Nations* is not as good as most college textbooks.

_____ Strongly agree 0 _____ Agree 1 _____ Undecided 2
_____ Disagree 3 _____ Strongly disagree 4

3. Adamic organizes and presents his material very well.

_____ Strongly agree 4 _____ Agree 3 _____ Undecided 2
_____ Disagree 1 _____ Strongly disagree 0

4. As a sociological treatise, Adamic's book does not rate very high.

_____ Strongly agree 0 _____ Agree 1 _____ Undecided 2
_____ Disagree 3 _____ Strongly disagree 4

5. Adamic does not discuss any one group in sufficient detail so that a student can obtain a real insight into problems of ethnic group relations in this country.

_____ Strongly agree 0 _____ Agree 1 _____ Undecided 2
_____ Disagree 3 _____ Strongly disagree 4

6. By providing a panorama of various groups, *A Nation of Nations* lets the student get a good perspective on ethnic group relations in this country.

_____ Strongly agree 4 _____ Agree 3 _____ Undecided 2
_____ Disagree 1 _____ Strongly disagree 0

7. *A Nation of Nations* is good enough to be kept as a textbook for this course.

_____ Strongly agree 4 _____ Agree 3 _____ Undecided 2
_____ Disagree 1 _____ Strongly disagree 0

Figure 3. The "Cornell" Technique for Scale and Intensity Analysis

I would go out on a date with a Negro.

Strongly agree	Agree	Undecided	Disagree	Strongly disagree
4	3	2	1	0

When possible, I would avoid sitting next to a Negro.

Strongly agree	Agree	Undecided	Disagree	Strongly disagree
0	1	2	3	4

Figure 4. "Cornell" Technique (Variation)

Apparent of any rating scale is its two-part composition: (1) an instruction, to orient the respondent to its subject and identify the property of the continuum, and (2) a scale, defining the points of gradation. The latter may be accomplished verbally or by graphics.[17]

"'Now I am going to ask you how you would judge a number of occupations. For example, a *railroad brakeman*—which statement on this card (HAND RESPONDENT [rating] CARD) *best gives your own* personal opinion of the *general standing* of a *railroad brakeman*? (PAUSE) What number on that card would you pick out for him?' (RECORD ANSWER.)

"'Try not to judge a job according to your own opinion of some one person you know who has such a job. Now, how would you judge a . . .?' (PROCEED THROUGH LIST OF OCCUPATIONS.)

"The rating card handed the respondent is reproduced below:

"'For each job mentioned, please pick out the statement that best gives *your own personal opinion* of the *general standing* that such a job has.

1. *Excellent* standing.
2. *Good* standing.
3. *Average* standing.
4. *Somewhat below average* standing.
5. *Poor* standing.

X. I don't know where to place that one.'"

Figure 5. The Descriptive Technique

Most conspicuous in the technique represented in Figures 5 and 6 is its weakness: It presumes an equal interval between description

gradations, and it limits respondent discrimination to such.[18]

THE OHIO
SOCIAL ACCEPTANCE SCALE
For the Intermediate Grades

Form C-2

Issued by
OHIO SCHOLARSHIP TESTS and DIVISION of ELEMENTARY SUPERVISION
STATE DEPARTMENT OF EDUCATION
COLUMBUS, OHIO

Prepared by
The Euclid Elementary Teachers
in Cooperation with
The College of Education, The Ohio State University

DIRECTIONS: On a separate sheet you will find the name of every student in your class. We want you to put a number in front of every name. The number you put down should be the number of one of the following paragraphs.

"My very, very best friends." 1 I would like to have this person as one of my very, very best friends. I would like to spend a lot of time with this person and would enjoy going places with this person. I would tell some of my troubles and some of my secrets to this person and would do everything I could to help this person out of trouble. I will give a NUMBER ONE to my very, very best friends.

* * * * *

"My other friends." 2 I would enjoy working and being with this person. I would invite this person to a party, and would enjoy going on picnics with this person and our friends. I would like to talk and make and do things with this person. I would like to work with this person and I would like to be with this person often. I want this person to be one of my friends. I will give a NUMBER TWO to every person who is my friend.

* * * * *

"Not friends, but Okay." 3 I would be willing to be on a committee with this person or to be in the same club. It would be all right for this person to be on the same team with me or to live in my neighborhood. I would be in a play with this person. I would just as soon work with this person in school. This person is not one of my friends, but I think this person is all right. I will put a NUMBER THREE in front of the name of every person I think is all right.

* * * * *

"Don't know them." 4 I do not know this person very well. Maybe I would like this person, maybe I wouldn't. I don't know if I would like to be with this person. I will put a NUMBER FOUR in front of the name of every person I don't know very well.

* * * * *

"Don't care for them." 5 I say "hello" whenever I meet this person around school or on the street, but I do not enjoy being with this person. I might spend some time with this person if I didn't have anything else to do, but I would rather be with somebody else. I don't care for this person very much. I will give a NUMBER FIVE to people I don't care for very much.

* * * * *

"Dislike them." 6 I speak to this person only when it is necessary. I do not like to work with this person and would rather not talk to this person. I will give a NUMBER SIX to every person I do not like.

* * * * *

Figure 6. The Descriptive Technique (Variation)

Graphic methods of scale portrayal, such as that in Figure 7, exhibit much more versatility and, properly constructed, have the

ability to at least minimize continuum segmentation.[19]

You are asked to rate your instructor in terms of a number of characteristics. Will you place a check mark (V) at the place on each line which you think best describes his usual manner of teaching?

1. In regard to enthusiasm for his subject, does he appear to be:

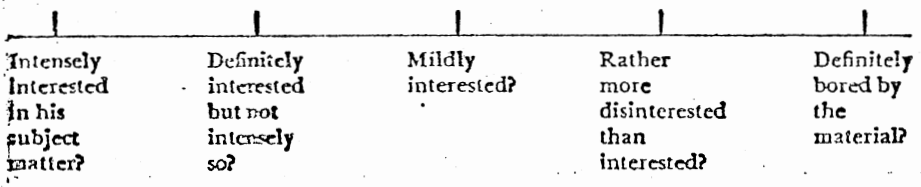


Figure 7. The Graphic Technique

Guilford, et.al., in discussing a scale's discriminatory power, maintained:

> In terms of psychometric theory, the advantage always is with using more rather than fewer steps. This is demonstrated by the numerous studies showing that the reliability of individual rating scales is a monotonically increasing function of the number of steps . . . Essentially the same principle is derivable from another body of evidence, that concerning relations between the number of scale steps and the information (or amount of discrimination) found in classical methods of psychophysical scaling.[20]

With enhanced use of graphic measurement scales came the facility to increase the number of steps--and, thus, discriminability (see Figure 8).[21]

It was not long before advantages of quantification of such scales were recognized.

> The numerical results provided by standardized measure have two advantages. First, numerical indices make it possible to report results in finer detail than would be the case with personal judgments. Thus the availability of thermometers makes it possible to report the exact increase in temperature

. . . rather than only that 'the temperature increase(d).' A second advantage of quantification is that it permits the use of powerful methods of mathematical analysis. This is essential in the elaboration of theories and in the analysis of experiments. Although it may be a long time off for psychology, it is reasonable to believe that all theories eventually will be expressed in mathematical form.[22]

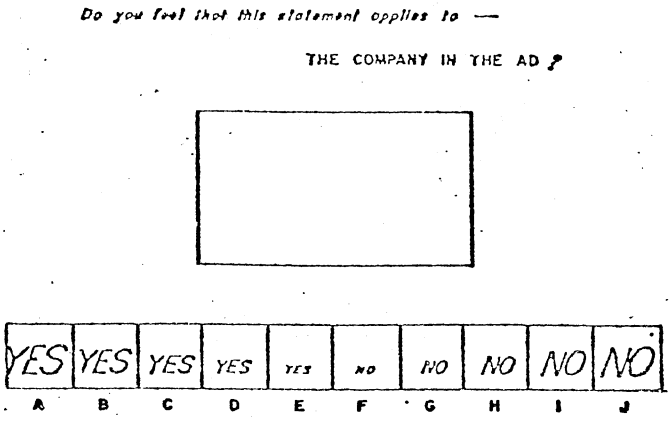


Figure 8. Scalar Device for Measuring Attitudes Toward a Company

Korzybski's contention (see Introduction) of the need for a "precise" language--such as mathematics--in the behavioral sciences was being incorporated in attitudinal measurement. An inexorable pattern of scalar construction began to evolve (see Figures 9 and 10).[23]

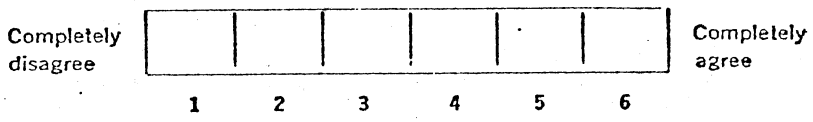


Figure 9. "Boxed" Graphic/Numerical

Figure 10. Graphic/Numerical

With Osgood's introduction of the semantic differential (S.D.) in the 1940's, graphic development of the S.D. measurement device itself (with minor modifications) largely has remained static. Since his landmark study (The Measurement of Meaning) was published in 1957, the conformative style of the Osgoodian scale has tended to exhibit a continuum bounded by polar opposites and segmented to denote linear gradation of intensity throughout its length.[24]

In conducting a study of social stereotypes with a method inspired by Warren,[25] Stagner and Osgood found, through factor analysis, a large grouping of descriptive scales that tended to resolve into three highly intercorrelated clusters.[26] These subsequently were termed evaluative, potency, and activity dimensions of meaning (see Introduction). The term "semantic differential" has since come to refer not to one but a battery of scales which, taken together, comprehensively measure affective judgment in all three areas, as shown in Figure 11.[27]

> Although it will be convenient here to speak of the semantic differential, the term is used in a generic sense to refer to (a) collection of rating scales anchored by bipolar adjectives. Rather than the semantic differential being a particular instrument (or test, as some have called it), it is a very flexible approach to obtaining measures of attitudes and other sentiments. The flexibility of the approach is one of its appealing features. The object that is rated is referred to as a concept, and anything that can be named can be rated, e.g., Winston Churchill, peach ice cream, labor unions, birth control, my best friend, and automobiles.[28]

SCHOOL

(*E*)	1. pleasant	:	:	:	:	:	:	unpleasant
(*A*)	2. angular	:	:	:	:	:	:	rounded
(*A*)	*3. passive	:	:	:	:	:	:	active
(*E*)	*4. ugly	:	:	:	:	:	:	beautiful
(*P*)	*5. delicate	:	:	:	:	:	:	rugged
(*A*)	6. fast	:	:	:	:	:	:	slow
(*E*)	7. good	:	:	:	:	:	:	bad
(*P*)	*8. weak	:	:	:	:	:	:	strong
(*A*)	*9. dull	:	:	:	:	:	:	sharp
(*P*)	10. deep	:	:	:	:	:	:	shallow
(*P*)	11. heavy	:	:	:	:	:	:	light
(*E*)	*12. dark	:	:	:	:	:	:	bright

(E) = Evaluative, (P) = Potency, (A) = Activity

Figure 11. The Semantic Differential

The Semantic Differential in Practice

The semantic differential, with all its recognition and acclamation, has not been an immaculate conception by any means. Despite its being afforded a proliferation of studies and reviews by psychologists, psycholinguists, semanticists and researchers of other pursuits over the past twenty-five years, there appear to this writer several ominous assumptions by practitioners of the method which must be addressed.

A review of literature lends credence to Osgood's assertion that S.D. technique generally satisfies criteria of measurement in categories of objectivity, reliability, validity, comparability and utility.[29] The conspicuous omission, as the author sees it, is the criterion of sensitivity. Osgood, early on, believed that

> The question of sensitivity of the method comes down to whether it is able to reflect as fine distinctions in meaning as are ordinarily made. We have incidental (sic) evidence that a semantic differential can tease out nuances in meaning . . .[30]

Other impinging factors considered equal, any enhanced refinement in measurement specificity in this criterion can have only a positive effect upon other criteria, and, therefore, the over-all accuracy of the technique. Sensitivity, then, was the core issue in this study.

