ABSTRACT

THE STABILITY OF VALUES
AND VALUE SYSTEMS

by Charles C. Hollen

The purpose of this research is to explore some contributing factors in the reliability of Rokeach's terminal and instrumental value scales. It is addressed to the following questions:

1. How reliable are the terminal and instrumental value scales, as a value measuring instrument?

2. To what extent is their reliability influenced by factors in the test, such as the degree of vagueness or clarity of the values used as stimuli, or by factors in the person taking the test, such as his degree of commitment to a system of personal values?

3. Are some values more reliable than others? If so, what are some systematic determinants of differential stability among values?

4. Are some people more reliable than others? If so, what variables may be used to predict individual differences in value system stability?

In the theoretical viewpoint which guided this research a value is defined as a belief "centrally located in one's belief system about how one ought or ought not to behave, or about some end state of existence worth attaining" (Rokeach, 1966b: 10). A value system is an hierarchical organization--a rank ordering--of a person's values in terms of their importance to him. A distinction is made between

terminal values (end-states worth attaining) and instrumental values (how one ought to behave). Following these definitions, two separate value ranking scales were devised to obtain measures of individuals' terminal and instrumental value systems.

Two investigations were conducted with students enrolled in introductory psychology at Michigan State University during the spring and fall terms of 1966. In both studies, questionnaires were administered containing the terminal and instrumental value scales and related questions. Subjects were retested seven weeks later for the purpose of obtaining test-retest stability data.

Among the findings of these investigations were the following: (a) Efforts to increase stimulus clarity by adding defining phrases to the values used as stimuli resulted in decreased rather than increased reliability of response. (b) For both terminal and instrumental values, it was found that those values which are very high or very low in an individual's value hierarchy are more stable over time than those in between. (c) There is a slight but significant relationship between the stability of an individual's terminal value system and his degree of instrumental value system stability. (d) No relation was found between open and closed-mindedness and value system stability. (e) Relationships between value system stability and five predictor variables (commitment, importance, vagueness, difficulty, and uncertainty) were found to be very slight and in general non-significant.

On the basis of these findings, it was concluded that:

1. Although the value ranking scales provide data of sufficient reliability to discriminate between groups, they are not reliable enough to warrant the use of correlational techniques.

2. Variations in the stability of value rankings are chiefly a function of the instability of the ranking scales, and only slightly influenced by factors in the person.

3. Value system stability may be considered to be present to a slight but significant extent, as a general personal characteristic common to both terminal and instrumental value systems.

4. No conclusions can be drawn regarding attempts to predict individual differences in value system stability. Due to the presence of excessive measurement error, the findings of non-significant relationships do not permit us to accept the null hypothesis--i.e., to conclude that the various predictor variables are not related to value system stability.

Finally, it is suggested that the continuing search for ways to improve the reliability of the ranking scales, along with further attempts to discover personality variables related to value system stability may prove fruitful.

Approved: Walter Schaad

THE STABILITY OF VALUES AND VALUE SYSTEMS

By

Charles C. Hollen

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1967

ACKNOWLEDGMENTS

I wish to express my deep thankfulness to all those who helped in the planning and execution of this research.

I am particularly grateful to my advisor, Dr. Milton Rokeach, for his always generous contributions of time, patience, incisive questions, and wise counsel. I would also like to thank the other members of my committee--Dr. Eugene Jacobson and Dr. Verling Troldahl--whose judicious observations and fair but firm criticisms were of distinct help. I greatly appreciate the aid of Miss Judy Goldbaum, Mr. Joseph Bivins, Mr. Dennis Skinner, and Miss Berna Pinner in the gathering and analysis of data and the preparation of the manuscript. This research was supported by a grant from the National Science Foundation.

Finally, I am deeply indebted to my wife Karen, who not only provided an abundance of encouragement, prodding, and general moral support, but also served as chief typist in the preparation of the final drafts. The value of her aid and comfort is incalculable.

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INTRODUCTION

Reliability may be considered as a characteristic of a test, a function of the reliability of each of its items. Alternatively, it may be considered as a characteristic of a population, a function of the reliability of each person within the population. The stability of subjects' responses to a test from one occasion to another is determined partially by characteristics of the test and partially by characteristics of the subject. It is often desirable to assess the extent to which reliability is a function of each of these component sources of variance. The purpose of the present research is to explore some contributing factors in the reliability of Rokeach's terminal and instrumental value scales. It is addressed to the following questions:

1. How reliable are the terminal and instrumental value scales, as a value measuring instrument?

2. To what extent is their reliability influenced by factors in the test, such as the degree of vagueness or clarity of the values used as stimuli, or by factors in the person taking the test, such as his degree of commitment to a system of personal values?

3. Are some values more reliable than others? If so, what are some systematic determinants of differential stability among values?

4. Are some people more reliable than others? If so, what variables may be used to predict individual differences in value system stability?

Definitions of Value

Conceptual definitions of value are numerous and varied. Tisdale (1961) suggests that psychological theory and research on values may

be classified into five categories according to which of the following variables is seen as critical to the definition of value: need, predisposition, choice behavior, concept, or situational relationships. In addition, a number of other concepts such as evaluative attitude, sentiment, ideal, and interest have been used to refer to the major dimension of value (Dukes, 1955). A considerable volume of research has been based upon conceptions of value which reflect agreement with Cantril and Allport that values are "pervasive, enduring, and generalized" rather than narrow, transient, and specific (1933: 272). Most theoretical conceptions of value would also tend to agree with Kluckhohn that an individual's values are relatively more concerned with the "ought" rather than the "is," with the "ideal" rather than the "real" (1951: 390).

Definitions of Terms Used in This Research

In the theoretical viewpoint which guided this investigation, values are seen as part of the belief system of an individual (Rokeach, 1966b). A belief is considered to be "any simple proposition, conscious or unconscious, inferred from what a person says or does, capable of being preceded by the phrase 'I believe that...'" (1966b: 3). A belief system consists of "the total universe of a person's beliefs about the physical world, the social world, and the self" (1966b: 10). In this system, a value is a type of belief "centrally located within one's belief system about how one ought or ought not to behave, or about some end state of existence worth attaining" (1966b: 10). A value system "is an hierarchical organization--a rank ordering--of ideals or values in terms of their importance" (1966b: 10).

Terminal values are distinguished from instrumental values as follows: a terminal value is a value which refers to an end state of existence which the individual feels is worth attaining. An instrumental value is a value which refers to how one ought to behave--to an ideal mode of conduct. Instrumental values are conceived as serving the interests of the terminal values, and as relatively less important than the terminal values. An individual's terminal value system is seen as separate from his instrumental value system, but related to it in that any of the instrumental values that an individual holds may operate in the service of any of his terminal values. There is no necessary one-to-one relationship between the instrumental and the terminal values.

This theoretical viewpoint differs from previous theories of values in its emphasis on the notion of a value system. It is assumed that an individual's values are not autonomous and isolated from one another, but that they are interrelated in hierarchical order, forming his personal value system. This value system is assumed to be more or less internally consistent, so that a change in one value will affect other values related to it. This is not to say that there cannot be inconsistencies or contradictions within the system of values, but only that there must be some potential means of resolving such inconsistencies when they arise.

This theoretical conception of a value system has important implications for the measurement of values and of value stability. Of particular relevance to the present study is the implication that value system stability may be meaningfully considered as a

concept separate and distinct from value stability, referring to the stability of one or more single values. In presenting the results of this study, the term value system stability will be used to refer to the change in a value system as a whole over time, i.e., the difference between the state of the entire system at Time 1 and at Time 2. Thus, terminal value system stability will refer to the degree of constancy of one individual's terminal value system. Similarly, instrumental value system stability will be used to refer to the constancy of one individual's instrumental value system. The term value stability will be used to refer to the constancy of one particular value over time, either for a single individual or within a group. Strictly speaking, the term stability could also be used to refer to a characteristic of a test. However, to avoid confusion due to too many different usages of this term, the term reliability will be used when referring to the stability of the measuring instrument.

Value Measuring Instruments

A number of instruments have been provided for measuring or describing the values of individuals and groups (Allport, Vernon and Lindzey, 1960; Cohen, 1941; Glaser and Maller, 1940; Harding, 1944; Hunt, 1933; Lurie, 1937; Ohnmacht, 1965; Rosenberg, 1956; Rothney, 1936; Smith, M.B., 1949; Van Dusen, Wimberly, and Mosier, 1939; White, R. K., 1951; Wickert, 1940; Woodruff, 1942). Most of these employ straight answer questionnaires rather than content analysis or judgmental classification of free response protocols, situational measures, or other more indirect measures. Many of them utilize ranking scales in one form or another.

One of the earliest attempts to apply quantitative measurements to the study of values was Vernon and Allport's application of Spranger's Types of Men (1928) in "A Study of Values" (Vernon and Allport, 1931). This test, later revised and updated, (Allport, Vernon and Lindzey, 1960) has become the most widely used and accepted instrument for the measurement of values. As Dukes points out, "Even though a number of standardized tests of values are available, the large majority of investigators employ the Allport-Vernon Study of Values" (1955: 26).

