A RELIABILITY TEST OF THE TWENTY STATEMENTS TEST

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY

Carole Overmier Bettinghaus

1966

ABSTRACT

A RELIABILITY TEST OF THE
TWENTY STATEMENTS TEST

by Carole Overmier Bettinghaus

The Twenty Statements Test (TST) is an instrument designed to access self-identification in a manner consistent with the "self" theories of George Herbert Mead. As such, its utility is limited to research which approaches the study of social psychology from the vantage point of sociology.

To administer the TST, the researcher asks a respondent to:

> Write twenty answers to the simple question "Who am I?" in the (twenty numbered) blanks. Just give twenty different answers to this question. Answer as if you were giving the answers to yourself, not to somebody else. Write the answers in the order that they come to you. Don't worry about logic or "importance". Go along fairly fast, for time is limited.

This open-ended question acts as a stimulus for a set of responses from each respondent. Each set of responses is called a protocol. Every protocol is content analyzed on the basis of a coded set of discrete categories.

The present study attempts to measure the reliability among six coders in the task of coding 150 such protocols. The 150 protocols were sampled randomly from a population of 1,528 protocols. The 1,528 protocols were obtained from respondents in a modified area probability sample of the United States general public (outside of institutions), age 21 or older.

The code applied to the protocols contained 46 discrete categories, and the coding task was done under independent conditions. Scores of reliability (range: from 0 to 10) were obtained for each response and each protocol. These reliability scores were based upon a combined scale of consistency (range: from 1 to 6), and agreements (range: from 0 to 15).

The basic hypothesis was: "there will be a high level of reliability among the six coders in the coding of the TST protocols, and responses." "High level of reliability" was defined as (1) 75 percent of all protocols will have a mean score of eight or better, and (2) 75 percent of all responses will have a score of eight or better, or the mean score of all responses will be eight or better. Results obtained were: (1) 72 percent of all protocols had a mean score of eight or better, (2) 72 percent of 858 responses had a score of eight or better, and the mean score of all responses was 8.19.

A set of sub-hypotheses was also tested. Sub-hypothesis one was: "when the difference between consensual responses and evaluative-consensual responses is overlooked, reliability is expected to appear as greater than under realistic conditions." Sub-hypothesis two was: "when differences between coders in the decision of whether to divide a statement into more than one statement for coding purposes are adjusted so that only the immediate statement is affected rather than all subsequent statements in the protocol, reliability is expected to appear as greater than under realistic conditions." Both of these hypotheses were supported. The suggestion is made that use of coders with training in grammar would contribute to an increase in inter-coder reliability.

Sub-hypothesis three was: "the more statements made on a TST protocol the lower will be the mean reliability score for that protocol." This hypothesis was supported. The conclusion was that reliability will be enhanced when fewer than 20 responses are sought.

Sub-hypothesis four was: "as content categories increase in complexity, reliability will lessen." This hypothesis was not supported.

Sub-hypothesis five was: "there will be no relationship in the pattern of coder agreements which contribute to unreliability." This hypothesis was used to operationalize the assumption made in the study that the coders were independent. The test of this hypothesis may be thought of as a validity check on the reliability study. Under one condition the hypothesis was rejected. This suggested that enhanced independence of coders in future studies may be expected to lower measured reliability.

Finally, some potentially useful forms of the TST are suggested: (1) structured and standardized with factor loadings for individual items, (2) a Q-sort form for self-styling by the respondent, and (3) unchanged, but with provision for self-coding by the respondent.

A RELIABILITY TEST OF THE TWENTY STATEMENTS TEST

By

Carole Overmier Bettinghaus

A THESIS

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

MASTER OF ARTS

Department of Sociology

1966

ACKNOWLEDGMENTS

I wish to express my appreciation to my advisor, Dr. Hideya Kumata, who guided me through the report writing stage of this research; and to the other members of my committee, Dr. Frederick Waisanen, and Dr. Mason Miller. During Dr. Kumata's absence from the country, Dr. Robert Stewart served as my advisor, and it is he to whom I am grateful for aiding me in the formative stages of the research.

The data used to check inter-coder reliability in this study, as well as the coding of the Gallup coder, were graciously made available to me by Dr. Charles Loomis, Dr. Kumata, Dr. Stewart, and Dr. Waisanen. I wish to thank them for this consideration.

The four individuals, other than myself, who recoded the data were Clark McPhail, Dr. Stewart, Charles Tucker and Dr. Erwin Bettinghaus. I thank them for conscientious attention to detail in a time-consuming, somewhat non-stimulating task. Any aspersions cast on their grammatical competence in Chapter IV, Conclusions, should be discounted. Grammar is not their recognized area of competence or training; and one would not expect those in the role of social scientist to exhibit the level of grammatical sophistication needed here.