Although Guilford maintained that discriminative ability of an affective measurement tool tended to increase with augmented incrementation to twenty steps,[31] others have contended, through the principle of summated scales, that such is not necessarily the case.[32] According to the latter theory, the number of steps on a single scale used to measure an attribute would be critical. However, if a battery of scales were utilized (say, a half dozen), all of which measure the same

attribute, six or seven steps would suffice. In this case, the reliability of the summed ratings is directly linked to the correlation among that group of scales.

This would seem to be the case with the semantic differential. It operates as a battery of scales represented by bipolar adjective identifiers--purported through factor analysis to measure three meaning dimensions.

Nonetheless, few analysts have been willing to gamble that this is always the case.[33] Indeed, some--directly addressing the S.D.--have endorsed an increase to a 20-point, or even a 30-point scale.[34] A complete overhaul of the present scale even has been suggested.[35] Scale incrementation thus formed the first of two major considerations in this study.

Korzybski's recognition of the need for a "precise" language--such as mathematics--in the social sciences, and the increased use of quantification in social research information retrieval was addressed previously. One outstanding feature of the semantic differential is its ability to quantify human affective/emotive judgment, and dispose it to the gamut of scientific operations available through mathematical permutation and analysis. Only to that extent will behavioral research findings and conclusions be as reliable as those of exact sciences. As Johnson has said, "The language of science is the better part of the method of science."[36]

It probably can be said safely that, once past the encoding stage, such information is in safe hands . . . at least until the quantitative results are decoded, or "interpreted" into our imprecise, multiordinal verbal language.[37] The challenge of the semantic differential or any

other retrieval system, then, must lie with what might be termed its "encoding efficiency."

The question in this case is: In the S.D.'s role as a tool to measure meaning, are its own variables interpreted the same by respondents and administrators alike? There already has been evidence to the contrary among administrators themselves. Attempts verbally to represent scale gradations in an equitable manner have caused considerable consternation among academicians, as is demonstrated in Figures 12 through 15.

There appear to be several items begging attention here. For instance, who is to say that "mostly X" in the scale of Figure 13 is equivalent to the "quite X" of the scale in Figure 14? Would an inquirer find anywhere near unanimity of agreement that "extremely Y" in the scale of Figure 14 equals "very Y" of the scale in Figure 15? Are "mostly/quite" and "very/extremely," respectively, congruent verbal barometers of intensity? In these two particular examples they are used to denote the same relative points of their respective continuums. One study lists as many as nine different adverbs used to earmark scalar gradation.[38]

Fortunately, there more recently has been a certain standardization to Osgood's battery of "linguistic quantifiers" depicted in Figure 14. This has solved part of the problem. Now it can be asked: Do respondents envision similar points in semantic space for Osgood's conventional linguistic quantifiers (neither/equally, slightly, quite, very/extremely)? This is actually a two-part question; the first being whether respondents interpret them in the order presented. Furthermore, are they seen as falling into the neat equidistant pattern stepped off

in the scale? Order and degree of the linguistic quantifier was the second major point addressed in this study.

The content of many complex linguistic assertions (e.g., "I don't think these Chinese Communists are to be trusted") can be reduced to the allocation of a concept to a scale, e.g.,

CHINESE COMMUNISTS:
trustworthy _:_:_:_:X:_ untrustworthy.

The greater the intensity of particular assertions (e.g., "These Chinese Communists are completely untrustworthy"), the more extreme becomes the allocation toward one or the other of the polar terms.

Source: C. E. Osgood and G. J. Suci, "Factor Analysis of Meaning," Journ. of Exper. Psych. (1955)

Figure 12. Rationale of the Semantic Differential (Early)

Completely disagree ___:___:___:___:___:___ Completely agree

1 2 3 4 5 6

In some instances, however, the numbers are defined and written in spaces opposite the objects to be rated, instead of having the appropriate numbers marked on a graphic scale. It is customary to refer to these as *numerical scales* rather than as graphic scales. The issue, however, usually concerns whether there will be numbers employed with a graphic scale or without a graphic scale. Numbers are used as anchors in most rating scales. The numbers must first be defined, e.g.:

1 Completely disagree

2 Mostly disagree

3 Slightly disagree

4 Slightly agree

5 Mostly agree

6 Completely agree

Source: J. C. Nunnally, Psychometric Theory (New York, 1967).

Figure 13. Semantic Differential with Linguistic Quantifiers (Variation)

The content of many complex linguistic assertions (e.g., "I don't think these Chinese Communists are to be trusted") can be reduced to the allocation of a concept to a scale, e.g.,

CHINESE COMMUNISTS:
trustworthy _:_:_:_:X:_ untrustworthy.

The greater the intensity of particular assertions (e.g., "These Chinese Communists are completely untrustworthy"), the more extreme becomes the allocation toward one or the other of the polar terms.

Source: C. E. Osgood and G. J. Suci, "Factor Analysis of Meaning," Journ. of Exper. Psych. (1955)

Figure 12. Rationale of the Semantic Differential (Early)

Completely disagree ___:___:___:___:___:___ Completely agree

1 2 3 4 5 6

In some instances, however, the numbers are defined and written in spaces opposite the objects to be rated, instead of having the appropriate numbers marked on a graphic scale. It is customary to refer to these as *numerical scales* rather than as graphic scales. The issue, however, usually concerns whether there will be numbers employed with a graphic scale or without a graphic scale. Numbers are used as anchors in most rating scales. The numbers must first be defined, e.g.:

1 Completely disagree

2 Mostly disagree

3 Slightly disagree

4 Slightly agree

5 Mostly agree

6 Completely agree

Source: J. C. Nunnally, Psychometric Theory (New York, 1967).

Figure 13. Semantic Differential with Linguistic Quantifiers (Variation)

Each item (pairing of a specific concept with a specific scale) presents the following situation:

(CONCEPT)

polar term X ____:____:____:____:____:____:____ polar term Y
(1) (2) (3) (4) (5) (6) (7)

in which the scale positions have already been defined for the subject in the instructions (see Chapter 3) as:

(1) *extremely* X
(2) *quite* X
(3) *slightly* X
(4) *neither* X nor Y; *equally* X and Y
(5) *slightly* Y
(6) *quite* Y
(7) *extremely* Y

We shall assume that, on the basis of a great deal of prior experience in encoding, the terms "extremely," "quite," and "slightly" as linguistic quantifiers have been associated with more or less equal *degrees* of intensity of whatever representational process (X or Y) happens to be elicited, and therefore, that the sign combinations "extremely X," "quite X," and so forth will elicit an r_m of the quality X and of the intensity given by the quantifier.

Source: C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, The Measurement of Meaning (Urbana, Ill., 1957).

Figure 14. The Semantic Differential and Attendant Linguistic Quantifiers (Osgoodian Convention)

"The purpose of this study is to measure the meanings of certain words to various people by having them judge each word against a series of descriptive scales. In taking this test, please judge the words on the basis of what they mean to *you*. Each numbered item presents a CONCEPT (such as DICTATOR), and a scale (such as *high–low*). You are to rate the concept on the 7-point scale indicated.

If you felt that the concept was *very closely associated* with one end of the scale, you might place your check mark as follows:

DICTATOR:

up___ : : : : : : X down.

"If you felt that the concept was *quite closely related* to one side of the scale, you might check as follows:

HOUSE:

straight___ : X : : : : : crooked.

"If the concept seemed *only slightly related* to one side as opposed to the other, you might check as follows:

CLOUD:

easy___ : : X : : : : difficult.

"If you considered the scale *completely irrelevant*, or *both sides equally associated*, you would check the middle space on the scale:

TREE:

idealistic___ : : : X : : : realistic.

Source: Osgood and Suci, Journ. of Exper. Psych. (1955).

Figure 15. Semantic Differential with Linguistic Quantifiers (Variation)

ENDNOTES

1. Fred N. Kerlinger, Foundations of Behavioral Research (New York, 1973), p. 426.

2. Jum C. Nunnally, Psychometric Theory (New York, 1967), p. 27.

3. Warren S. Torgerson, Theory and Methods of Scaling (New York, 1967), pp. 14-15.

4. Lyle V. Jones, "The Nature of Measurement," in Educational Measurement, ed. Robert L. Thorndike (Washington, 1971), p. 336.

5. Ibid.

6. Wendell Johnson, People in Quandaries (New York, 1946), pp. 30-31.

7. J. P. Guilford, Psychometric Methods (New York, 1936), p. 264.

8. Ibid., p. 265.

9. Pauline Young, Scientific Social Surveys and Research, (Englewood Cliffs, N.J., 1966), p. 351.

10. L. Thurstone and E. Chave, The Measurement of Attitudes (Chicago, 1929), pp. 61-78.

11. Nunnally, p. 51.

12. William J. Goode and Paul K. Hatt, Methods in Social Research (New York, 1952), p. 272.

13. Ibid.

14. Ibid., p. 256.

Young, p. 352.

15. Louis H. Guttman, "The Cornell Technique for Scale and Intensity Analysis," Educational and Psychological Measurement, Vol. 7, pp. 248-279.

Keith S. Griffiths, "The Construction of a Scale to Measure Attitudes Toward Defined Racial and Religious Groups" (Unpub. M.S. thesis, University of Washington, 1949).

16. Young. p. 362.

17. Goode and Hatt, p. 257.

18. Ibid., pp. 259, 252.

19. Ibid., p. 257.

20. Nunnally, p. 521.

21. Harry W. O'Neill, "Advertising with the Differential Attitude Technique," Journal of Marketing, Vol. 27, p. 21.

22. Nunnally, pp. 5-6.

23. Ibid., pp. 524, 521.

24. Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Ill., 1957).

25. H. C. Warren, Dictionary of Psychology (New York, 1934).

26. Charles E. Osgood, "The Nature and Measurement of Meaning," Psychological Bulletin, Vol. 49, pp. 197-237.

27. Kerlinger, p. 572.

28. Nunnally, p. 535.

29. Osgood, pp. 230-231.

30. Ibid.

31. Nunnally, pp. 521-522.

32. Ibid., pp. 522-523.

33. Ibid.

34. Harold Gulliksen, "How to Make Meaning More Meaningful," Contemporary Psychology, Vol. 3, p. 116.

35. Ibid.

36. Johnson, p. 50.

37. Ibid., pp. 155-158.

38. Norman Cliff, "Adverbs as Multipliers," Psychological Review, Vol. 66, p. 29.

CHAPTER III

METHODOLOGY AND DESIGN

Introduction

If measurement is the means by which mathematics may be applied to a science, it follows that to the extent behavioral research emends its measurement tools, it refines itself as a science.

It has been noted that all sciences begin empirically--as bodies of correlated observation and data--and progress toward the theoretical.[1] This "evolution" of the sciences can be seen in their relative status today. The "exact" sciences such as physics, mathematics and chemistry, which found their origins tens of hundreds of years ago in philosophy, are now relatively far more sophisticated than the behavioral sciences. Thus biology and psychology are yet largely correlational pursuits whereas physics and chemistry are for the most part theoretical.

> It is more than a mere coincidence that the sciences would order themselves in largely the same way if they were classified on the basis of the degree to which satisfactory measurement of their important variables has been achieved. The development of a theoretical science . . . would seem to be virtually impossible unless its variables can be measured adequately. And one of the primary differences between the social and behavioral sciences on the one hand and the physical sciences on the other lies in the procedures used for measuring their important concepts.[2]

In behavioral sciences where there is certainly a wealth of observables but few precise instruments, the great promise of the semantic differential as a controlled association and scaling device to measure

connotative meaning, probably cannot be overemphasized. The task of this work was to add some degree of articularity and specificity--and thus emendment--to the scale's design.