The Allport-Vernon test, designed to tap Spranger's six categories of values (theoretical, economic, political, aesthetic, social and religious) utilized multiple choice questions providing alternative answers which the subject is asked to rank in order of preference. Each of the alternative responses is assumed to represent one of the six value types. The items represent a wide variety of specific interests and preferences. Applying the label "theoretical" (or "economic" or "social") to the sum of a collection of preferences is an inference or abstraction, the validity of which depends upon how well a person's preferences can be assumed to reflect his values. Allport and Vernon do not address themselves directly to this underlying assumption--they provide no theoretical framework for the assumed connection between interests and values. Beech points out that "it is easy to conceive of two individuals, both with interests in civil rights, but who see their interest in civil rights as relevant to, or in the service of, entirely different values..." (1966: 5). Thus the validity of this operational definition of values may be somewhat open to question.

The emergence of the Allport-Vernon test aroused the interests of psychologists in empirical studies of values, and stimulated the construction of numerous other tests of values. Some of these were attempts to modify and improve the Allport-Vernon test, while others represented new approaches to the measurement of values.

Lurie (1937) constructed a test of evaluative attitudes based on Spranger's six categories of values, in which he used material very similar to that of the Allport-Vernon test. Instead of ranking scales, however, he used seven-point Likert-type rating scales. This scoring system avoids one disadvantage inherent in ranking scales, i.e., the fact that the rankings of different individuals or groups may not be directly comparable with those of others, because of unknown differences in the absolute importance which different individuals and groups attach to certain values or to values in general. Strictly speaking, the results of a ranking scale cannot be interpreted as indicating which values the individual or group feels are important and unimportant, but only which values are felt to be relatively more important and less important than others. The rating scale method, by allowing each response to vary independently of all others, allows comparisons between the absolute importance placed upon each value by different individuals and groups. However, the rating scale introduces other sources of error. For example, one group of subjects may interpret the scale categories consistently higher or lower than another group of subjects, or may tend to respond to them more at the extremes (Van Dusen, Wimberly, and Mosier, 1939). Biasing effects such as these have been found to be present in a test of values using rating scales (Lurie, 1937: 22).

The primary strength of the ranking method is that it seems to reflect the fundamental nature of the conflict and decision processes so often encountered in life situations, where the maximization or attainment of one value or goal necessarily implies the sacrifice of one or more others. As Duffy has pointed out, "The individual taking the test is forced to choose. He may, as in life situations, favor one interest only at the expense of another" (1940: 607). This seems especially applicable to the role of values in everyday life. Indeed, Rokeach has recently suggested that one of the primary functions of values consists of their use as rules for deciding what to do, i.e., as overriding principles for resolving conflicts of interest and preference (1967: 7).

There have been numerous other attempts to construct standardized tests of values. Rothney (1936) constructed a revised version of the Allport-Vernon test for use with high school students. Van Dusen, Wimberly and Mosier (1939) constructed a Standards Inventory for the measurement of evaluative attitudes based on Lurie's test and using Lurie's method of scaling, i.e., seven-point rating scales. Glaser and Maller (1940) constructed a modified form of the Allport-Vernon scale called the Interest-Values Inventory, using a combination of ranking and rating scales and designed to measure theoretical, aesthetic, social and economic values. Wickert (1940) designed a test for personal goal-values using items similar to those used by Allport and Vernon but designed to measure ten personal goal values utilizing four-alternative ranking scales. Harding (1944) constructed a Value-Type Generalizations Test containing ten subscales and using five-point Likert rating scales. The foregoing tests

represent chiefly attempts to improve upon Allport and Vernon's operationalization of Spranger's category system or to provide revisions of the category system itself.

All of the tests cited thus far reflect the influence of Allport and Vernon's Study of Values in that they measure values indirectly. Items are chosen which measure a person's specific preferences or interests, which are assumed to represent his values. In most cases the use of the indirect method is implicitly justified by the well-known difficulties inherent in direct self-rating tests. The direct self-rating technique, because it depends heavily upon the degree of conscious self-awareness, accuracy of self-perception, and understanding of the conceptual meaning of the trait being evaluated, has frequently proven to be an unreliable approach to the measurement of values. Allport and Vernon, comparing the scores obtained by their test with both judges' ratings and self-ratings for the six value types found low correlations (ranging from .06 to .69) between test scores and ratings. They conclude that "empirical validity cannot be found satisfactorily by comparing the test results with ratings, owing to the unfamiliarity of the average rater with the conceptual nature of the values" (Vernon and Allport, 1931: 245).

Other investigators, however, have succeeded in devising methods more direct than those used by the aforementioned tests, yet not as point blank as self-rating. Smith, for example, used a method which was more phenomenological, and which led to a noticeably different conception of values (Smith, M. B., 1949).

Instead of starting with a set of *a priori* abstract categories and constructing test items to represent them, he simply asked his subjects an open-ended question about "what you think is important in life" (1949: 479). Responses to this question were classified into content categories. The resulting list of personal values, in order from most to least frequently mentioned, are: economic security, home and family life, liberty and freedom, health, getting along well with people, world peace, happiness, enjoyment of life, comforts of life, education, religion, pride in work, doing job well, economic advancement, getting ahead on job, recreation, sports and hobbies, decency and morality, being a good citizen, community participation, and racial and ethnic tolerance. It is evident from this list that the phenomenological method of measuring values leads to a considerably more concrete notion of values--as personal attributes or things that people strive for rather than as abstract, generalized categories such as social, theoretical, or aesthetic.

Woodruff (1942) constructed a generalized test of values consisting of three personal problem situations each offering eight possible courses of action which the subjects were asked to evaluate. This test provides information concerning the relative importance of twelve values, each value represented in two of the three problems. Woodruff's twelve values, which are remarkably similar to the sixteen values later obtained by Smith are: friends, home life, social service, personal attractiveness, intellectual activity, security, political power, comfort, society, wealth, excitement, and religion.

Occasionally investigators interested in obtaining measures of the importance of certain specific types of values among selected

groups have utilized direct ranking scales, in which the stimuli to be ranked are the values themselves. Hunt (1933) in a study of the relative value of certain ideals, presented subjects with a list of seventeen ideal traits (i.e., cheerfulness, cleanliness, courage, honesty, obedience, thrift, etc.) and asked them to rank these ideals in order of importance. More recently Ohnmacht (1965) used a list of ten educational objectives (i.e., the acquisition of important information, the development of effective methods of thinking, etc.) which the subjects were asked to assign numbers from one to ten.

A variation of the direct technique of value measurement is the recent work of Rosenberg (1956, 1960a, 1960b) who used thirty-five value items constructed on the basis of White's value-analysis technique (1951) and Murray's analysis of major needs (1938). Subjects are given thirty-five cards each containing a value statement (i.e., being looked up to by others, having a steady income, having interesting work to do, etc.) and are asked to sort these into twenty-one categories from "gives me maximum satisfaction" to "gives me maximum dissatisfaction" (1956: 368).

The generalized test of values which is the basis of the present research differs from Allport and Vernon's test in two ways. It combines the more concrete and phenomenological conception of values exemplified by Smith, Woodruff, and others, with the direct ranking method used by Hunt, Ohnmacht, and others. It differs from the tests used by these investigators chiefly in that it consists of two separate value ranking scales designed to measure terminal and instrumental values.

Lovejoy (1950) in a philosophical essay, has drawn a distinction between "terminal" and "adjectival" (instrumental) values which

reflects the philosophers' historic concerns with means and ends. No psychological investigation of values thus far has embodied this distinction. It can readily be seen that the type of values used by Hunt (cleanliness, courage, honesty) are of a different nature than those used by Smith (economic security, home and family life, liberty and freedom). The former seem to describe personal attributes while the latter describe states of existence. Many psychological investigations of values have more or less indiscriminately mixed these two types of values in a single scale. The present research is based on an explicit conceptualization of each of these two types of values represented in two separate value scales.

The measuring instrument initially chosen by Rokeach as an operational test of his conception of terminal and instrumental value systems consisted of two simple ranking scales, each containing twelve values. Subjects were asked to rank each set of values according to "the relative importance of these values for you" (Beech, 1966). The actual sets of values used in the two ranking scales were selected on the basis of a series of pilot studies. (For a description of these studies, see Beech, 1966). The criteria used in the selection of the final lists of values were designed to eliminate synonomous or highly correlated values, values which were uniformly ranked very high or very low, and values which failed to discriminate between known subgroups in a college population as defined by various demographic characteristics (Beech, 1966:95). The final list of terminal values consisted of: A comfortable life, A meaningful life, A world at peace, Equality, Freedom, Maturity, National Security, Respect for others,

Respect from others, Salvation, True friendship, and Wisdom. The final list of instrumental values included: Broadminded, Clean, Cooperative, Courageous, Forgiving, Honorable, Intellectual, Polite, Responsible, Self-disciplined, Tender, and Trustful. These value scales will be referred to in this report as the Form A value scales.

The Reliability of Value Measuring Instruments

The reliabilities of a number of value measuring instruments are shown in Table 1, along with the scaling technique and the method of determining reliability used by each investigator. It may be noted that although success in devising reliable tests of values has varied somewhat, there are a number of tests which achieve a reasonably high level of reliability. Of particular interest to the present study are the findings of Allport, et al. and of Rosenberg, both of whom used a test-retest method such as is used in this study. Allport, Vernon and Lindzey determined repeat reliability coefficients for two populations, one after a one-month interval, the other after a two-month interval. The mean repeat reliability coefficient for the Study of Values was .89 for the one-month interval and .88 after two months (1960:10). Rosenberg, who used a rank order correlation to determine the stability of each individual's value rankings, reports a median correlation of .88 for twelve subjects retested after a four- to five-week interval (1956, 368).