Hypotheses

Segmentation versus Sustentation

Figures 16 and 17 understate the need for enhanced resolution on the S.D. scale. There are a number of analyses that may be run on scores derived from the semantic differential, but in each case computations begin with the assignment of numbers 1 through 7 (more or less depending on incrementation) to scale positions as shown in Figure 16.[3]

Means from the total sampling of scores may then be tabulated. Such a tabulation, replotted over a master grid, takes on a geometric form as depicted in Figure 17. As can be ascertained, group scores rarely fall neatly into the stepped positions of the Osgoodian scale. Rather, they tend to ignore any segregation of space and retrofit into a type of continuum.[4]

This plot, plus Osgood's own report that changes/differences in group means as small as .5 scale units are significant at the .05 level of reliability, appeared to be major indicants that further study in scale design was warranted.[5] Other substantiating material, having evidential bearing on what this writer earlier termed "encoding efficiency" of the S.D. scale, was located. Notable among these were two works. Tukey states:

> Many times scale makers and measurers stop far short of the limit (of measurement discriminability) set by the danger of non-cooperation (from respondents) . . . If more than 10% are exact checks (on test-retest scores), either the scale is too coarse, or the duplicates are not independent.[6]

Bad (1):(2):(3):(4):(5):(6):(7) Good

Note: Either bi-polar adjectives or direction of number sequence must be aligned to the concept rated during tabulation--the direction of weighting being determined by favorableness or unfavorableness of the individual scale.[7]

Figure 16. Analysis of Semantic Differential Data

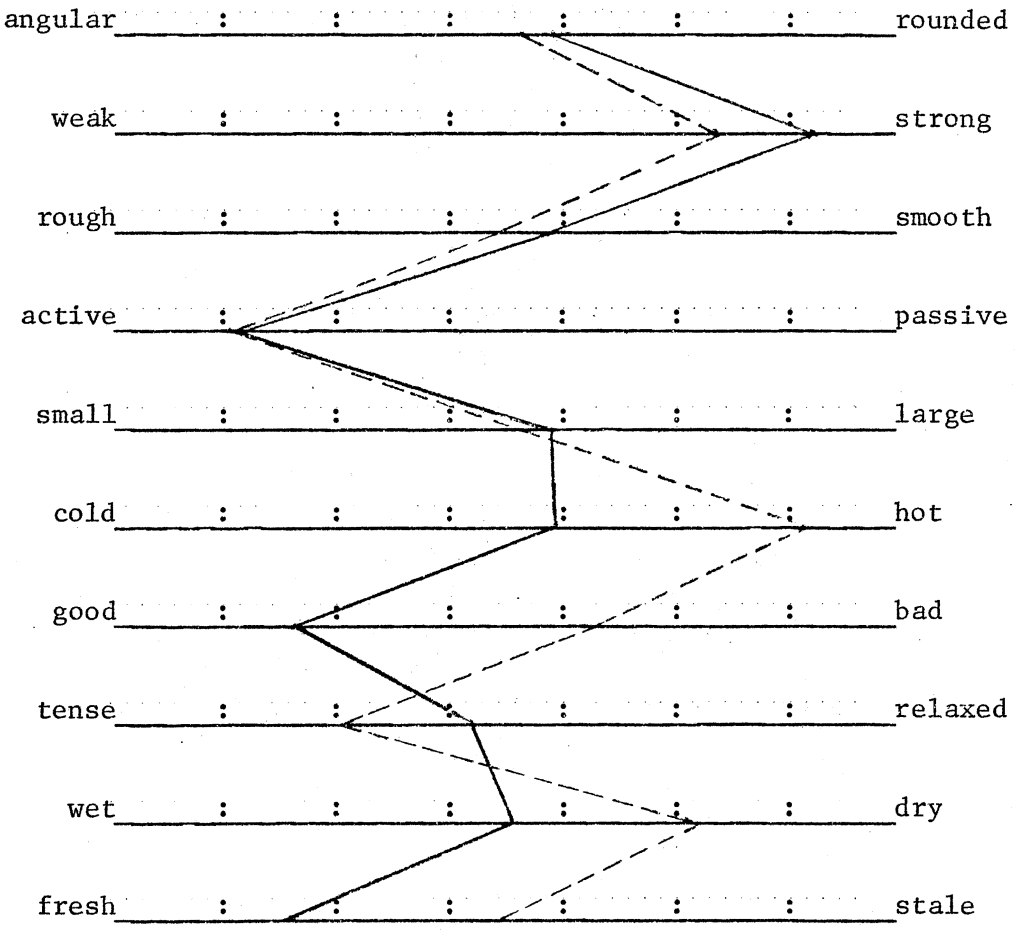


Figure 17. Plotted Medians for Two Groups of 20 Subjects Differentiating the Adjectives "Eager" and "Burning"

Gulliksen applied this to the semantic differential:

> In Osgood's data (The Measurement of Meaning, pp. 129-132), 54.0% of complete agreements indicates extremely coarse grouping for determining the standard error of measurement. In terms of a normal curve this middle interval is at least 1.4 sigma in width. The problem is pointed out by Osgood himself . . . 'On many individual items . . . the variance approaches zero.' Clearly, it is not possible to determine accuracy of measurement when such coarse grouping is used. For any measurement one needs a unit so fine that a reasonable determination of error is possible . . . If more than 10% are exact checks, either the scale is too coarse, or the duplicates are not independent. Even if one sets a standard of 20% or 30% identical re-measurements, instead of only 10%, it is still clear that the 7-category scale results in far too coarse grouping.[8]

Linguistic Quantifiers Questioned

This rationale was carried a step further. During the course of literature review, a definite evolutionary trend toward scales exhibiting variations of a continuum was indicated. If scalar segmentation was hypothesized to be a culprit of design bias, it would be reasonable to also regard as suspect any system of insular emphasis within the semantic scale space. This practice is obviously the norm in the conventional 7-point S.D. scale, where subjects are mandated to respond to segregated space according to the legend (1) very/extremely X, (2) quite X, (3) slightly X, (4) neither X nor Y; equally X and Y, (5) slightly Y, (6) quite Y, (7) very/extremely Y.

Osgood (1957) says he assumes the linguistic quantifiers are associated with more or less equal degrees of intensity of the quality of X or Y elicited.[9] Furthermore, he maintains that these quantifiers, radiating in both directions from a central origin, do yield nearly equal psychological measurement units in the judgment process.[10]

The question posed at this point was of two parts. First, are

Osgood's linguistic quantifiers actually perceived in the order of intensity he states? For instance, is it a justifiable presumption that "very" and "extremely" are semantically interchangeable affective terms? Is "quite" always viewed as falling above "slightly" and below "very" or "extremely" on a gradient of intensity? Second, do these linguistic quantifiers represent to the respondent the precise points and intervals Osgood has so delineated on the S.D. scale?

These questions paved the way for setting down logic around which subsequent work in this study revolved. The first hypothesis actually dealt with two points. It tested respondent weightings of Osgood's adverbial measures of intensity, and compatibility of those ratings with the 7-step equidistant scale. The second hypothesis served as a check on Osgood's assumption that certain adverbs are interchangeable as intensity measures. This postulation, while not a direct odds with Osgood's second hypothesis regarding equivalent experimental continua, would nonetheless probably justify a modification in subject application if vindicated.[11]

The hypotheses were:

1. Allocation of perceived linguistic quantifier midpoints to S.D. scales of increased resolution will result in significant variation from the Osgoodian equal-interval standard.

2. The adverbs "very/extremely" and "equally/neither", respectively, represent significantly different points on unprescripted scales.

The Questionnaire

Osgood himself has recognized,

> An instrument is sensitive to the degree that it renders discriminations commensurate with the natural units of the

> material being studied; ideally it should yield distinctions as fine, or even finer, than those made on common sense grounds. Sensitivity thus implies both reliability and validity.[12]

Why practitioners haven't applied this ascription to the design of the S.D. scale is, to this writer, an anomaly. Says Nunnally:

> Any (student) . . . will learn . . . that the reliability estimate obtained in any particular study is independent of the number of persons in the study, but in any study the reliability is directly related to the number of items on the test.[13]

If a survey's reliability can be increased by increasing its items, why shouldn't item sensitivity be increased by increasing its measurement resolution?

With this notion in mind, work was initiated on an instrument design which could capture the information that would be needed in test of the hypotheses set down. A list of considerations to account in composition of the device was assembled:

1. Number of Steps or Points
2. Scale Segmentation
3. Nature and Size of the Continuum
4. Inclusion of Graphic or Numerical Anchors
5. Orientation of Scale
6. Instructions/Descriptive Phrases
7. Adaptability of Retrieved Data to Quantification and Analysis

Literature was reviewed for counsel and guidance on decisions made within each of these areas. Not surprisingly, however, conflicting disclosures were found due to the myriad of designs devised over the past fifty to seventy-five years. In these cases, as in all, methods appearing most acclimated to the present problem were adopted. Several questionnaire designs were critiqued through written and open discussion

in a communication research designs class instructed by Dr. Walter J. Ward at Oklahoma State University in the spring of 1978. A number of suggestions were adopted, and are so cited in certain sections below.

Number of Steps or Points

The gravity of the entire study probably rested more with this than any other single consideration pertaining to questionnaire design. The challenge was to meet all criteria of enhanced measurement sensitivity parameters desired of the S.D. scale itself. If too few steps were utilized, scale efficiency would be tempered due to coarseness. On the other hand, there would be no point in resolving gradations so fine that the scale extended beyond discriminative power of the central nervous system.[14] Nunnally reported another impinging factor to be considered is fatigue.

> The only exception to the rule that reliability increases with the number of scale steps would occur in instances where a large number of steps confused subjects or irritated them to the point where they became careless. Then it would be possible to find the reliability coming back down with, say, as many as 20 steps.[15]

Nevertheless, of paramount importance was the use of a high-resolution device, if the intended purpose of critiquing the 7-step S.D. scale and attendant linguistic quantifiers was to be effected. Early, the notion was seized to utilize a straight-line, relatively unencumbered continuum. Such a design would make possible an infinite number of geometric points while not fatiguing the respondent into a state of carelessness.

This is by no means an entirely new model. Goode and Hatt report use of a similar graphic continuum in which the respondent placed a check mark at a point on the line corresponding to his attitude on a

subject (see p. 18). Related devices have been cited.[16] In all those cases, however, there were numerical or descriptive weightings attendant to the scales.

As a cross-check on the design, a separate section utilizing a 100 percent scale (numbers themselves constitute the continuum) was administered.[17] This system seemed nearest, as numerical scales go, to the fluid concept strived for in the graphic scale.

Scale Segmentation

The prime consideration here was inclusion of graphic reference points. Consistent with rationale above, no standard geometric/equidistant gradation marks were utilized. Thought was given to use of two specific references: (1) a midpoint hash to divide the positive and negative portions of the rule,[18] and (2) an orientation hash-mark linked to a given linguistic quantifier position (this position identified by Osgoodian convention) from which reference the respondent supplied the position of a neighboring quantifier as he so conceived it.

A number of "pros" and "cons" were assembled for each, but the overriding view of the research design class--because the principal objective of this study was to decrease design bias of the S.D. scale--was that any form of reference mark constituted some measure of bias in and of itself. As such, any and all should be dropped in favor of the unencumbered straight-line delineated above.

Nature and Size of the Continuum

Since the "unencumbered straight-line" scale concept was settled, the principal consideration here was physical length of the scale. A

number of researchers have endorsed Guilford's five-inch scale, and the dimension was adopted for this application.[19]

Inclusion of Graphic or Numerical Anchors

The rationale of the first two sections (above) was carried over to this area also. The suspicion that any method of weighting or segmentation adversely could sway respondent replies was judged more critical by the designs class and author than fear of insufficient parley of the scale to a respondent (as discussed further below).

Orientation of Scale

The foremost consideration here was what quality or value would be affixed to, and represented by, bipolar points of the scale. To resolve this, Osgood's Table of Rotated Factor Loadings was consulted, and the adjective pair "good-bad" was chosen because of its high evaluational factor loading.[20]

Since the particular adjective orientors were of little significance in this work, compared to the study of their modifying adverbs (linguistic quantifiers), it was decided that one pair could be used for the entire battery of scales. There were several advantages to this route, including lessened likelihood of respondent confusion or an adverb/adjective multiplicative effect as has been hypothesized by some researchers.[21]

Finally there was the question of whether to alternate "good" and "poor" ends of the scale to prevent what has been termed "response bias tendency."[22] However, this phenomenon tends most often to present itself in association with the "halo effect,"[23] and was not felt to be a

significant factor here, since neither attitudes nor personalities were involved.[24] On the other hand, possibility of respondent confusion or oversight of the scale reversals loomed as a substantial possibility. Accordingly, all scales were graduated left to right, "bad" to "good."