A previous study conducted in the spring of 1966 by Rokeach using the Form A value scales yielded repeat reliability coefficients of .65 for the terminal scale and .65 for the instrumental scale, for a population of 210 college students retested after a seven-week interval. The procedure used to determine an estimate of the

Table 1. The reliabilities of several value measuring instruments.[1]

Author (s)	measurement technique	reliability method	reliability coefficient
Rothney (1936)	indirect ranking scales	method unreported	.42
Van Dusen, Wimberly and Mosier (1939)	7-point Likert scales	split-half	.81*
Glaser and Maller (1940)	ranking and rating scales	test-retest (10 days)	.82
Wickert (1940)	indirect ranking scales	split-half	.42*
Cohen (1941)	Thurstone scaling	alternate forms	.85
Harding (1944)	5-point Likert scales	split-half	.88*
Rosenberg (1956)	direct sorting and ranking	test-retest (4 to 5 weeks)	.88
Allport, Vernon and Lindzey (1960)	indirect ranking scales	test-retest (1 mo.) (2 mo.)	.89 .88
Rokeach (1966a)	direct ranking scales	test-retest (7 weeks)	.65

[1]Listed chronologically

*Obtained by averaging the reported reliabilities of several subscales. Reliability of total scale unreported.

reliability of the value scales was similar to that used by Rosenberg. A Spearman rank correlation coefficient was computed for each individual subject between his initial rankings and his retest rankings of twelve values. The reliability coefficients cited above are the medians of the two distributions of individual rho correlations, for terminal and instrumental values, respectively. In order to attain an understanding of the empirical import of these reliability coefficients, a Spearman rank correlation was computed for a hypothetical subject, all twelve of whose values changed three units (i.e., from 1 to 4, 2 to 5, etc.) from test to retest. The correlation under these conditions was found to be .62, a figure closely approximating the median rho correlation obtained for each of the two value scales. Thus it is seen that the reliability coefficients of the value scales represent a considerable amount of shift, on the average, from test to retest. Based on these results, the present study undertook to investigate some factors contributing to the reliability of the value scales, including (1) the extent to which all of the values in general may be ambiguous or ambivalent stimuli, evoking different responses on different occasions; (2) the extent to which some particular values are more stable over time than others; (3) the extent to which differential stability of personal values can be considered a meaningful characteristic of a college population; and (4) the extent to which individual differences in value system stability can be predicted or accounted for by certain other variables.

The Clarity or Ambiguity of Values Used as Stimuli

Guilford (1954) has pointed out that the degree of clarity or ambiguity of the stimuli used in a scale will partially determine

the reliability of responses to those stimuli. The lists of values used as stimuli in the Form A terminal and instrumental value scales consisted of single words or phrases referring to broad general concepts, i.e., a comfortable life, clean, etc. These words and phrases may have been open to differing interpretations, not only by different individual subjects, but by the same subject on different occasions, thereby affecting the reliability of response to the scales. Informal reports from subjects during administrations of the Form A value scale tended to confirm the supposition that many of the values were, in fact, ambiguous or subject to more than one interpretation. Therefore, in the Form B revision of the value scales defining phrases or synonyms were added in parentheses following each value to provide clarification of the meaning intended by the investigators. In revising Form A it was also decided that certain values should be removed and replaced because they had largely failed to discriminate between criterion groups in a college population. It was expected that the addition of defining phrases to the values would increase the repeat reliability of the test by insuring greater uniformity in the way each value was interpreted by an individual on two separate occasions.

The Stability of Values as a Function of Their Position in a Hierarchy

The values which are used in the terminal and instrumental ranking scales are all positive values for most people. Therefore, when a person is asked to rank these values in order according to their relative importance, he may encounter difficulty deciding which of the values are more important than others to him. However, for most

people it is expected that one or two values will stand out from the list as clearly more important than the others and likewise that one or two values will be immediately perceived as substantially less important than the rest. Further, it is expected that an individual's choices of the one or two values which are most important and the one or two which are least important of the twelve will vary less from one occasion to another than will his rankings of those values in between the two extremes. This should produce a curvilinear relationship between the ranked position of a value and its stability from test to retest. Those values which are ranked at both extremes on the scale should be significantly more stable from initial test to retest than those values ranked between the two extremes.

The Stability of Personal Values

In order to draw conclusions about the reliability of a test based on two administrations separated by a time interval, it is necessary to provide a certain amount of empirical justification for the underlying assumption that the trait being measured by the test does not change markedly during the time between test and retest. Cattell (1957) has pointed out that a test-retest reliability coefficient may be considered as either a measure of the reliability of the test or as an index of the constancy or stability of the trait being measured, depending on the length of the time interval between tests. Duffy (1940) points out that in considering tests of values which were administered at intervals of one year or more, the coefficients of correlation between test and retest "may be

taken to reflect, not the unreliability of the test, but the shifting of the students' values, as brought about in part by their college training" (1940: 599). There have been several studies of the constancy or stability of personal values over extended periods of time (Hilton and Korn, 1964; Jacob, 1957; Kelly, 1955; Whitely, 1938).

Whitely (1938) administered the Allport-Vernon test to eighty-four college students on four occasions separated by intervals of one year--once at the beginning of their freshman, sophomore, junior and senior years in college. Computing the Pearson product-moment correlations between all possible pairs of administrations, he found that the average of these correlations was .625. There was a slight tendency for the coefficients of stability to increase in the later administration. Whitely states that "it may be concluded that the values are fairly stable" (1938: 406).

In an attempt to measure changes in personal values over a much shorter period of time, Hilton and Korn (1964) submitted the Allport-Vernon test to thirty college seniors and graduate students seven successive times at one-month intervals. They report an increase in reliability coefficients in later administrations, the mean repeat reliability coefficient for the first four administations being .79 and for the last three administrations being .93. Despite this high level of stability, they found statistically significant change trends during the seven-month interval which they interpreted as evidence of significant changes in values in the members of the group during this period.

Jacob (1957) reports a large scale study of changing values in college, in which evidence gathered with a wide variety of testing techniques led him to the conclusion that "basic values remain largely constant through college" (1957: 56), although certain consolidating and sharpening trends occur.

Kelly (1955), using the Allport-Vernon test, retested 446 adult men and women after a time interval of twenty years. He found repeat reliability coefficients for the Allport-Vernon subscales ranging around .50 even after this lengthy span of time.

The foregoing studies seem to demonstrate that an individual's values are a stable and enduring aspect of his personality, even during his college years. However, these studies were all based on Allport and Vernon's operational definition of values. The failure of the above studies to show substantial changes in values during college days may be due in part to the insensitivity of Allport and Vernon's instrument as a measure of value change. Presumably, a change in a person's underlying values may or may not be reflected in a change in his specific behavioral preferences as measured by the Allport-Vernon test. On the other hand, the direct ranking method of measuring values used by Rosenberg and by Rokeach may be relatively sensitive to small shifts in value importance, even momentary and transient ones. Evidence gathered thus far using Rokeach's twelve-item ranking scales seems to indicate that at least among college students, the relative importance of various "values" expressed as situation-free abstract concepts fluctuates considerably from one occasion to another. Because of the nature of the direct ranking method, Rokeach's test may be highly sensitive to

value change but at the cost of decreased stability over time.

In general, empirical evidence on the stability of personal values over time corroborates the theoretical conceptions of values proposed by Rokeach (1966b), Kluckhohn (1951) and others in which an individual's values are characterized as a particularly stable and enduring portion of his belief system.

Individual Differences in Value Stability

To this writer's knowledge, there have been no previous studies of individual differences in value stability. There have, however, been numerous studies of individual differences in general personality stability. Reviewing the literature concerned with personal stability and fluctuation, Cattell found "major evidence of personal unreliability" as a personality characteristic across a variety of trait measures (1957: 603). It would seem reasonable to assume that there also are significant individual differences in the stability of value systems--that some individuals possess highly stable personal values to which they are deeply and firmly committed, while others' values are less deeply rooted, more superficial and unstable. Data from the previous study using the Form A value scales (Rokeach, 1966a) indicate that this assumption may indeed be empirically justified. It was found that the distribution of individual value system stability scores ranged from -.34 to +.97, forming a somewhat skewed, inverted U-shaped curve. This distribution of scores could have been obtained either because of differences between individuals in value system stability or because of random error. A major aim of the present study is to estimate the extent to which each of these two

determining factors influences the variation in individual value system stability scores. To the extent that the range of stability scores found in the previous study is due to systematic factors "in the person," one would expect a high correlation between terminal and instrumental value system stability. An individual who has highly stable terminal values should also have highly stable instrumental values, if value stability operates as a general characteristic.

Determinants of Value System Stability

If there are significant individual differences in value system stability, then what personality characteristics or other explanatory variables may account for these differences? One personal characteristic which may function as a significant determinant of the stability of an individual's values is the extent to which he tends to be open-minded or closed-minded. According to Rokeach's theory of open and closed belief systems (1960), one of the characteristics of the relatively closed-minded individual is his need to reduce threat and allay anxiety, which leads to an overreliance on authority. While the relatively open-minded person will tend to evaluate incoming information on the basis of its cognitive correctness or logical consistency, and selectively accept or reject the influences of authority according to their congruence with his own beliefs, the behavior of the closed-minded individual "reflects a tendency not to relate beliefs to /his own/ inner requirements... but to assimilate them wholesale, as fed by /his/ authority figure" (1960: 61). Thus, in regard to his values, along with other types of beliefs, the closed-minded person may tend to accept *en bloc* the

values of his current authority figure, as he perceives them. In other words, the relatively closed-minded person is more likely to base his value judgments on a frame of reference provided for him by an external authority, while the open-minded person is more likely to possess a stronger, more independent, internalized system of values of his own.

To the extent that the closed-minded person's authority figure remains constant, his value judgments should remain relatively constant over time. However, most members of modern industrial society meet with constantly changing demands and exhortations from a wide variety of authorities. The closed-minded person, faced with this problem of multiple roles, may develop a number of isolated and relatively superficial sets of values, each one derived from a particular authority and used in his dealings with that authority. A closed-minded college student, for example, might behave according to differing sets of values in the presence of his parents, his peers, and his professors. In addition, he may be unaware or only dimly aware of the contradictions between these isolated sets of values and related behaviors. When he is asked, as in the present study, to respond to a value scale in terms of his own value preferences, without reference to any specific authority figure or reference group, he may experience difficulty and uncertainty in deciding the relative importance of various values in his life. To the extent that he lacks a strong and stable commitment to a single system of values, buttressed by an internalized, independent system of personal convictions, his responses to a value scale will be less stable or consistent from one occasion to another.

It has been suggested above that a relatively closed-minded person will experience a greater degree of difficulty and uncertainty in responding to the value scales. The degree of difficulty a person has and the degree of uncertainty he feels as he is ranking the twelve values can be seen as intervening or mediating variables between Dogmatism (or other influencing factors) and value system stability. Other mediating variables of this type are the degree of commitment a subject expresses toward his rank ordering of the values, the degree of importance he attaches to all the values in general, and the degree of vagueness he perceives the values to possess. It is expected that to the extent that an individual (1) has a high degree of commitment to a particular ordering of values, (2) feels that values in general are very important, (3) has clear and definite meanings for most of the values, (4) experiences little difficulty in ranking a set of positive values, and (5) feels little uncertainty about his rankings, his rankings of those values will be more stable over time.

Hypotheses

A. To the extent that the defining phrases and synonyms provide clarification of the meanings of the values used as stimuli, repeat reliability will be significantly increased.

B. Those values which a person ranks at both extremes on the scale will be significantly more stable over time than those ranked between the two extremes.

C. To the extent that value system stability operates as a general personality characteristic, there will be a significant correlation between terminal and instrumental value system stability.

D. To the extent that an individual tends to be closed-minded, his value system will be less stable over time.

E. To the extent that an individual expresses a high degree of commitment to the way he has ordered the values, his value system will be more stable over time.

F. To the extent that an individual feels that the values are very important to him, his value system will be more stable over time.

G. To the extent that an individual sees the values as vague or ambivalent to him, his value system will be less stable over time.

H. To the extent that an individual experiences a high degree of difficulty in ranking the values, his value system will be less stable over time.

J. To the extent that an individual expresses a high degree of uncertainty about the rankings he gave the values, his value system will be less stable over time.

METHOD

To test the above hypotheses, analyses were performed on two sets of data. The first of these, referred to below as Study I, consists of a set of data previously collected by Rokeach and others, using the Form A value scales (Rokeach, 1966a). The second set of data were collected by the present author, using the Form B value scales, and will be referred to as Study II. To test Hypothesis A, a comparison was made between relevant data from Study I and Study II. Hypotheses B and C were tested through analyses of both Study I and Study II, considered as replicates. Data pertaining to Hypotheses D, E, F, G, H and J were collected in Study II.

Study I

A questionnaire containing the two Form A value scales was administered to 444 introductory psychology students at Michigan State University in the spring of 1966.* Seven weeks later 210 of the students were retested for the purpose of obtaining test-retest reliability data.

Study II

In the fall of 1966, 440 introductory psychology students at Michigan State University were administered a questionnaire containing the two Form B value scales along with Troldahl and Powell's 20-item short form (1965) of the Dogmatism Scale (Rokeach, 1960). The Form B terminal and instrumental value scales, along with their instructions, are reproduced below.

---

*The Form A terminal and instrumental value scales are reproduced with their instructions in Appendix A.

Form B Terminal Value Scale

Below is a list of 12 values arranged in alphabetical order. We are interested in finding out the relative importance of these values for you.

Study the list carefully. Then place a 1 next to the value which is most important for you, place a 2 next to the value which is second most important to you, etc. The value which is least important relative to the others should be ranked 12.

When you have completed ranking all of the values, go back and check over your list. Please take all the time you need to think about this so that the end result is a true representation of your values.

_____ A COMFORTABLE LIFE (a pleasurable, successful life)

_____ A WORLD AT PEACE (a world free of war and conflict)

_____ A WORLD OF BEAUTY (beauty of nature and the arts)

_____ EQUALITY (brotherhood, equal opportunity for all)

_____ FREEDOM (independence, free choice)

_____ FRIENDSHIP (close relationship with another)

_____ NATIONAL DEFENSE (protection from attack)

_____ PERSONAL SECURITY (safe, free from worry)

_____ RESPECT FROM OTHERS (looked up to, admired by others)

_____ SALVATION (saved, eternal life)

_____ SELF-FULFILLMENT (developing myself fully)

_____ WISDOM (mature understanding of life)

Form B Instrumental Value Scale

Below is a list of another 12 values. Rank these in order of importance in the same way you ranked the first list on the preceding page.

_____ BROADMINDED (open-minded)

_____ CAPABLE (competent, effective)

_____ CAREFUL (cautious, prudent)

_____ CLEAN (neat, tidy)

_____ COURAGEOUS (standing up for your beliefs)

_____ INDUSTRIOUS (hard-working, ambitious)

_____ INTELLECTUAL (intelligent, rational)

_____ PATIENT (calm, willing to wait)

_____ POLITE (courteous, well-mannered)

_____ PRACTICAL (down-to-earth, realistic)

_____ TENDER (gentle, warm-hearted)

_____ TRUSTING (not suspicious, trustful of others)

Immediately after they had ranked the 12 terminal values and again after ranking the 12 instrumental values, the subjects were asked the following question designed to tap their degree of commitment to the way they had ordered the values:

> Now we are interested in knowing how you feel about the way you ranked these 12 values on the preceding page in general. Please circle one number on the following scale:

1	2	3	4	5	6	7	8	9
I care very much about the order in which I ranked these values.								It does not make much difference which order I put them in.

Seven weeks later 218 of the students were retested for the purpose of obtaining test-retest reliability data. On this occasion data were obtained on the degree of importance, vagueness, difficulty and uncertainty of the value scales. To obtain these data subjects were asked the following questions regarding each of the two value scales:

It is very important that we find out some of your reactions to this set of values. Think back for a moment about how you felt toward these values as you were ranking them.

1. How many of these values are very important to you?

ALL of them	MOST of them	MANY of them	SOME of them	A FEW of them	NONE of them

2. How many of these values are very unimportant to you?

ALL of them	MOST of them	MANY of them	SOME of them	A FEW of them	NONE of them

3. How many of these values are vague or ambivalent to you?

ALL of them	MOST of them	MANY of them	SOME of them	A FEW of them	NONE of them

4. How difficult was it to decide how to rank these values?

EXTREMELY difficult	VERY difficult	QUITE difficult	SOMEWHAT difficult	SLIGHTLY difficult	NOT AT ALL difficult

5. How uncertain do you feel about the rankings you gave the 12 values?

EXTREMELY uncertain	VERY uncertain	QUITE uncertain	SOMEWHAT uncertain	SLIGHTLY uncertain	NOT AT ALL uncertain

Individual value system stability coefficients were obtained by computing the Spearman rank-order correlation between each individual's initial rankings and his retest rankings of the twelve values. Thus a terminal and an instrumental value system stability coefficient was obtained for each individual. A perfect correlation of 1.00 would indicate that the individual ranked the twelve values in exactly the same order on both administrations of the test. To the extent that

his rankings differed from Time 1 to Time 2, his stability coefficient would be reduced. The sampling distribution of r_s (the Spearman rank correlation coefficient) is unimodal and symmetric, and approaches the normal form with N large, i.e., "at least greater than or equal to 10" (Hays, 1963: 646). Therefore, with N=12 in the present study, it was felt that the stability coefficients could justifiably be used as a set of scores to be correlated with measures of other variables.

RESULTS

The frequency distributions, medians, means and standard deviations of the two sets of individual value system stability scores are shown in Table 2. They range from -.34 to .99. Although a few individual's initial rankings of the values were negatively correlated or only slightly correlated with their retest rankings, the majority of the stability coefficients are substantially positive (beyond .60).