Instructions/Descriptive Phrases

To properly acquaint respondents with the questionnaire and orient them so that accurate, representative information was received was vitally important, for two reasons. First, the specific design of the scales was, if not unique, most likely one to which respondents were not accustomed. Second, (in attempt to minimize or eliminate all biasing factors) the absence of numerical, verbal, or graphic weighting/quantifying labels could leave respondents baffled. Thus, considerable care was devoted to composition of the pretest instructions.

Several early ideas and suggestions were rigorously adhered to. Use of the word "test" was avoided throughout the questionnaire, since this term has been known to not only evoke tension, but a range of Pavlovian responses in various human as well as infra-human subjects.

Brevity was desired and subscribed to as much as possible. The rough draft of these instructions was submitted to the research design class, with a further reduction of nearly half its former size the result.

Instructions comprised four parts: (1) an orientation, or background, on the subject being researched, (2) a statement of purpose of the work, (3) the situation, or subject being addressed in the questionnaire, and (4) three examples to acclimate respondents. In all three examples, the adjective pairs (bipolar terms) and adverbs (linguistic

quantifiers) were unlike those utilized in the body of the questionnaire. These precautions, hopefully, minimized possibility of pre-trial bias.

Adaptability of Retrieved Data to Quantification/ Analysis

This area presented no special problems or obstacles. Some scientists have shied away from unlabeled graphic scales, specifically because they do not facilitate analysis of data.[25] Sufficient space already has been devoted to justification for adoption of the device; therefore, this section only addresses the way in which scales were quantified and data extracted.

Since an unsegmented continuum of any dimension implies an infinite number of geometric points, the only physically-limiting factor in this type scale's resolution is the ratio of the width of the hash-mark made by a given respondent's pen or pencil to the actual length of the scale. It was determined that the smallest practical unit of measure with this system is 1/32nd of an inch. Since the conventional Osgoodian S.D. scale has seven points, theoretically this method offers an increased resolution factor of about 22 times (5 ÷ 1/32 ÷ 7).

To quantify scales, then, it was only necessary to subject them individually to a ruler. (Scale quantification was expedited with an overlay grid segregated into base units of 32nds of an inch.)

The backup "100 percent" scale presented no quantification problems, since the number submitted by the respondent constituted the value used for analysis. (The survey instrument is reprinted at full length in Appendix A.)

Survey Procedure

One-hundred questionnaires were printed and distributed, with a sufficient number of backup surveys for non-returns. A numbering system was used in conjunction with respondent identification to eliminate names and extraneous marks from the questionnaire proper.

Since this was not a survey of attitudes or opinions, a random sampling, per se, was not felt necessary.[26] Nonetheless, a heterogeneous sample was sought in age, occupation, ethnicity and geographic location. Geographic distribution of the survey ranged from Stillwater, Oklahoma, to Catania, Sicily. Ethnic backgrounds varied from Anglo-American to Chinese (all were English-speaking respondents, however). Ages of respondents differed widely, and occupational pursuits ranged from receptionists to research scientists. (A complete tabulation of respondents is posted in Appendix B.)

To minimize any effects of test sensitization, respondents were requested not to return to previously-marked scales to change responses. Subjects were directed to use ink pens as a further measure of enforcing this effort.

No time limit was established for questionnaire completion. The typical response time fell between ten and fifteen minutes.

ENDNOTES

1. Lyle V. Jones, "The Nature of Measurement," Educational Measurement (Washington, 1971), p. 335.

2. Warren S. Torgerson, Theory and Methods of Scaling (New York, 1967), p. 2.

3. Fred N. Kerlinger, Foundations of Behavioral Research (New York, 1973), p. 572.

4. Charles E. Osgood, "The Nature and Measurement of Meaning," Psychological Bulletin, Vol. 49, p. 229.

5. _____, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Ill., 1957), p. 329.

6. John W. Tukey, "Symposium on Statistics for the Clinician," Journal of Clinical Psychology, Vol. 6, p. 64.

7. Pauline Young, Scientific Social Surveys and Research (Englewood Cliffs, N.J., 1966), p. 359.

8. Harold Gulliksen, "How to Make Meaning More Meaningful," Contemporary Psychology, Vol. 3, p. 116.

9. Osgood, Suci, and Tannenbaum, p. 29.

10. Ibid., p. 327.

11. Osgood, p. 227.

12. Osgood, Suci, and Tannenbaum, p. 166.

13. Jum C. Nunnally, Psychometric Theory (New York, 1967), p. 10.

14. J. P. Guilford, Psychometric Methods (New York, 1936), p. 267.

15. Nunnally, p. 521.

16. Guilford, p. 271.

17. William J. Goode and Paul K. Hatt, Methods in Social Research (New York, 1952), p. 258.

18. Guilford, p. 271.

19. Ibid.

Goode and Hatt, p. 257.

20. Charles E. Osgood and George J. Suci, "Factor Analysis of Meaning," Journal of Experimental Psychology, Vol. 50, p. 331.

21. Norman Cliff, "Adverbs as Multipliers," Psychological Review, Vol. 66, pp. 27-44.

22. Kerlinger, pp. 572-573.

23. Ibid., p. 548.

24. Ibid.

25. Nunnally, p. 526.

26. Kerlinger, pp. 118-120.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

Data were treated in three segments: (1) absolute values assigned by respondents to linguistic quantifiers, (2) midpoint values of linguistic quantifiers, as perceived by respondents, and (3) a comparison of absolute values assigned to linguistic quantifiers on a 5-inch, compared to a 100-percent scale.

How much distance on the 5-inch experimental scale did respondents allot each of the six linguistic quantifiers? In the first major analysis, distance was measured in 32nds of an inch. Each 22/32nds of an inch converted to one meaning unit on the Osgood 7-point scale. (This was the case in all analyses of the 5-inch scale.) In other words, responses in this study showed the number of Osgood unit values assigned to varying degrees of good and bad.

The author compared respondents' average unit values to Osgood's theoretical unit values for each linguistic quantifier of good and bad. For example, if respondents assigned an average unit value of 1.025 to "quite" bad, did this differ significantly from Osgood's theoretical value of 2.00?

Further analysis in section one sought to determine if corresponding quantifiers were assigned similar unit values. That is, were "quite" good and "quite" bad placed about the same distance from the respective opposite ends of the scale by the average respondent, as

assumed by Osgood?

A major concern in this first of the three-part analysis was Osgood's assumption of equal interval scaling; i.e., one unit of meaning space between each contiguous pair of quantifiers. If respondents assigned a unit distance of, say, 1.759 between "slightly" and "quite" bad, was this a significant departure from Osgood's theoretical distance of 1.00? A further check on the equal interval assumption came with comparison of observed and theoretical distances between each possible pair of linguistic quantifiers. Osgood, for example, assumes that four units separate "very" good and "slightly" bad. By how many units did the respondents separate this pair of quantifiers?

The second major analysis centered on midpoints of Osgood's scale units, rather than on whole- or absolute-unit values. Concern here was in how closely respondents' perceptions of linguistic quantifier absolute values were to their perceptions of midpoint values.

The final analysis dealt with the feasibility of using a 7-unit and 100-unit scale interchangeably. Would, for example, a particular linguistic quantifier be assigned about the same relative space on a 7- as a 100-point (or percentage) scale?

Absolute Unit Distances

Between Bipolar Adjectives and Linguistic Quantifiers

Given the bipolar and presumably linear continuum of the good-bad evaluative dimension, and assuming maximum intensity of meaning in the opposing adjectives, respondents were asked to designate degrees as described by adverbs serving as linguistic quantifiers.

For example, if the "bad" end represented the maximum of "bad," where on the 5-inch continuum would a respondent mark "slightly" bad? Such designations were made with a vertical hash mark along the solid-line continuum. Distances were reported in 32nds of an inch. Each 22/32nds of an inch on the author's 5-inch continuum represented one meaning unit on the Osgood scale.

Linguistic quantifiers of good-bad were: "extremely," "very," "quite," "slightly," "neither," and "equally." The first four adverbs were applied to both sides of the continuum.

A treatments-by-subjects variance analysis showed significant differences between all descending adjacent pairs of linguistic quantifiers; in other words, between adjacent intensities of good-bad ($F=909.86$, $df=14/1386$, $p<.001$). Critical difference for pairs of adverbs was .2613, $p<.01$, $df=99$. Table II shows observed Osgoodian unit values which the average respondent assigned the varying degrees of good and/or bad.

Noteworthy observations from Table II involve the two extreme degrees and the middle degrees. Osgood, in his interpretation scale, considers "extremely" as the most positive (and negative) degree, followed by "quite." Yet, in his instructions to respondents, he suggests the highest value be given on basis of a concept having "very" much of the adjective quality. Table II indicates that "extremely" and "very" comprise significantly different degrees of intensity in the minds of the 100 study respondents (6.726 vs. 6.213 and .688 vs. .291, both $p<.01$, $df=99$).

The study did support Osgood in his equating of "neither" and "equally." No significant unit differences occurred between the two

nor were there differences between the repeated unit differences assigned to "neither" and "equally." The latter point implies, too, that respondents were consistent in perception of distances described by each of the midpoint linguistic quantifiers. Response consistency also was indicated on the reported measurements of "slightly" and "quite" bad.

TABLE II

OSGOODIAN UNIT VALUES ASSIGNED TO VARIOUS DEGREES OF GOOD AND BAD BY THE AVERAGE RESPONDENT

Adverbial Degrees	Osgoodian Units Assigned by Respondent
Extremely Good	6.726
Very Good	6.213
Quite Good	5.659
Slightly Good	4.149
Neither (Good nor Bad)	3.550
Neither (Bad nor Good)	3.573
Equally (Good and Bad)	3.501
Equally (Bad and Good)	3.547
Slightly Bad	2.713
Slightly Bad	2.784
Quite Bad	.949
Quite Bad	1.100
Very Bad	.688
Extremely Bad	.291

Theoretical vs. Respondent Unit Values for Linguistic Quantifiers

Though the respondents' assigned values for adjacent quantifiers

conform to Osgood's theoretical hierarchy, the mean assigned value differed significantly from that used by Osgood, as shown in Table II ($t=6.29$, $p<.01$, $df=9$).

All of the differences in the righthand column of Table III reflect lesser unit values than those prescribed by Osgood. Even the "extremely" quantifier drew from nearly one-third to seven-tenths lesser unit values from respondents than from Osgood. The five largest departures from Osgood's unit values were: Quite Bad, .975; Slightly Good, .851; Very Good, .787; and Extremely Bad, .709 units.

TABLE III

COMPARISON OF OSGOODIAN AND RESPONDENT UNIT VALUES OF LINGUISTIC QUANTIFIERS ALONG THE GOOD-BAD CONTINUUM

Linguistic Quantifier	Osgood's Unit Value	Respondent Unit Value	Differences
Extremely Good	7.00	6.726	.274
Very Good	7.00	6.213	.787
Quite Good	6.00	5.659	.341
Slightly Good	5.00	4.149	.851
Neither Good nor Bad	4.00	3.562	.438
Equally Good and Bad	4.00	3.524	.216
Slightly Bad	3.00	2.784	.975
Quite Bad	2.00	1.025	.312
Very Bad	1.00	.688	.709
Extremely Bad	1.00	.291	.476

Note: Identical values were given to the "very" and "extremely" degrees, and to the "neither" and "equally" degrees under the Osgood Unit-Value column, because he uses them interchangeably, as mentioned earlier.

To this point, the average respondent's perception of relative unit values of linguistic quantifiers is consistent with Osgood's, with the exception that "very" earns a lesser value than "extremely." The four quantifiers on which repeated measurements were taken seem to be reliable, in that respondents assigned insignificantly different values to them. This, again, indicated that perceptions of unit values of quantifiers were consistent.

Corresponding linguistic quantifiers were perceived to connote less intense degrees of "good," but more intense degrees of "bad." The differential was most evident with the corresponding "quite" quantifiers. "Quite" good was rated .314 less good than theorized, while "quite" bad was rated nearly one unit lower (.975). So, the value assigned to "quite" differed more from the theoretical on the "bad" end of the continuum than on the "good."