Results Pertaining to Hypothesis A

According to Hypothesis A, it was predicted that a defining phrase or synonyms would clarify the meaning of each value, thereby helping to stabilize an individual's responses to the values from one occasion to another. To test this hypothesis the reliabilities of the Form B value scales were compared with those of Form A. It was found that the Form B terminal value scale is not significantly more reliable than the Form A terminal scale, while the Form B version of the instrumental value scale is significantly less reliable than the previous form ($p < .01$, t test). Thus the hypothesis is not confirmed for terminal values, and is contradicted for instrumental values.

The significant decrease in stability of the Form B instrumental value scale may be partially explained by the results shown in Table 3. The Form B instrumental and terminal scales were compared with regard to stability, commitment, importance, vagueness, difficulty and uncertainty. First, the Form B instrumental value scale was found to be significantly less stable than the Form B

Table 2. Distributions of individual stability coefficients[1] for Form A and Form B terminal and instrumental values[2]

	Frequencies										
Interval range	-.4 to -.2	-.2 to 0.0	0.0 to .2	.2 to .4	.4 to .6	.6 to .8	.8 to 1.0	Total	Median	Mean	s.d.
Form A terminal	1	2	9	14	59	77	48	210	.65	.63	.22
Form A instrumental	2	4	11	24	42	85	42	210	.65	.60	.25
Form B terminal	3	3	3	26	42	74	67	218	.69	.64	.25
Form B instrumental	3	8	18	25	49	80	35	218	.62	.55	.28

[1]Spearman rank-order correlation between an individual's initial test and retest value rankings.

[2]The Pearson product-moment correlations between terminal and instrumental value stability are .20 for Form A and .24 for Form B, both of which are significantly different from 0 at the .01 level of probability.

Table 3. Differences between Form B terminal and instrumental values in degree of stability, commitment, importance, vagueness, difficulty, and uncertainty.

	Terminal		Instrumental		
	Mean	s.d.	Mean	s.d.	Difference
Stability	.64	.25	.55	.28	.09*
Commitment	3.78	2.01	3.38	2.07	.40*
Importance	2.71	1.56	2.67	1.67	.04
Vagueness	1.86	0.98	1.86	1.09	.00
Difficulty	3.43	1.41	3.74	1.39	-.31*
Uncertainty	2.95	1.21	3.35	1.24	-.40*

*$p < .01$, t test for correlated measures

terminal scale ($p < .01$, t test for correlated measures). Second, although there was no significant difference in the degree of importance and the degree of vagueness attributed to instrumental and terminal values, the instrumental values elicited significantly less commitment than terminal values. Finally, subjects found the instrumental values significantly more difficult to rank and were significantly less certain about their rankings of them. Although the data do not permit a direct comparison between the degree of difficulty, uncertainty and commitment of Form A and Form B, they do imply that the decreased reliability of the Form B instrumental scale as compared with terminal values may have been due to decreased commitment and increased difficulty and uncertainty of the Form B version. This interpretation of the results will be explored in the subsequent discussion of these and other results.

An alternative explanation for the decreased reliability of the Form B instrumental scale concerns the values which were deleted and those which were substituted in place of them in the change from Form A to Form B. Five values (Cooperative, Forgiving, Honorable, Responsible and Self-disciplined) were removed from the Form A instrumental scale because they had failed to discriminate between subgroups within the sample. Five new values (Capable, Careful, Industrious, Patient and Practical) were selected to replace them in the Form B revision. In considering possible reasons for the significant decrease in reliability from Form A to Form B, it was hypothesized that perhaps the values which had been removed from Form A were highly stable ones, while those which were substituted in Form B were highly unstable values. If this were true,

it would account for the difference in reliability between the two forms. To test this speculation, the stability of each of the Form A and Form B instrumental values was computed. The amount of change from Time 1 to Time 2 was computed for each of the twelve values, summing across all individuals and disregarding direction of change. The results of this analysis are shown in Table 4. Those values which were changed in the revision are starred. It may be seen that three of the five values removed from Form A were the three least stable values on that scale, while the other two were fourth and seventh in rank order of stability. Thus the speculation that the difference in the reliability of the two forms was due to the removal of stable values and the substitution of unstable ones is not supported. Rather, it appears in Table 4 that there was a general tendency for each of the Form A instrumental values to become less stable with the addition of the defining phrases and synonyms in the Form B revision. All but one of the six values which were identical on both forms (Tender, Polite, Courageous, Clean and Intellectual) were less stable on Form B than they were originally, while only one (Broadminded) became more stable.

Table 5 shows the results of the same analysis of the terminal value scale. In general, the range of value stabilities for Form A is very similar to that of Form B. Of the seven values which were identical on both forms, two (Equality and Wisdom) were more stable in the Form B revision, two (A world at peace and Freedom) were less stable, while three (Salvation, A comfortable life and Respect from others) remained exactly the same. The three new terminal values (A world of beauty, Self-fulfillment, and Personal security) were,

Table 4. Stability of each of the Form A and Form B instrumental values.[1]

Form A instrumental values	Median change[2]	Form B instrumental values	Median change
Tender	1.22[a]	Tender	1.56[b]
Courageous	1.41	Careful*	1.58
Polite	1.42	Polite	1.69
Responsible*	1.46	Broadminded	1.74
Clean	1.71	Trusting	1.97
Trustful	1.76	Courageous	1.99
Forgiving*	1.76	Capable*	2.02
Broadminded	1.80	Patient*	2.07
Intellectual	1.83	Intellectual	2.10
Honorable*	1.92	Clean	2.20
Self-disciplined*	1.98	Practical*	2.35
Cooperative*	2.06	Industrious*	2.40

[1]Listed in order from most to least stable.

[2]Median of the differences between initial test and retest, taken regardless of direction.

*Starred values are those which were changed in the revision from Form A to Form B

[a]The differences in amount of stability among the Form A instrumental values are significant beyond the .001 level of probability (Chi Square=33.59, df=11, median test for k related groups).

[b]The differences in amount of stability among the Form B instrumental values are significant beyond the .001 level of probability (Chi Square=39.21, df=11, median test for k related groups).

Table 5. Stability of each of the Form A and Form B terminal values[1]

Form A terminal values	Median change[2]	Form B terminal values	Median change
Salvation	0.48[a]	Salvation	0.48[b]
A comfortable life	1.56	National defense	1.27
National security	1.62	A world of beauty*	1.48
True friendship	1.65	Equality	1.50
A world at peace	1.67	A comfortable life	1.56
Freedom	1.67	Respect from others	1.72
Respect from others	1.72	A world at peace	1.77
Equality	1.74	Self-fulfillment*	1.81
A meaningful life*	1.79	Freedom	1.81
Maturity*	1.91	Wisdom	1.89
Respect for others*	1.96	Friendship	1.94
Wisdom	2.10	Personal security*	2.04

[1]Listed in order from most to least stable.

[2]Median of the differences between initial test and retest, taken regardless of direction.

*Starred values are those which were changed in the revision from Form A to Form B.

[a]The differences in amount of stability among the Form A terminal values are significant beyond the .001 level of probability (Chi Square=53.64, df=11, median test for k related groups).

[b]The differences in amount of stability among the Form B terminal values are significant beyond the .001 level of probability (Chi Square=58.40, df=11, median test for k related groups).

on the average, more stable than the three which they replaced (A meaningful life, Maturity, and Respect for others).

Results Pertaining to Hypothesis B

According to Hypothesis B, it was predicted that those values which a person ranks at both extremes on the scale would be more stable over time than those ranked between the two extremes. To test this hypothesis, the amount of change from Time 1 to Time 2 (regardless of direction) was computed for each individual's "number 1" value, whichever value that might be. The median of these changes was taken, as an estimate of the change from test to retest of value number 1. The same procedure was followed for values second through twelfth, generating a measure of the stability of values by initial rank position. The results of this procedure, applied to both the Form A and Form B terminal and instrumental values, are shown in Table 6. These same results are illustrated graphically in Figure 1 and Figure 2. Figure 1 shows that both the Form A and the Form B terminal values conform remarkably well to a curvilinear function. A significant difference was found between the stabilities of values according to their rank position both for Form A and Form B terminal and instrumental values.* It may be seen in Figure 2 that the instrumental values showed a weaker but nevertheless significant tendency toward the curvilinear relationship predicted. It may be noted that in the Form B revision, those values ranked first and

---

*X^2 values are 142.83 (Form A terminal); 71.76 (Form A instrumental); 124.13 (Form B terminal); and 45.33 (Form B instrumental). All four of these are significant beyond the .001 level of probability with 11 degrees of freedom (median test for k related groups, Mood, 1950).