"Very," on the other hand, departed most from Osgood on the positive side. It was rated .787 units "less good," but only .312 units "more bad." Conversely, "extremely" good was only .274 units less good than claimed, while "extremely" bad was viewed .709 units "more bad." "Slightly" deviated most from Osgood on both ends of the continuum. It was rated .851 units less positive than claimed, but 1.216 units more negative.

The above data suggest that "bad" connotatively may be more negative, while "good" is connotatively less positive. And the degree of "more" and "less" varies. "Extremely," "quite," and "slightly" deviate less from Osgood on the positive than the negative side, while "quite" deviates more on the negative than the positive.

The greater-than-expected negative connotation of "bad" was most

evident when quantified by "slightly," "quite," and "extremely," respectively. The less-than-expected positive connotation of "good" came mostly from the quantifiers "slightly" and "very," in that order.

Thus far, Osgood's most over-estimated unit values, compared with respondents' perceptions, were on "very" good and "slightly" good. Most underestimated in intensity were "slightly," "quite," and "extremely" bad.

At this point, the question arises about the comparative quantitative meaning intensity of the presumed bipolar "good" and "bad." Perhaps "bad" more nearly approaches zero in raters' minds, while "good" falls relatively short in connoting the ultimate positive. At any rate, the relative distance between adjacent linguistic quantifiers appears related to which side of the continuum serves as referent. The same could be said for intensity of meaning elicited by corresponding quantifiers.

A scan of the "respondent-unit-value" column in Table III indicates respondents adhere more to unit midpoints than to Osgood's prescribed whole-unit values. This matter is discussed in more detail on page 57.

Observed vs. Theoretical Distances Between Adjacent Pairs of Linguistic Quantifiers

On a solid line representing a single good-bad continuum, each of Osgood's linguistic quantifiers is allotted an equal unit of space, totaling seven in all. In other words, one unit of space would lie between each pair of contiguous quantifiers. But in Table IV, about half the perceived unit distances exceed one unit, while the remainder span less than one unit.[1]

Distances between "quite" and "slightly" bad, for example, were perceived as 1.759 units--.759 more than theorized. More than half a unit (.513) separated "extremely" and "very" good, while .397 units were perceived between "extremely" and "very" bad. Since all the above differences exceeded chance, doubt is cast on whether "extremely" and "very" carry equal intensity of meaning or that "slightly" and "quite" elicit only one unit distance in meaning.

On the other hand, distance between "slightly" and the contiguous midpoints of "neither" and "equally" may have been over-estimated to date. Only about three-fourths a unit was perceived by respondents to separate "slightly" from "neither" and "equally" bad (.778 and .740, respectively). "Slightly" good stands only a respective .625 and .587 units from "equally" and "neither."

The one-unit expected difference between "extremely" and "quite" good was supported (1.000 vs. 1.067, $p<.05$), but respondents' perceived difference between "extremely" and "quite" bad was only .743 units--.266 units less than expected. Meaning gap between "very" and "quite" bad was perceived to be only a little more than one-third unit (.337), or .663 units short of the theoretical. And respondents saw the "very"-to-"quite" good gap as .554 units--.446 units below the expected.

These findings lend further support to the previous implication that "very" probably quantifies somewhere between "extremely" and "quite"; therefore, it should not be used interchangeably with "extremely" in Osgood's instructions.

The midpoint quantifiers of "neither" and "equally" were seen as interchangeable by respondents. The .038 unit distance was not significant.

TABLE IV

NUMBER OF MEANING UNITS BETWEEN PAIRS OF LINGUISTIC QUANTIFIERS ASSIGNED BY OSGOOD, COMPARED WITH CORRESPONDING DISTANCES ASSIGNED BY STUDY RESPONDENTS

Pairs of Linguistic Quantifiers	Osgood's Theoretical Distances	Observed Distances	Differences
Slightly from Quite Bad	1.000	1.759	.759
Extremely from Very Good	.000	.513	.513
Slightly from Quite Good	1.000	1.510	.510
Extremely from Very Bad	.000	.397	.397
Extremely from Quite Good	1.000	1.067	.067
Neither from Equally	.000	.038	.038
Slightly Bad from Neither	1.000	.778	-.222
Slightly Bad from Equally	1.000	.740	-.260
Extremely from Quite Bad	1.000	.734	-.266
Slightly Good from Equally	1.000	.625	-.375
Slightly Good from Neither	1.000	.587	-.413
Very from Quite Good	1.000	.554	-.446
Very Bad from Quite Bad	1.000	.337	-.663

On the positive side of the good-bad continuum; i.e., from the midpoint to "extremely" good, there was an average departure of .387 units from the theoretical, with more than half being due to over-estimation of the theoretical distance between adjacent pairs. Osgood assigns larger meaning distances than did the respondents between "slightly" good and the midpoints, and between "very" and "quite" good. Underestimated theoretical distances were most notable between the contiguous pairs of "extremely" and "very," and "slightly" and "quite" good.

From the midpoint to "extremely" bad, respondents' average deviation from Osgood's theoretical distances between adjacent linguistic quantifiers was .856 units, considerably greater than the average deviation on the positive end of the scale. Furthermore, nearly three-fourths the deviation from Osgood on the scale's negative side was due to greater distances assigned by respondents than by Osgood. Respondents saw notably larger meaning distance between "quite" and "slightly" bad, as well as between "extremely" and "very" bad. On the other hand, they perceived lesser meaning space between "slightly" bad and the midpoint, between "extremely" and "quite" bad and between "very" and "quite" bad.

Observed vs. Theoretical Distances Between All Pairs of Linguistic Quantifiers

More insight into Osgood's claim of equal interval scale points is offered in Table V. Theoretical unit distances between each possible pair of linguistic quantifiers is compared with the corresponding distance assigned by the average respondent.

Table V shows that the relative magnitude of theoretical distance

TABLE V

COMPARISON OF THEORETICAL AND PERCEIVED DISTANCES
BETWEEN 36 PAIRS OF LINGUISTIC QUANTIFIERS

Pairs of Linguistic Quantifiers	Theoretical Distance	Perceived Distance	Difference
Very Good-Slightly Bad	4.00	2.83	-1.17
Slightly Good-Very Bad	4.00	3.07	- .93
Very Good-Extremely Bad	6.00	5.17	- .83
Slightly Good-Slightly Bad	2.00	2.83	.83
Extremely Good-Slightly Bad	4.00	3.21	- .79
Extremely Good-Very Bad	6.00	5.21	- .79
Very Good-Equally	3.00	2.30	- .70
Quite Bad-Slightly Bad	1.00	1.61	.61
Quite Good-Slightly Bad	3.00	2.41	- .59
Slightly Good-Extremely Bad	4.00	3.41	- .59
Quite Good-Very Bad	5.00	4.41	- .59
Very Good-Quite Good	1.00	.42	- .58
Very Good-Quite Bad	5.00	4.44	- .56
Very Bad-Equally	3.00	2.53	- .47
Slightly Bad-Equally	1.00	.53	- .47
Slightly Good-Equally	1.00	.54	- .46
Extremely Good-Extremely Bad	6.00	5.55	- .45
Very Bad-Quite Bad	1.00	.39	- .39
Extremely Good-Very Good	0.00	.38	.38
Extremely Bad-Very Bad	0.00	.34	.34
Quite Good-Slightly Good	1.00	1.34	.34
Extremely Bad-Slightly Bad	2.00	2.34	.34
Extremely Good-Equally	3.00	2.68	- .32
Slightly Good-Quite Bad	3.00	2.68	- .32
Extremely Bad-Quite Bad	1.00	.73	- .27
Quite Good-Extremely Bad	5.00	4.75	- .25
Very Good-Slightly Good	2.00	1.76	- .24
Extremely Good-Quite Good	1.00	.80	- .20
Extremely Good-Quite Bad	5.00	4.82	- .18
Very Good-Very Bad	6.00	4.83	-1.17
Quite Bad-Equally	2.00	2.14	.14
Extremely Good-Slightly Good	2.00	2.14	.14
Extremely Bad-Equally	3.00	2.87	- .13
Quite Good-Equally	2.00	1.88	- .12
Quite Good-Quite Bad	4.00	4.02	.02
Very Bad-Slightly Bad	2.00	2.00	.00

between pairs of quantifiers is nearly the same (r=.96, df=34). That is, if the theoretical distance between "extremely" good and "extremely" bad is greater than that between "slightly" good and "slightly" bad, then the same held true, generally, for the respondents--and to about the same degree.

However, the mean theoretical distance was significantly greater than the mean perceived difference (2.89 vs. 2.59 units distance, t= 4.68, df=35, $p<.001$).

What this indicates is that the average theoretical distance between the 36 possible pairs of quantifiers is 2.89 units--.30 units greater than that assigned by respondents. The picture becomes clearer, though, in the "difference" column of Table V. Distance between 27 of 36 pairs of quantifiers was judged by respondents to be less than that theorized by Osgood. Eleven pairs showed greater distance than Osgood suggests.

The greatest theoretical over-estimates were between quantifying pairs on opposite sides of the continuum, while all but two underestimates were between pairs on the same side of the continuum from the middle. In other words, the farther apart the pairs were linguistically, the more Osgood tended to over-estimate the quantitative distance between them and vice versa. For instance, distance between contiguous pairs on the positive half of the scale was over-estimated an average of only .08 units and on the negative side by .06 units. But the distance between linguistic pairs on the opposite sides of the midpoint was over-estimated about one-fourth unit, on the average.

Compared with respondents, Osgood most over-estimated distances between "very" good and "slightly" bad, and "very" good to "very" bad

(by more than a unit). He assigns from one-half to more than .90 units additional to 10 other pairs than did the average respondent. Underestimated distances were most notable between "slightly" good and "slightly" bad (.61).

Midpoint Distances

Observed vs. Theoretical Midpoints of Linguistic Quantifiers

On the semantic differential scale, a respondent's checkmark yields a whole unit value (1, 2, 3, 4, 5, 6, 7). Thus, if he marks the extreme space on the positive side, his score is 7, designating a meaning intensity of 7. Yet, the most positive space could represent any intensity from 6 through 7, with a midpoint of 6.5, just as the midpoint of, say, the fourth or middle space, would yield a meaning space from 3 through 4, with a midpoint of 3.5 (the halfway point on a scale of 7).

In the second part of the study, each respondent marked on the 5-inch experimental scale line the closest point to good where he perceived "extremely" good to lie. Then he marked the farthest point from good where "extremely" good was perceived to lie. The author then took the midpoint of the two checkmarks and computed its total distance from good. This task was performed on four linguistic quantifiers, using both the good and bad end of the experimental scale as points of origin. The midpoints of "equally" and "neither" were omitted in the analysis.[2]

The average respondent's midpoints, as compared to Osgood's, are shown in Table VI. All but two observed midpoints differed significantly from the theoretical. The midpoint of "quite," as perceived by respondents, nearly duplicated Osgood's--on both sides of the scale.

For "quite" good, the average respondent's midpoint was 5.49 meaning units, compared with Osgood's 5.50. On the bad side, respondents differed from Osgood only .03 units from the theoretical midpoint of "quite" (1.47 vs. 1.50).

TABLE VI

DIFFERENCES BETWEEN OBSERVED AND OSGOODIAN MIDPOINT VALUES FOR FOUR LINGUISTIC QUANTIFIERS ALONG EACH SIDE OF THE GOOD-BAD CONTINUUM

Linguistic Quantifier	Observed Unit Midpoint	Osgood Midpoint	Difference	t-ratio	p
Extremely Good	6.29	6.50	-.21	4.2683	<.001
Very Good	5.91	6.50	-.59	10.5925	<.001
Quite Good	5.49	5.50	-.01	.1645	n.s.
Slightly Good	4.15	4.50	-.35	6.3177	<.001
Slightly Bad	3.08	2.50	.58	11.8609	<.001
Quite Bad	1.47	1.50	-.03	.4644	n.s.
Very Bad	1.08	.50	.58	9.1483	<.001
Extremely Bad	.74	.50	.24	3.5139	<.001

In other observed-theoretical comparisons, respondents assigned lower-valued midpoints to three linguistic quantifiers on the good side and higher values to three on the bad side. And there was an inverse relation between midpoint value and scale direction. For example, the

midpoint of "very" good was seen as .59 units below the theoretical, while "very" bad was assigned .58 units above. Similarly, "slightly" good's midpoint fell .35 units short of Osgood, while "slightly" bad earned .58 more units of meaning space. The "extremely" quantifier was judged short on the good and long on the bad (.21 and .24 units, respectively).