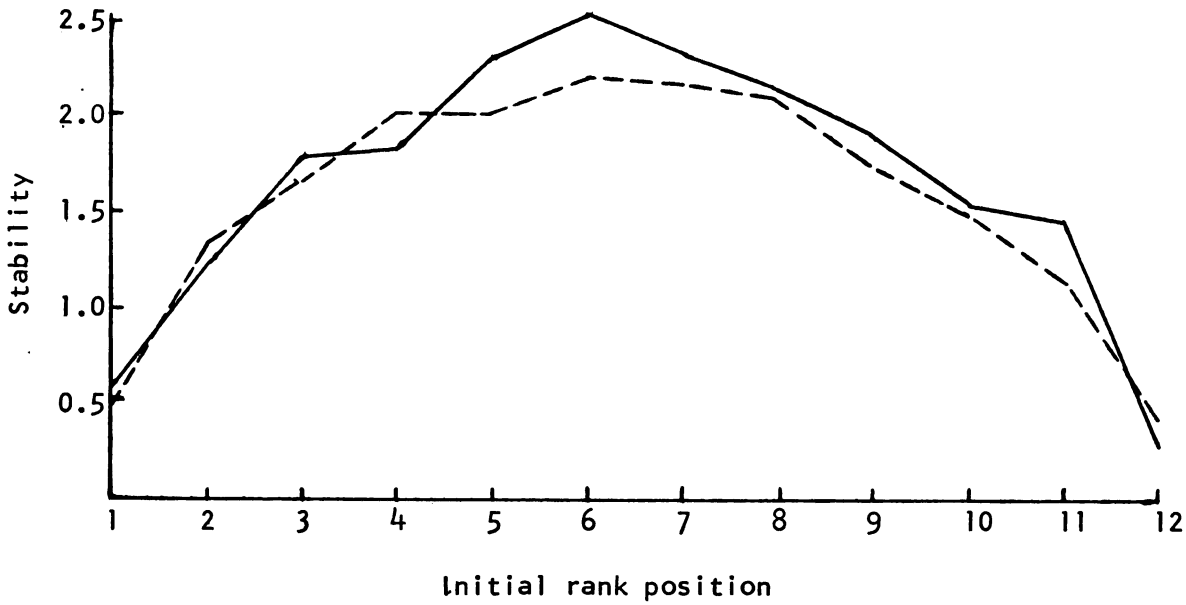
Table 6. Stability[1] of values by initial rank position.[2]

Initially ranked	Terminal Form A	Terminal Form B	Instrumental Form A	Instrumental Form B
1	0.55	0.50	1.17	1.16
2	1.23	1.29	1.33	1.38
3	1.78	1.65	1.58	1.97
4	1.82	2.00	1.93	2.11
5	2.32	2.00	2.14	2.28
6	2.52	2.19	2.00	2.40
7	2.38	2.18	2.14	2.52
8	2.13	2.14	2.05	2.28
9	1.92	1.75	1.96	2.28
10	1.55	1.52	1.52	1.91
11	1.48	1.18	1.20	1.40
12	0.33	0.42	0.92	1.42

[1]Median of the differences between initial test and retest, taken regardless of direction.

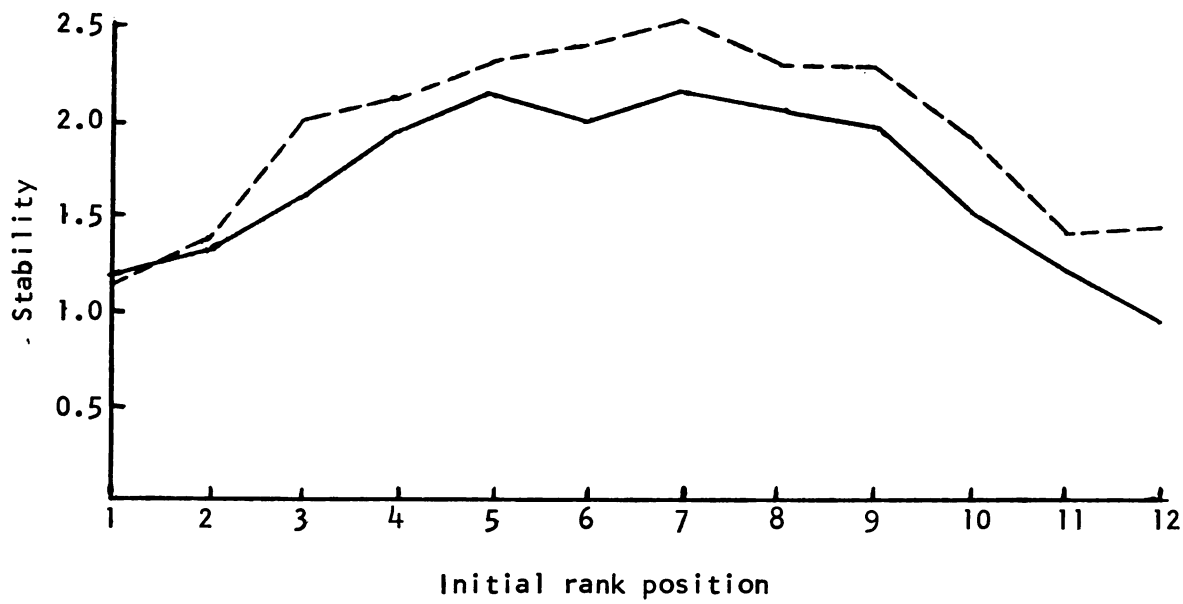
[2]Differences in stability among each of the four sets of values were found to be significant beyond the .001 level of probability, using the median test for k related groups.

Figure 1. Stability of terminal values as related to initial rank position[1]



[1]Solid line=Form A, broken line=Form B.

Figure 2. Stability of instrumental values as related to initial rank position[1]



[1]Solid line=Form A, broken line=Form B.

second were as stable as their Form A counterparts while those ranked at the lower extreme tended to be less stable.

Results Pertaining to Hypothesis C

According to Hypothesis C, it was predicted that there would be a significant correlation between stability of terminal and instrumental values, due to a general value system stability factor. To test this hypothesis a Pearson product-moment correlation was computed between the instrumental and terminal value system stability scores of the 210 subjects in Study I (Form A). This analysis was replicated using the 218 subjects in Study II (Form B). These correlations were found to be .20 for Form A and .24 for Form B, both of which are significantly different from zero at the .01 level of probability. In order to estimate the extent to which these correlations were affected by the unreliability of the test, each of the two correlations was corrected, using the formula for complete correction for attenuation (Guilford, 1954: 400). It was found that the "true" correlations between instrumental and terminal value system stability were estimated at .33 (Form A) and .40 (Form B). Using these corrected correlations, and taking the square of the correlation coefficient as an estimate of the proportion of variance in Y attributable to X, it may be seen that roughly 10% (Form A) to 16% (Form B) of the variance in instrumental value system stability is shared by variance in terminal value system stability, assuming a perfectly reliable test. Although this finding supports the presence of value system stability as a general characteristic, it appears

that this characteristic cannot alone account for differences in an individual's values from one occasion to another.

Results Pertaining to Hypothesis D

According to Hypothesis D, it was predicted that relatively closed-minded individuals' value systems would tend to be less stable over time that those of relatively open-minded individuals. To test this hypothesis, correlations between Dogmatism and terminal and instrumental value system stability were computed, using the data from Study II. In addition, the correlations between Dogmatism and each of the five "predictor" variables (commitment, importance, vagueness, difficulty, and uncertainty) were computed, both for terminal and for instrumental values. These results are shown in Table 7. In general, the relationships found are weak and insignificant. The only correlation which achieves statistical significance is the one between Dogmatism and importance of terminal values. Because the same relationship does not hold for instrumental values, it is likely that this barely significant result occurred by chance. The data were further analyzed to check whether any of these low correlations could be concealing a curvilinear relationship. However, this was not found to be the case.

Results Pertaining to Hypotheses E, F, G, H and J

According to Hypotheses E, F, G, H, and J, it was predicted that value system stability would be significantly positively related to commitment and importance, and negatively related to vagueness, difficulty, and uncertainty. To test these hypotheses intercorrelations (Pearson product-moment) were computed between

Table 7. Correlations between Dogmatism and stability, commitment, importance, vagueness, difficulty and uncertainty for terminal and instrumental values.[1]

Correlations between Dogmatism and:	Of terminal values	Of instrumental values
Stability	-.01	-.16
Commitment	.02	.08
Importance	.20	.07
Vagueness	-.01	.15
Difficulty	.06	.13
Uncertainty	.06	-.02

[1]Any correlation greater than .18 is significantly different from 0 beyond the .01 level of probability with N=218.

value system stability scores and ratings of commitment, importance, vagueness, difficulty, and uncertainty, for terminal and instrumental values (the responses to the two questions regarding "importance" and "unimportance" were combined to form an index of perceived importance). Table 8 shows that the correlations between value stability and each of the predictors (subscripted e, f, g, h, and j) are in general insignificant, though all of them are in the predicted direction. Other results shown in Table 8 provide some additional insights. First, those correlations which are subscripted n indicate systematic positive relationships (ranging from .42 to .58) between instrumental and terminal ratings on each of the five predictors. For example, if a subject expresses a high degree of commitment to his ordering of the terminal values, he is also likely to be strongly committed to his instrumental value rankings (r=.58). Second, those correlations subscripted k show that there is a highly positive relationship between the degree of perceived difficulty and the degree of felt uncertainty, both for terminal and instrumental values (r=.64 and r=.75).

Table 8. Intercorrelations[1] between value system stability, commitment, importance, vagueness, difficulty, and uncertainty, for Form B Terminal and Instrumental Values.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Stability (T)[2]		.14$_e$	.00$_f$	-.05$_g$	-.11$_h$	-.18$_j$	.24	.03	-.11	.05	-.05	-.02
2. Commitment (T)			.07	-.08	-.24	-.21	.06	.58$_n$	.05	-.01	-.07	-.09
3. Importance (T)				-.38	.24	.04	.12	-.17	.42$_n$	-.19	.03	.03
4. Vagueness (T)					.08	.25	-.07	.07	-.14	.44$_n$	.06	.09
5. Difficulty (T)						.64$_k$	.01	.08	.27	.01	.55$_n$	.33
6. Uncertainty (T)							.01	.10	.08	.08	.45	.45$_n$
7. Stability (I)								.15$_e$	-.04$_f$	-.06$_g$	-.15$_h$	-.11$_j$
8. Commitment (I)									.19	-.06	-.11	-.20
9. Importance (I)										-.36	.23	.03
10. Vagueness (I)											.22	.35
11. Difficulty (I)												.75$_k$
12. Uncertainty (I)												--

[1]Any correlation greater than .18 is significantly different from 0.0 beyond the .01 level of probability with N=218.

[2](T)=terminal, (I)=instrumental

DISCUSSION

Stimulus Ambiguity and Test Reliability

It would appear that efforts to increase stimulus clarity resulted in decreased rather than increased reliability of response. The defining phrases and synonyms which were added in the Form B revision not only failed to increase reliability (in the case of the terminal values), but apparently resulted in a significant decrease in reliability (of instrumental values). This result runs counter to expectations concerning stimulus ambiguity and response stability. Of course, it may be argued that the significant decrease in reliability of the instrumental scale was a "rare event," was due to sampling error, or a result of unknown contaminating influences. However, there are grounds for concluding that the significant decrease in reliability was a direct result of the added defining phrases. The reasoning behind this post hoc conclusion is as follows: In the original Form A version, each value was presented as a single word or phrase, with no definition provided. These relatively ambiguous stimuli may have functioned somewhat as a projective test, in allowing the subject considerable freedom to project onto them his own meanings and interpretations. We may assume that for each subject, several values were interpreted more or less unfavorably--i.e., were given negative definitions. Faced with the difficult task of discriminating between twelve stimuli, a subject would almost inevitably interpret some values negatively in order to resolve the dilemma of deciding how to rank them. In the Form B version, however, the subject is faced with a set of values, each of which has been given a positive, favorable definition.

The defining phrases provided for each value may prevent him from interpreting some values unfavorably, thereby rendering more difficult his task of rank ordering them in terms of favorability. To the extent that the clarification of formerly ambiguous stimuli results in a reduction of the psychological distance between stimuli, discriminations between them will be less stable. If the psychological distance between stimuli were increased, for example, by adding the words Disrespectful and Cowardly to the instrumental value scale, discriminations between these stimuli and the others would no doubt be highly reliable. One may reasonably assume that most people would rank these two "values" 11th and 12th, respectively, in comparison with the other instrumental values. The extreme stability of these two values would increase the overall reliability of the instrumental value scale. By contrast, the addition of defining phrases to clarify the meanings of the values, by reducing the psychological distance between them and making discrimination more difficult, may tend to reduce the repeat reliability of the value ranking scales, other factors being equal.

Value Stability and Hierarchical Position

For both terminal and instrumental values, it was found that those values which are very high or very low in an individual's value hierarchy are more stable than those in between. This result may be interpreted as an indication that people possess a clearer conception of which values they consider extremely important or extremely unimportant. It may also be interpreted as an indication of an underlying "normal distribution" of psychological distances between values in a hierarchy, with intervals becoming larger as

the distance from the center of the distribution increases. This result would be expected for a variety of reasons, and a failure to find it would be surprising.

Value System Stability as a General Personal Characteristic

It was estimated that approximately 10 to 16% of the variance in instrumental value system stability could be accounted for by variance in terminal value system stability, if error variance due to test unreliability were removed from both measures. Apparently, value system stability as a general characteristic of an individual's terminal and instrumental values may be tentatively considered as a meaningful concept. However, a considerable portion of variance remains to be explained. Two possible explanations may be offered to account for the relatively low correlations found between terminal and instrumental stability: First, it is possible that some individuals have highly stable terminal values but relatively unstable instrumental values, while for others the reverse is true. Second, it is possible that differential systematic changes took place during the seven-week time interval between test and retest, due to influences from the individuals' environment which changed one or more of a person's terminal values without affecting his instrumental values, or vice versa. Neither of these explanations seems particularly plausible from any theoretical standpoint. At worst, it may be that there simply are random differences between individuals in the relation between instrumental and terminal value stability. At best, it must be said that further research is needed, utilizing more reliable tests of values.

Predictors of Differential Value System Stability

No relation was found between open and closed-mindedness and value system stability. It had been argued that the relatively closed-minded person, having acquired multiple superficial cognitively isolated sets of values from his various authority figures, would be less able to reliably describe his own value system from one occasion to another. There may be several alternative reasons why the data fail to support this hypothesis. First, it may be that many closed-minded individuals solved the dilemma of multiple value systems by simply reacting to the value-ranking task in a way they felt would be approved by the professor, whom they knew to be the recipient of their test responses. If this were true, then the psychological presence of the experimenter on both testing occasions would tend to minimize whatever instability might have arisen due to conflict between contradictory role-identifications. Second, it is possible that the hypothesized effect did in fact occur to some extent, but was contravened by other effects such as a tendency for open-minded subjects to be more open to mind-changing influences from their environment during the time intervening between test and retest. If this were true, one would expect to find two different types of "instability," i.e., systematic changes in values in the open-minded subjects and random fluctuations among the relatively closed-minded subjects. Finally, it must be noted that the hypothesis is actually an inference from an untested theoretical assumption. The extent to which relatively closed-minded individuals do in fact tend to possess multiple, superficial, cognitively isolated sets of values is an interesting

empirical question in and of itself, and has yet to be investigated. The extent to which this condition, if true, would affect value system stability can be seen as a separate issue based on the former one. Further research on the structure of value systems may succeed in providing evidence to resolve the above questions and speculations.

Relationships between value system stability and the five predictor variables (commitment, importance, vagueness, difficulty, and uncertainty) were found to be very slight. The fact that reasonably high correlations were found between terminal and instrumental predictors indicates that these variables apparently were measured with some degree of reliability. None of them, however, was found to be significantly related to value system stability. Evidence presented earlier suggests that this may have been partially due to error variance in the instrument used to measure value system stability, namely, the value ranking scales. The low reliabilities (median stability coefficients) of the ranking scales and the low correlation between terminal and instrumental value system stability indicate that individual differences in value stability were more a function of momentary fluctuations or random error than of differences along these five dimensions or other systematic influences.

Reliability of the Value Scales

One of the major conclusions implied by the results of this study pertains to the relative instability of the value ranking scales. Apparently the inherent difficulty of ranking twelve positive stimuli produces a considerable degree of fluctuation in rank

orderings over time. In the theory which guided this research it is assumed a priori that individuals possess a relatively stable and enduring hierarchical ordering of values. This assumption seems justified by theoretical considerations as well as by common sense. This poses the problem, then, of explaining why the findings do not fit the theory. First, it may be that individuals' value decisions in everyday life are frequently made "instinctively" without conscious articulation of value criteria. If this is the case, then when confronted with the task of ranking the values themselves, expressed as abstract concepts, the individual may lack a previously thought-out, articulated "value system." Second, in real-life value decisions, much depends on the situation in which the decision is faced, as a source of definitions of the specific value criteria to be considered. If this is true, then the values, presented situation-free, may seem to some extent meaningless or may seem to possess multiple meanings which bedevil the judgment process. Third, the theoretical assumption of the existence of an ordered "value system" seems logically to presuppose that at some point in his previous history the individual has had to face, at least once, a conflict between each of the possible pairs of values, in order to establish which one is more important. In the case of a "personal" value such as Wisdom and a "societal" value such as A world of beauty, such a conflict may arise very rarely. In other words, individuals may possess separate subsets of values, each subset internally ordered but not including explicit relations between every member of subset A (personal) with every member of subset B (societal). If so, then asking subjects

to rank both "personal" and "societal" values together on the same scale may, to some extent, be a strange and unfamiliar task, similar to ranking kinds of vegetables and flavors of ice cream on a single scale. To the extent that this is true, rankings of the values would tend to be unstable from one occasion to another. Finally, the overall instability of the ranking scales may have been affected by restrictions in both the range of stimuli and the range of subjects used in these studies. One of the criteria for selection of the values used as stimuli was the elimination of values uniformly ranked high or low, because they failed to differentiate between various types of groups. As argued earlier, this restriction of the psychological distance between stimuli to a narrow range would be expected to produce instability of stimulus rankings from one occasion to another. Further, the use of college students as subjects sharply restricted the range of ages sampled. It may be that college students (especially students in a beginning-level psychology class, the majority of whom are freshmen) are a relatively unstable group, compared to their elders. If it were found that older subjects' value rankings remain highly stable over a seven-week interval, then the conclusions drawn from the present study would concern the instability of college students' values more than the unreliability of the value ranking scales. Further research will be aimed at overcoming these two restrictions in the present data.