Table VI, in essence, says that quantifier midpoints are not as good as theorized, nor as bad. In other words, quantifiers of good and bad tend to be "squashed" toward the center of a purported seven-point, equal-interval scale.

Noteworthy, too, in Table VI are the extraordinarily large gaps in meaning between "quite" and "slightly," on both ends of the scale. Here, the midpoint analysis supports earlier reported findings. Also, the midpoints of "very" fall between "extremely" and "quite" on both sides of the scale, as reported earlier.

Meaning Distances on Two Scales

Comparison of 5-Inch and 100-Point Scales

In the third part of this study, six linguistic quantifiers were rated on a scale of 100 (or, "100-percent" scale). The quantifiers were: "extremely," "very," "quite," "slightly," "equally" and "neither." A respondent's rating automatically told what percentage of the scale was allotted to each quantifier. All comparisons in this phase were analyzed in percentage, rather than units, of meaning. However, to compare the 100-percent scale data with those gathered via the 5-inch scale, the author had to standardize the measuring procedure.

To illustrate, say respondent "A" wrote 90 on the 100-percent scale

for his rating of "extremely." That consumed 90 percent of the scale. What percent of the 5-inch scale did he allot to "extremely"? Extremely what? Good or Bad? The 5-inch scale was divided in half, leaving 2.5 inches or 80/32nds of an inch on either side. So, 80/32nds of an inch comprised 100 percent of the good side of the scale. (Likewise, for the bad side.) If respondent "A" had checked "extremely" good 70/32nds of an inch off center toward the good side of the scale, he would have rated "extremely" good 90 out of 100 percent. (Likewise, had that been the case from the midpoint toward the bad end of the scale.)

Table VII gives the comparative percentages of the good and bad portions of the 5-inch and 100-percent scales assigned to six linguistic quantifiers. On the good side, the only significant difference in scales is on the "quite" quantifier, which netted 8.5 percent more space on the 100-percent than the 5-inch scale ($p<.001$).

On the bad side, the quantifiers on the 5-inch scale were allotted a higher percentage of space over-all; significantly so, in one case. "Quite" bad got 4.2 percent more space on the 5-inch scale ($p<.05$). Conversely, "slightly" bad got 3.6 percent more space on the 100-percent scale ($p<.02$).

Any summary as to the probability of using the 5-inch and 100-percent scales interchangeably would be that very little over-all difference in space was allotted to the average quantifier. Specifically, however, respondents saw "quite" good as comprising a larger percentage of meaning space on the 100-percent scale, while "quite" bad was allotted more space on the 5-inch scale. "Slightly" good and bad tended to consume more space on the 100-percent scale. "Neither" and "equally" were viewed similarly on both scales.

TABLE VII

COMPARATIVE PERCENTAGES OF GOOD AND BAD PORTIONS OF SCALES ASSIGNED TO SIX LINGUISTIC QUANTIFIERS

	Bipolar Adjective					
Linquistic Quantifier	Good 5-inch scale	Good 100-pct. scale	dif	Bad 5-inch scale	Bad 100-pct. scale	dif
Extremely	92.73	94.28	-1.55	92.56	94.56	-2.00
Very	82.21	81.96	.25	84.50	82.12	2.38
Very	80.07	82.53	-2.46	---	---	---
Quite	65.37	73.95	-8.58	74.46	71.58	2.88
Quite	---	---	---	76.36	72.15	4.21
Slightly	20.52	23.14	-2.62	19.20	22.55	-3.35
Slightly	---	---	---	18.37	21.98	-3.61
Neither	---	---	---	50.91	50.51	.40
Neither	49.38	48.89	.49	---	---	---
Equally	49.21	49.88	- .67	---	---	---
Equally	---	---	---	50.29	50.49	- .20

The paradigm and data of Table VIII revolve around equality of intervals in both the 5-inch and 100-percent scales. Theoretically, the distance in percentage of meaning space should be "equal" between each pair of adjacent quantifiers. This would apply to both the good and bad sides of the scale. For the 5-inch and 100-percent scales to be used interchangeably, then, nearly equal percentages of space between corres-

TABLE VIII

DISTANCE IN SCALE PERCENTAGE POINTS BETWEEN ADJACENT LINGUISTIC QUANTIFIERS*

Adjacent Linguistic Pairs			Bipolar Adjectives: Good % of 5-inch scale	Good % of 100-pct. scale	Bad % of 5-inch scale	Bad % of 100-pct. scale
Extremely	minus	Very	11.59	12.03	8.06	12.44
Extremely	minus	Quite	27.36	20.33	17.15	22.69
Very	minus	Quite	15.77	8.30	9.09	10.25
Quite	minus	Slightly	44.85	50.81	54.95	51.28
Slightly	minus	Neither	19.90	22.03	21.37	21.10
Slightly	minus	Equally	19.63	23.02	20.75	21.08

*The average of repeated measures of "slightly" bad, "very" good and "quite" bad are utilized in this table.

ponding adjacent quantifiers should appear.

The obvious departure from equidistant meaning space in Table VIII is between "quite" and "slightly." The percentage of scale space between them averages more than halfway across the good and bad ends of both scales. This is double what it should be.

Again, data imply that "very" probably falls nearly equidistant from "extremely" and "quite." But if "very" were disregarded and the scales were equal interval, then the difference between the adjacent quantifiers should consume about 33 percent of scale space. For example, if respondents' perceived distance between "extremely" and

"quite" good as 27.36 percent of the scale, then the presumed 33 percent difference represents an over-estimate of 21.8 percent, as shown in Table IX.

TABLE IX

PERCENT OF THEORETICALLY OVER- OR UNDERESTIMATED SPACE BETWEEN ADJACENT LINGUISTIC QUANTIFIERS ON THE 5-INCH AND 100-PERCENT SCALE

Adjacent Linguistic Quantifiers			5-Inch Scale	100-Percent Scale
Extremely-	minus	Quite-Good	18.0%-O	39.0%-O
Extremely-	minus	Quite-Bad	51.5%-U	32.0%-O
Quite-	minus	Slightly-Good	34.5%-U	52.4%-U
Quite-	minus	Slightly-Bad	64.8%-U	53.9%-U
Slightly-Good	minus	Neither	40.3%-O	23.0%-U
Slightly-Bad	minus	Neither	36.0%-O	37.0%-O

O=Percentage Over-estimated; U=Percentage Underestimated

On the 5-inch scale, the largest over-estimation of meaning space on the part of Osgood was between the "slightly" and "neither" quantifiers: 40.3 percent on the good side and 36 percent on the bad. He also over-estimated the distance between "extremely" and "quite" good by 18 percent.

On the 100-percent scale, the equal-interval assumption over-estimated the observed distance between "extremely" and "quite" good by 39 percent. This was more than double the over-estimation between those two quantifiers on the 5-inch scale. The inflated difference gap between "slightly" bad and "neither," however, nearly was the same on the 100-percent as the 5-inch scale--37 percent compared with 36 percent.

Theoretically-underestimated meaning spaces occurred in both scales between "quite" and "slightly." However, on the good side of the scale, the underestimate was not as great on the 5-inch scale, while on the bad side, the 100-percent scale showed less underestimation.

The scales were differentiated in two other instances. Percentage of the 100-percent scale allotted for meaning space between "extremely" and "quite" bad was over-estimated 32 percent. It was underestimated more than 50 percent on the 5-inch scale. Conversely, the space between "slightly" good and "neither" was underestimated on the 100-percent scale (34 percent), but over-estimated on the 5-inch continuum.

In summary, the two scales both over-estimated distances between two sets of quantifiers. On two other sets, one scale over-estimated, while the other underestimated. (This would indicate that the two scales should probably not be used interchangeably.)

ENDNOTES

1. In Table IV, "extremely" and "very" are presumed synonymous; therefore, not to differ in unit distance. Likewise, with "neither" and "equally." Both are used interchangeably by Osgood. Note, too, that ten linguistic quantifiers are listed in Table III, instead of seven listed by Osgood. The purpose was to check on the feasibility of using "extremely" and "very" interchangeably, and "neither" and "equally" interchangeably.

2. Since respondents were asked to mark the points closest and farthest from good that were "equally" good and bad, a midpoint could not be determined. The midpoint from good would only be offset by the midpoint from bad.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

This exploratory study examined two primary constructs of the semantic differential, with a view toward improving its measurement sensitivity. The semantic differential is a combined associational and scaling procedure used to measure human affective and emotive meaning. Sensitivity is one of six criteria proposed by its author, Charles E. Osgood, as an evaluation of the method.

The specific items addressed were the equal-interval assumption of the S.D. scale, and the function of its adverbial units of intensity, or "linguistic quantifiers." Although a considerable amount of research has been expended on various facets of the system of semantic differentiation, it appeared to this writer that little critical attention had been directed toward these two particular areas; this, despite sizeable evidence from researchers, scholars and academicians which would warrant it in both instances.

To accomplish the study's objectives, a largely original scale design was devised, and a 33-item survey questionnaire composed. It was subsequently administered to a heterogeneous sample of respondents. Results were compiled in three groups, consistent with data furnished by the questionnaire.

Conclusions

Fairly clearcut answers emerged to each of the study's hypotheses, dealing with sensitivity of the semantic differential.

First, linguistic quantifiers were perceived in the ascending order of intensity espoused by Osgood, with the exception of the adverb "very." Each successive quantifier was assigned a higher meaning unit value, ascending from "extremely" bad through the scale center-point to "extremely" good. But, as shown in Table II, page 48, respondents placed "very" more than half of a unit below "extremely" good and nearly .40 units above "extremely" bad, in respective cases. "Very," then, conclusively fell between "extremely" and "quite" on the positive and negative sides of the scale, and must be viewed as a semantically unique linguistic quantifier with respect to others in this listing.

A second concern explored Osgood's claim that his scale represents an interval measurement level, and that the seven intervals are equidistant. In other words, the S.D. is purported to be an equal-interval scale. Psychological measurement units assigned by this study's respondents clearly indicated that the scale was not equal-interval to them.

From Table III, page 49, the average unit of meaning space the respondents assigned to the quantifying adverbs differ significantly from Osgood. Generally, quantifiers were seen to have less intensity of meaning on the good side of the scale, but more intensity on the bad side. Notably, "slightly" and "very" were viewed as comprising lesser degrees of "good" than Osgood claims, while "quite" and "extremely" bad were "more bad" than he states.

Continued evidence mounted that defied the equal interval assump-

tion. For example, in Table IV, page 53, meaning distances between several adjacent pairs of quantifiers varied widely, differently and erratically from Osgood's assumptions. Most conspicuous were the meaning distances between "quite" and "slightly." Contrary to the assumed 1-unit gap, 1.510 units separated "quite" and "slightly" good, while 1.759 units stood between "quite" and "slightly" bad. "Slightly" also fell short one-half to three-fourths units from the scale's midpoint. Assumed distance was 1 unit. Table IV shows that in both directions, Osgood most underestimates the meaning distance between "quite" and "slightly."

Further data at this point also pointed to the ill-advisement of treating "extremely" and "very" synonymously. On the good side, a half unit stood between them. Similarly, a four-tenth unit spread appeared between "very" and "extremely" bad.

Contrary to the hypothesis, "neither" and "equally" were assigned about the same degree of meaning space, indicating they can be used interchangeably.

More evidence challenging the equal-interval assumption emerged from Table V, page 55, which reported perceived meaning distances between each of the 36 combinations of linguistic quantifiers. Respondents saw less distance than did Osgood between 27 of the 36 pairs. The farther apart the quantifying pairs, the more Osgood overestimates the quantitative meaning distance; e.g., the assumed distance between "slightly" good and "quite" bad (3.00 Osgoodian units) is more likely to be an over-estimation than the assumed unit spaces between "extremely" and "slightly" good (2.00 Osgoodian units).

The final concern of the study was whether perceived midpoint

values of linguistic quantifiers varied significantly from the equal-interval standard. This research indicated they do. Table VI, page 58, showed that all but two respondent-designated midpoints varied significantly from Osgood's assumed midpoints. The theoretical midpoints of "quite" good and bad (5.50 and 1.50, respectively) did not differ beyond chance from those assigned by respondents. As with absolute unit values, midpoints of "quite" and "slightly" showed greater gap in semantical meaning than Osgood assumes. So, too, the midpoint of "very" good and bad were perceived to lie between "extremely" and "quite."

Perceived midpoints of quantifiers do not extend out toward bipolar ends of the scale to a degree assumed (see Figures 18 and 19), but exhibit a mild clustering effect toward the center of the scale as proposed by Hollingworth more than sixty years ago.[1]

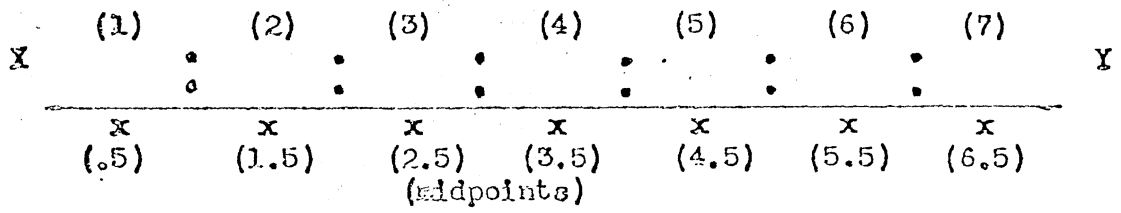


Figure 18. The Semantic Differential and Plot of Attendant Linguistic Quantifier Midpoints According to Osgood

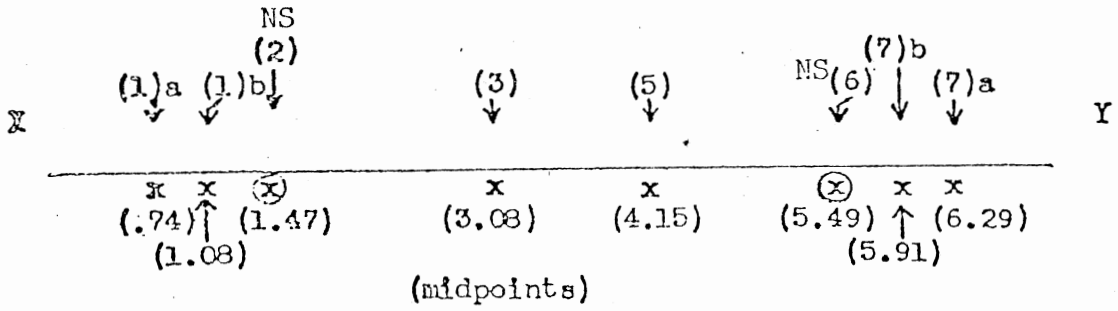


Figure 19. The Semantic Differential and Plot of Attendant Linguistic Quantifier Midpoints According to Study Respondents

Recommendations

> Science (is) the policy of subjecting The Word to the test of experience and of revising it accordingly, no matter how old The Word may be or who defends it . . . (a) method based not on the authority of age and prestige, but rather on the authority of observation and experiment.[2]

Conclusions derived from this research would suggest serious consideration be given to several design modifications in the method of semantic differentiation. All have to do with what the study termed "encoding efficiency" of the S.D. measurement instrument.

According to derivations of the present research, the scale falters due to design limitations in two categories. First, assumption that a given group of linguistic quantifiers is geometrically represented by a respective set of equidistant points on the scale is errant. Accuracy of information retrieval is thus lost or stunted because of design bias inherent in the scale. Second, as proposed in numerous other works and evinced in this study, confinement of the scale to seven points or choices representing the spectrum of a stated quality is insufficient.

Because of the significant variance between respondent-perceived and

Osgoodian adverbial midpoints on the S.D. scale, this study concludes that one of two avenues regarding the handling of linguistic quantifiers is mandated. Either these verbal weighting factors should be dissociated from the scale, or the method of segmentation of the scale should be re-analyzed. Results of the research indicate the system as presently conceived is not internally consistent or compatible.

Possible routes of effecting these modifications include selection of a revised group of adverbs which can be conclusively proven to represent an equidistant perceptual pattern on the scale. Otherwise, the adoption of the "unsegmented continuum" scale design utilized in this work is recommended as an alternative to the widely-recognized 7-step S.D. scale. An unsegmented design was seen as offering the enhanced resolution of a highly-segmented graphic scale while not risking the danger of respondent fatigue, which many scholars associate with the latter device.

Suggestions for Further Research

The results of this study intimate a variety of further research efforts related to semantic differential sensitivity. A listing of some of these as conceived by the writer include:

1. Additional research on the "unsegmented continuum" scale theory of semantic differentiation. The design incorporated in this study could be juxtaposed to several segmented scale models, and, with direct comparisons of scales with varying degrees of step division, more exacting determination of the ideal S.D. measurement instrument made.

2. Extended tabulation of adverbial candidates, with the view of establishing a comprehensive list of linguistic quantifiers proven to

represent specific points along the scale continuum. This would aid both the encoding process of the respondent vis-a-vis the instructional element, and the decoding process in data analysis and rationale of findings.

3. Ascertainment of degree of need for, and effectiveness of the instructional element in semantic differential surveys. The composition of the survey instructions were adjudged by this writer as critical, since improper orientation may confuse and/or bias respondents, with the consequent compromise of retrieved data.

4. Re-analysis of Osgood's List of Adjectival Pairs[3] for respective conformity of opposite polar intensities--possibly using this method, as reflected on page 51 of the study.

ENDNOTES

1. H. L. Hollingworth, "Experimental Studies in Judgment," Archeological Psychology, No. 29 (1913), pp. 44 ff.

2. Wendell Johnson, People in Quandaries (New York, 1946), p. 31.

3. Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Ill., 1957), p. 37.

BIBLIOGRAPHY

Carroll, John B. "Review of the Measurement of Meaning." Language, 35 (1959), p. 66.

Cliff, Norman. "Adverbs as Multipliers." Psychological Review, 66 (1959), pp. 27-44.

Conklin, E. S. "The Scale of Values Method for Studies in Genetic Psychology." University of Oregon Publication, 2 (1923).

Ford, LeRoy H. and Murray Meisels. "Social Desirability and the Semantic Differential." Educational and Psychological Measurement, 24 (1965), pp. 465-475.

Goode, William J. and Paul K. Hatt. Methods in Social Research. New York: McGraw-Hill Book Co., 1952.

Griffiths, Keith S. "The Construction of a Scale to Measure Attitudes Toward Defined Racial and Religious Groups." (Unpub. M.S. thesis, University of Washington, 1949.)

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Co., 1936.

Gulliksen, Harold. "How to Make Meaning More Meaningful." Contemporary Psychology, 3 (1958), p. 116.

Guttman, Louis H. "The Cornell Technique for Scale and Intensity Analysis." Educational and Psychological Measurement, 7 (1947), pp. 248-279.

Heise, David R. "Semantic Differential Profiles for 1000 Most Frequent English Words." Psychological Monographs, 79-8 (1965), p. 1.

Hollingworth, H. L. "Experimental Studies in Judgment." Archeological Psychology, No. 29 (1913).

Hull, C. L. Aptitude Testing. Yonkers: World Book Co., 1928.

Johnson, Wendell. People in Quandaries. New York: Harper and Brothers, 1946.

Jones, Lyle V. "The Nature of Measurement." Educational Measurement. Ed. Robert L. Thorndike. Washington, 1971, p. 336.

Kendler, Howard H. Basic Psychology. New York: Appleton-Century-Crofts, 1963.

Korzybski, Alfred. Manhood of Humanity. New York: E. P. Dutton and Co., 1921.

_____. Science and Sanity. Lancaster, Pa.: The International Non-Aristotelian Library Publishing Co., 1933.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1973.

McGinniss, Joe. The Selling of the President 1968. New York: Trident Press, 1969.

Newcomb, T. "An Experiment Designed to Test the Validity of a Rating Technique." Journal of Educational Psychology, 22 (1931), pp. 279-289.

Nunnally, Jum C. Psychometric Theory. New York: McGraw-Hill Book Co., 1967.

O'Neill, Harry W. "Advertising with the Differential Attitude Technique." Journal of Marketing, 27 (1963), p. 21.

Osgood, Charles E. "Semantic Differential Technique in the Comparative Study of Cultures." American Anthopologist, 66 (1964), pp. 171-200.

_____. "Semantic Space Revisited." Word, 15 (1959), pp. 192-200.

_____. "The Nature and Measurement of Meaning." Psychological Bulletin, 49 (1952), pp. 197-237.

_____ and George J. Suci. "Factor Analysis of Meaning." Journal of Experimental Psychology, 50 (1955), pp. 325-338.

_____, _____, and Percy H. Tannenbaum. The Measurement of Meaning. Urbana, Ill.: University of Illinois Press, 1957.

Sprague, Elmer and Paul W. Taylor. Knowledge and Value. New York: Hartcourt, Brace and World, 1959.

Symonds, P. M. "On the Loss of Reliability in Ratings Due to Coarseness of the Scale." Journal of Experimental Psychology, 7 (1924), pp. 456-461.

Thorndike. E. L. "A Constant Error in Psychological Ratings." Journal of Applied Psychology, 4 (1920), pp. 25-29.

Thurstone, L. L. and E. J. Chave. The Measurement of Attitudes. Chicago: University of Chicago Press, 1929.

Torgerson, Warren S. Theory and Methods of Scaling. New York: John Wiley and Sons, 1967.

Tukey, John W. "Symposium on Statistics for the Clinician." Journal of Clinical Psychology, 6 (1950), p. 64.

Warren, H. C. Dictionary of Psychology. New York: Houghton-Mifflin, 1934.

Wells, F. L. "A Statistical Study of Literary Merit." Archeological Psychology, No. 7 (1907).

Young, Pauline. Scientific Social Surveys and Research. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

APPENDIX A

QUESTIONNAIRE

QUESTIONNAIRE INSTRUCTIONS

Please write your name on the outside of the folder, below the printed number. Write that number, not your name, on the upper right-hand corner of this page. Mark all responses in ink.

Since this is a formal experiment, it is necessary to have your close cooperation. Your replies should represent your own judgment. Do not discuss items in the survey with those around you. There are no necessarily right or wrong answers. If you do have a question, refer it to the administrator.

The most effective and important means with which we communicate with one another is by the written or spoken word. Little scientific data, however, has been accumulated regarding how people express and interpret words.

The purpose of this experiment is to discover relative meanings of certain words by getting ratings of the words on a set of descriptive scales.

For instance, in describing Marilyn Monroe, one might say "She was an exceptionally beautiful woman." If you never knew of Marilyn Monroe, how beautiful would you feel she was--given this description? Providing you interpreted "exceptionally" to mean a high degree of beauty, you might accordingly (but not necessarily) mark a scale which represents the entire range of this quality, or lack of it, thus:

Ugly_____/____Beautiful

This is a three-part questionnaire. The first section consists of a number of scales labeled with a given quality. Imagine that the scale represents the entire spectrum of that quality.

For instance, in a scale marked

Dull_____Sharp

the point at the extreme left end of the line would represent the extreme value of dull, or "dullest." Similarly, the point at the right end of the line would represent the "sharpest" value of sharp.

Thus, if the description paired with the scale said "moderately" dull, you would place a hash mark at the point on the scale which would most approximately represent moderately dull to you.

In this questionnaire, it is important that you give your first impression to the modifier described. Please do not return to a scale you have previously marked and change your response.

1)

Slightly Bad:

Bad_____Good

2)

Quite Good:

Bad_____Good

3)

Very Good:

Bad_____Good

4)

Neither Bad
nor Good:

Bad_____Good

5)

Slightly Good:

Bad_____Good

6)

Slightly Bad:

Bad_____Good

7)

Extremely Bad:

Bad_____Good

8)

Equally Good
and Bad:

Bad_____Good

9)

Quite Bad:

Bad_____Good

10)

Very Good:

Bad_____Good

11)

Equally Bad
and Good:

Bad_____Good

12)

Very Bad:

Bad_____Good

13)

Extremely Good:

Bad_____Good

14)

Neither Good
nor Bad:

Bad_____Good

15)

Quite Bad:

Bad_____Good

The second part of this questionnaire is similar to the section you have just completed in that, once again, we ask for your immediate impressions to the items described.