One possible solution to the problem of instability would be to abandon the ranking scales entirely and seek to measure values by the "indirect" technique used by Allport and Vernon and others. Another would be to radically increase the number of items and utilize a

card-sorting task, a la Rosenberg. Both of these alternatives have their drawbacks--the validity of the "indirect" technique depends on the validity of items chosen to represent values, and the card-sorting technique is time-consuming and logistically cumbersome. Another approach would be to attempt to "improve" the ranking scales themselves, by increasing the number of items, separating the present scales into subscales, resorting to a paired comparisons technique, etc. The answer to the stability-instability problem inevitably lies in the experimenter's choices of criteria for what does and does not constitute sufficient precision of measurement. These choices, in turn, depend on the purposes for which the data are intended. Although responses to the Rorschach test, for example, vary considerably from one occasion to another, nevertheless the test provides valuable information and is widely used. On the other hand, in studies which depend on correlational statistics, a "reliability coefficient" of .80 or better is desired and sought. The value ranking scales have already proven capable of discriminating gross differences between various groups (Rokeach, 1966a). Results of the present study indicate that they are not sufficiently precise as a technique for measuring and predicting individual differences in value system stability--using correlational statistics. However, to expect these simple ranking scales to achieve such precision is in a sense like using a simple kitchen measuring cup in a chemistry laboratory--it is forcing them to serve a purpose for which they were never intended.

Conclusion

Returning to the series of questions posed in the introduction,

it appears that this study has provided at least a tentative answer for each one:

1. How reliable are the value scales, as a measuring instrument? It may be concluded that although the ranking scales provide data of sufficient reliability to discriminate between groups, they are not reliable enough to warrant the use of correlational techniques.

2. To what extent is their reliability influenced by factors in the test or by factors in the person? It appears that variations in the stability of value rankings are chiefly a function of the instability of the ranking scales, and only slightly influenced by factors in the person.

3. Are some values more reliable than others; if so, what are some determinants of differential stability among values? It was demonstrated that one powerful determinant of a value's stability was its position in a hierarchy--those high or low being more stable than those toward the middle.

4. Are some people more reliable than others; if so what variables may predict individual differences in value system stability? Value system stability, as a general personal characteristic, was found to be present to a slight but significant degree. No conclusions can be drawn regarding the attempts to predict it. Due to the presence of excessive measurement error, the finding of nonsignificant relationships does not permit us to accept the null hypothesis--i.e., to conclude that the various predictor variables are not related to value system stability.

In addition to providing these and other answers, this research has raised many questions for further inquiry. Among these are the

continuing search for ways to improve the reliability of the ranking scales, and further attempts to discover personality variables related to value system stability. Woodruff (1942) and others have called our attention to the fact that increased social psychological knowledge of people's values and value systems may help us to better understand the functions which attitudes serve in an integrated personality system. One realm in which further systematic knowledge seems particularly needed is the study of individual differences in value system stability. It is hoped that this study will lead to further inquiry into this feature of human values.

REFERENCES

Allport, G.W., Vernon, P.E., and Lindzey, G. A Study of Values. revised ed. Boston: Houghton, Mifflin, 1960.

Allport, G.W., and Odbert, H.S. "Trait Names: A psycholexical study," Psychological monographs, 1936, 47, 1-171.

Anderson, H. H. Likeableness Ratings of 555 Personality-Trait Adjectives. Unpublished manuscript. University of California, Los Angeles, June, 1964.

Beech, R.P. "Value Systems, Attitudes, and Interpersonal Attraction." Unpublished Ph.D. Thesis, Michigan State University, 1966.

Cantril, H., and Allport, G.W. "Recent applications of the Study of Values. J. abnorm. & soc. psychol., 1933, 28, 259-273.

Cattell, R.B. Personality and Motivation Structure and Measurement. Yonkers-on-Hudson, New York: World Book, 1957.

Cohen, J.B. "A scale for the measurement of attitude toward the aesthetic value." J. Psychol., 1941, 12, 75-79.

Duffy, E. "A critical review of investigations employing the Allport-Vernon Study of Values and other tests of evaluative attitude." Psych. bull., 1940, 37, 598-612.

Dukes, W.F. "Psychological studies of values," Psych. bull., 1955, 52, 24-50.

Glaser, E.M. and Maller, J.B. "The measurement of interest values. Charac. & Pers., 1940, 9, 67-81.

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

Harding, L. W. "A value-type generalizations test. J. soc. Psychol., 1944, 19, 53-79.

Hays, W.L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Hilton, T.L. and Korn, J.H. "Measured change in personal values," Educ. psychol. Measmt., 1964, 24, 609-622.

Hunt, A.McC. "A study of the relative value of certain ideals." J. abnorm. soc. Psychol., 1935, 30, 222-228.

Jacob, P.E. Changing Values in College. New York: Harper and Brothers, 1957.

Kelly, E. L. "Consistency of the adult personality," Amer. Psychologist, 1955, 10, 659-681.

Kluckhohn, C. "Values and value-orientations in the theory of action." pp. 388-433 in Parsons, T. and Shils, E.A. (eds), Toward a General Theory of Action. Cambridge: Harvard University Press, 1951.

Lovejoy, A.O. "Terminal and adjectival values." Journal of Philosophy, 1950, 47, 593-608.

Lurie, W.A. "A study of Spranger's value types by the method of factor analysis." J. soc. Psychol., 1937, 8, 17-37.

Mood, A. Introduction to the Theory of Statistics. New York: McGraw-Hill, 1950.

Murray, H.E. Explorations in Personality. New York, Oxford University Press, 1938.

Ohnmacht, F.W. "Factor analysis of ranked educational objectives: an approach to value orientation," Educ. psychol. Measmt., 1965, 25, 437-447.

Rokeach, M. "Attitudes, Values, and Political Behavior," paper read at the meetings of The American Political Science Association, New York City, September 9, 1966a.

Rokeach, M. The Open and Closed Mind. New York: Basic Books, Inc., 1960.

Rokeach, M. "The nature of attitudes." To be published in the International Encyclopedia of the Social Sciences. 1966b.

Rokeach, M. "A Theory of Organization and Change Within Value and Attitude Systems," Presidential Address, Society for the Psychological Study of Social Issues, to be presented at the annual meeting, Washington, D. C., September 3, 1967.

Rosenberg, M. J. "Cognitive structure and attitudinal affect," J. abnorm. soc. Psychol., 1956, 53, 367-372.

Rosenberg, M.J. "Cognitive reorganization in response to the hypnotic reversal of attitudinal affect," J. Pers., 1960a, 28, 39-63.

Rosenberg, M.J. "An analysis of affective-cognitive consistency." pp. 15-64 in Rosenberg, M.J., Hovland, C.I., et al. Attitude Organization and Change. New Haven: Yale University Press, 1960b.

Rothney, J.W.M. "Evaluative attitudes and academic success." J. Educ. Psychol., 1936, 27, 292-298.

Smith, M.B. "Personal values as determinants of a political attitude." J. Psychol., 1949, 28, 477-496.

Spranger, E. Types of Men. (Trans. of 5th German ed. by P.J.W. Pigors.) Halle: Niemeyer, 1928.

Tisdale, J.F. "Psychological Value Theory and Research: 1930-1960," Unpublished Ph.D. Thesis, Boston University, 1961.

Troldahl, V.C., and Powell, F.A. "A short-form dogmatism scale for use in field studies," Social Forces, 1965, 44, 211-214.

Van Dusen, A.C., Wimberly, S. and Mosier, C.I. "Standardization of a value inventory." J. Educ. Psychol., 1939, 30, 53-62.

Vernon, P.E. and Allport, G.W. "A test for personal values," J. abnorm. and soc. Psychol., 1931, 26, 231-248.

White, R.K Value-Analysis, the Nature and Use of the Method. Glen Gardner, N.J.: Libertarian Press, 1951.

Whitely, P.L. "The constancy of personal values," J. abnorm. soc. Psychol., 1938, 33, 405-408.

Wickert, F. "A test for personal goal-values." J. soc. Psychol., 1940a, 11, 259-274.

Woodruff, A.D. "Personal values and the direction of behavior," Sch. Rev., 1942, 50, 32-42.

APPENDIX A

Form A Terminal Value System Scale and Instructions

Below is a list of 12 values arranged in alphabetical order. We are interested in finding out the relative importance of these values for you. Study the list of values below carefully. Which of these values do you feel to be the most important for you? Place a 1 on the blank line to the left of this value. Now, cross this value off your list and look carefully at the remaining 11 values. Which one of these values is second most important for you? Place a 2 etc. Cross this value off your list and look carefully at the remaining 10 values. Place a 3 etc. Now, rank all of the remaining values in order of importance to you. The value which is least important, relative to the others, should be ranked 12. When you have completed ranking all of the items, go back over your list to make sure they are in the proper order.

_____ A comfortable life

_____ A meaningful life

_____ A world at peace

_____ Equality

_____ Freedom

_____ Maturity

_____ National security

_____ Respect for others

_____ Respect from others

_____ Salvation

_____ True friendship

_____ Wisdom

Form A Instrumental Value System Scale and Instructions

Below is a list of another 12 values. Rank these in order of importance, in the same way you ranked the first list on the preceding page.

_____ Broadminded

_____ Clean

_____ Cooperative

_____ Courageous

_____ Forgiving

_____ Honorable

_____ Intellectual

_____ Polite

_____ Responsible

_____ Self-disciplined

_____ Tender

_____ Trustful