Here you are asked to respond to "maximum limits" in each direction on the scale where a given description could lie.

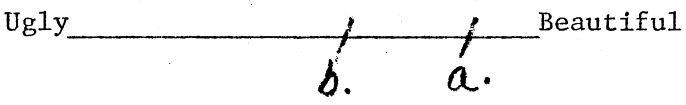
For instance:

Ugly_____Beautiful

a) Place a mark at the closest point to beautiful where "moderately" beautiful could lie, and label "a".

b) Place a mark at the furthest point from beautiful where "Moderately" beautiful could lie, and label "b".

Your resultant scale might accordingly (but not necessarily) appear as follows:



Again, as in the first part of this survey, the point at the extreme left end of the line would represent the extreme value of ugly, or "ugliest." Similarly, the point at the right end of the line would represent the "most beautiful" value of beautiful.

1)

Bad_____Good

a) Place a mark at the closest point to Bad where you feel "Quite Bad could lie, and label "a".

b) Place a mark at the farthest point away from Bad where you feel "Quite" Bad could lie, and label "b".

2)

Bad_____Good

a) Place a mark at the closest point to Good where you feel "Extremely" Good could lie, and label "a".

b) Place a mark at the farthest point away from Good where you feel "Extremely" Good could lie, and label "b".

3)

Bad_____Good

a) Place a mark at the closest point to Good where you feel "Equally" bad and Good could lie, and label "a".

b) Place a mark at the farthest point away from Good where you feel "Equally" Bad and Good could lie, and label "b".

4)

Bad_____Good

a) Place a mark at the closest point to Bad where you feel "Extremely" Bad could lie, and label "a".

b) Place a mark at the farthest point away from Bad where you feel "Extremely" Bad could lie, and label "b".

5)

Bad_____Good

a) Place a mark at the closest point to Neither Good nor Bad where you feel "Slightly" Bad could lie, and label "a".

b) Place a mark at the farthest point away from Neither Good nor Bad where you feel "Slightly" Bad could lie, and label "b".

6)

Bad_____Good

a) Place a mark at the closest point to Bad where you feel "Very" Bad could lie, and label "a".

b) Place a mark at the farthest point away from Bad where you feel "Very" Bad could lie, and label "b".

7)

Bad_____Good

a) Place a mark at the closest point to Good where you feel "Very" Good could lie, and label "a".

b) Place a mark at the farthest point away from Good where you feel "Very" Good could lie, and label "b".

8)

Bad_____Good

a) Place a mark at the closest point to Equally Good and Bad where you feel "Slightly" Good could lie, and label "a".

b) Place a mark at the farthest point away from Equally Good and Bad where you feel "Slightly" Good could lie, and label "b".

9)

Bad_____Good

a) Place a mark at the closest point to Good where you feel "Slightly" Good could lie, and label "a".

b) Place a mark at the farthest point away from Good where you feel "Slightly" Good could lie, and label "b".

10)

Bad_____Good

a) Place a mark at the closest point to Good where you feel "Quite" Good could lie, and label "a".

b) Place a mark at the farthest point away from Good where you feel "Quite" Good could lie, and label "b".

11)

Bad_____Good

a) Place a mark at the closest point to Bad where you feel "very" Bad could lie, and label "a".

b) Place a mark at the farthest point away from Bad where you feel "Very" Bad could lie, and label "b".

12)

Bad_____Good

a) Place a mark at the closest point to Equally Bad and Good where you feel "Slightly" Good could lie, and label "a".

b) Place a mark at the farthest point away from Equally Bad and Good where you feel "Slightly" Good could lie, and label "b".

Finally, on a scale of 0 (least) to 100 (most), where would you rate the following words:

Very _____

Equally _____

Quite _____

Extremely _____

Neither _____

Slightly _____

APPENDIX B

SURVEY POPULATION

TABLE X

SURVEY POPULATION

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
1.	M	56	Research Chemist	Caucasian	Bartlesville, Okla
2.	F	56	Housewife	Caucasian	Bartlesville, Okla
3.	F	85	Retiree	Caucasian	Aline, Okla
4.	F	63	Retiree	Caucasian	McWillie, Okla
5.	F	27	Secondary Education	Caucasian	Catania, Sicily
6.	M	26	Naval Officer	Caucasian	Oakland, Calif
7.	M	29	Businessman	Caucasian	Dallas, Texas
8.	F	27	Court Reporter	Caucasian	Allen, Texas
9.	M	29	Air Force Officer	Caucasian	Omaha, Nebraska
10.	F	25	Broadcasting	Caucasian	Wichita, Kansas
11.	M	30	Engineer	Chinese	Stavanger, Norway
12.	M	30	Police Detective	Caucasian	Rose Hill, Kansas
13.	F	28	Nurse	Caucasian	Wichita, Kansas

TABLE X (Continued

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
14.	M	30	Technician	Caucasian	Tulsa, Okla
15.	M	52	Common Laborer	Negroid	Denver, Colorado
16.	M	29	Common Laborer	Caucasian	Dewey, Okla
17.	M	56	Clerk	Caucasian	Bartlesville, Okla
18.	M	50	Banker	Caucasian	Stillwater, Okla
19.	F	27	Clerk	Caucasian	Houston, Texas
20.	M	21	Golf Pro	Caucasian	Stillwater, Okla
21.	F	46	Housewife	Caucasian	Bartlesville, Okla
22.	F	78	Retiree	Caucasian	Waco, Texas
23.	F	73	Retiree	Caucasian	Bartlesville, Okla
24.	M	42	Engineer	Caucasian	Dewey, Okla
25.	F	47	Housewife	Caucasian	Bartlesville, Okla
26.	M	46	Chemist	Caucasian	Bartlesville, Okla
27.	F	54	Housewife	Caucasian	Bartlesville, Okla

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
28.	F	56	Clerk	Caucasian	Tulsa, Okla
29.	F	59	Typist-Receptionist	Caucasian	Bartlesville, Okla
30.	M	25	Student	Caucasian	Stillwater, Okla
31.	M	49	Receptionist	Caucasian	Pawhuska, Okla
32.	F	24	Housewife	Caucasian	Terre Haute, Indiana
33.	M	31	Veterinarian	Caucasian	Terre Haute, Indiana
34.	F	22	Dietician	Caucasian	Amorita, Okla
35.	M	23	Farmer	Caucasian	Byron, Okla
36.	F	22	Student	Caucasian	Arkansas City, Kansas
37.	M	55	College Professor	Caucasian	Stillwater, Okla
38.	M	35	Photographer	Caucasian	Tulsa, Okla
39.	F	33	Elementary Education	Caucasian	Miami, Florida
40.	F	34	Salesperson	Caucasian	Indianapolis, Ind

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
41.	M	46	Technician	Caucasian	Indianapolis, Ind
42.	M	19	Student	Caucasian	Tulsa, Okla
43.	F	18	Student	Negroid	Oklahoma City, Okla
44.	F	17	Student	Caucasian	Stillwater, Okla
45.	M	20	Reporter	Caucasian	Stillwater, Okla
46.	M	45	Plant Foreman	Caucasian	Norman, Okla
47.	F	21	Student	Negroid	Tulsa, Okla
48.	F	19	Receptionist	Caucasian	Tulsa, Okla
49.	M	21	Clerk	Caucasian	Oilton, Okla
50.	F	20	Student	Puerto Rican	Stillwater, Okla
51.	M	19	Student	Caucasian	Enid, Okla
52.	M	20	Student	Negroid	Stillwater, Okla
53.	F	18	Student	Caucasian	Tulsa, Okla

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
54.	M	19	Student	Negroid	Lawton, Okla
55.	M	25	Editor	Caucasian	Altus, Okla
56.	F	20	Student	Mexican	Stillwater, Okla
57.	M	22	Student	Caucasian	Stillwater, Okla
58.	M	23	Store Manager	Negroid	Ponca City, Okla
59.	F	19	Student	Caucasian	Stillwater, Okla
60.	M	35	College Professor	Caucasian	Stillwater, Okla
61.	F	20	Student	Chinese	Stillwater, Okla
62.	M	53	Clerk	Caucasian	Pawhuska, Okla
63.	M	48	College Professor	Caucasian	Stillwater, Okla
64.	F	19	Receptionist	Caucasian	Shawnee, Okla
65.	F	57	Secretary	Caucasian	Akron, Ohio
66.	M	39	Film Producer	Caucasian	Marianna, Ark

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
67.	F	27	Secretary	Caucasian	Bartlesville, Okla
68.	F	51	Secretary	Caucasian	Bartlesville, Okla
69.	M	52	Engineer	Caucasian	Dewey, Okla
70.	F	33	Chemist	Caucasian	Indianapolis, Ind
71.	M	37	Archaeologist	Caucasian	Indianapolis, Ind
72.	M	52	College Professor	Caucasian	Glencoe, Okla
73.	M	17	Student	Caucasian	Glencoe, Okla
74.	F	63	Housewife	Caucasian	Santa Rosa, Calif
75.	M	64	Agricultural Specialist	Caucasian	Santa Rosa, Calif
76.	F	36	Cashier	Caucasian	Los Angeles, Calif
77.	M	37	Researcher	Caucasian	Los Angeles, Calif
78.	F	52	Secretary	Caucasian	Bartlesville, Okla
79.	M	16	Student	Caucasian	Bartlesville, Okla

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
80.	F	68	Retiree	Caucasian	Medford, Okla
81.	M	69	Retiree	Caucasian	Medford, Okla
82.	F	71	Retiree	Caucasian	Bartlesville, Okla
83.	M	73	Retiree	Caucasian	Brownwood, Texas
84.	M	62	Banker	Caucasian	Stillwater, Okla
85.	F	54	Housewife	Caucasian	Stillwater, Okla
86.	F	19	Student	Polynesian	Stillwater, Okla
87.	F	19	Student	Caucasian	Ponca City, Okla
88.	F	18	Clerk	Caucasian	Cushing, Okla
89.	F	19	Student	Caucasian	Tulsa, Okla
90.	F	19	Student	Caucasian	Jet, Okla
91.	F	21	Clerk-typist	Caucasian	Wichita, Kansas
92.	M	58	Lawyer	Caucasian	Morrison, Okla

TABLE X (Continued)

Respondent Number	Sex	Age	Occupation	Ethnicity	Geographic Location
93.	F	23	Secretary	Caucasian	Stillwater, Okla
94.	F	22	Salesperson	Caucasian	Stillwater, Okla
95.	F	26	Secretary-Receptionist	Caucasian	Miami, Okla
96.	F	32	Secretary	Caucasian	Stillwater, Okla
97.	M	20	Army Medic	Caucasian	San Antonio, Texas
98.	F	39	Accountant	Caucasian	San Francisco, Calif
99.	F	37	Housewife	Caucasian	Santa Rosa, Calif
100.	M	59	Technician	Caucasian	Bartlesville, Okla

VITA[2]

Steven Merrill Schnitzer

Candidate for the Degree of

Master of Science

Thesis: THE SEMANTIC DIFFERENTIAL: EXPLORING THE EQUAL-INTERVAL ASSUMPTION

Major Field: Mass Communications

Biographical:

Personal Data: Born in Cherokee, Oklahoma, July 20, 1948, the son of Dr. and Mrs. A. Merrill Schnitzer.

Education: Graduated from College High School, Bartlesville, Oklahoma, in May, 1966; received Bachelor of Science degree in Business Administration from Oklahoma State University in May, 1970; attended Wichita State University, 1972; took extension through Eastern Washington State University, 1973; completed requirements for Master of Science degree at Oklahoma State University in May, 1979.

Professional Experience: Commissioned Second Lieutenant, United States Air Force, May, 1970. Discharged with rank of Captain in April, 1975; graduate teaching assistant, Oklahoma State University School of Journalism and Broadcasting, 1976-77.