My sincere thanks go to Mrs. Shirley Sherman, who typed the final manuscript for this thesis, and was a great help with many formal details.

Finally, I wish to thank all the individuals, within my acquaintance, whose consistent expectations have been that this research and report either should be completed, or had already been completed. This "generalized other" has had great influence on the personal perseverence which I brought to the task.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

BACKGROUND

Introduction

Any open-ended item in an interview schedule depends upon its code, rather than its form, for reliability. Therefore, this study is designed to test a comparatively complicated, formal code system for the content analysis of responses to a fairly simple, informal question, "Who am I?", asked as if the respondent were soliloquizing (16, p. 69). The test is one of reliability.

Reliability is operationalized as very good agreement between six individual coders in the task of coding identical sets of responses to the above question. Following explanations shall refer to each set of responses as a protocol, and to the question as one form of a self identification problem or SIP (23, p. 74). This specific SIP is known as the Twenty Statements Test, or TST. Thus, a TST-SIP protocol is understood to be a set of responses made by an individual respondent to the question "Who am I?" At times the six coders may be referred to in pragmatically equivalent terms as experimental subjects. A discussion of the high level of reliability in Chapter II will deal with the concise operational definition of the above phrase "very good agreement".

In this chapter, I will (1) discuss the need for a SIP which creates minimum bias in operationalizing Meadian self-theory within any experimental design context, (2) review the characteristics of several types of SIP, (3) discuss the history of the TST-SIP which asks a direct question, while creating a non-directive bias, and (4) review previous

investigations into the reliability of content analyses.

Relation of Self Identification Instrument to Meadian Theory

From the standpoint of a researcher in behavioral science, any verbalization designed to elicit communication from a respondent is a tool of the trade, or an instrument. From the vantage point of an individual respondent such a verbalization creates or increases the imbalance between work required and work completed. That is, for the respondent, it creates a task or problem to deal with. Since it is the same verbalization from different standpoints, we will adopt the practice of using the term "self identification instrument" as interchangeable with SIP. A SIP is, then, an instrument for eliciting communication from a respondent about the concept of self; specifically, what objects he identifies or associates self with as well as what objects he identifies or defines self as.

The concept of self is central to the self-role theories of George Herbert Mead. He held that the self mediated between society and the individual and allowed the development of the mind. The self was composed of two phases, the "I", and the "Me". The "I" is the impulsive tendency of the individual which begins social acts and gives propulsion to them. The "Me" comprises the incorporation of other individuals within each individual. It gives direction to each social act, and is the final state of a social act. The "Me" or object phase of the self is formed through the process of role-taking (26, p. 15). Kuhn tells us that Mead saw the self as carrying on an internal conversation between the "I" and the "Me" (14, p. 629). Meltzer interprets Mead's

belief that "the ability (of the "I" phase of self) to act toward oneself (in the "Me" phase) makes possible an inner experience which need not reach overt expression ... which constitutes mind." (26, p. 18)

Others, especially Charles Horton Cooley, a contemporary of Mead, and Gordon Allport, a later reviver of Meadian theory, have added their contributions to self-theory (14, p. 629). Cooley characterizes self as feeling or primitive emotion and claims that the feeling of self "is, perhaps, to be thought of as a more general instinct, of which anger, (fear, grief, etc.) etc. are differentiated forms..." (6, p. 171).

Allport, speaking of the study of personality says,

> ...all psychological functions commonly ascribed to a self or ego, must be admitted as data in the scientific study of personality. These functions are not, however, co-extensive with personality as a whole. They are rather the special aspects of personality that have to do with warmth, with unity, with a sense of personal importance...I have called them "propriate" functions. If the reader prefers, he may call them self-functions, and in this sense self may be said to be a necessary psychological concept (1, p. 55).

Of social psychology Allport says,

> Scholars interested in culture and personality deal primarily with the function of ego-extension, for their task is to account for the process of socialization (1, p. 57).

Of the "propriate" function of ego-extension he has this to say,

> Soon the process of learning brings with it a high regard for possessions, for loved objects and later for ideal causes and loyalties. We are speaking here of whatever objects a person calls "mine." They must at the same time be objects of importance, for sometimes our sense of "having" has no affective tone and hence no place in the proprium. As we grow older we identify with groups, neighborhood, and nation as well as with possessions, clothes, home (1, pp. 41 ff.).

However, Allport evinces a view contrasting to Mead's, that the self directs and controls an individual's behavior, gives rise to the mind, and is necessarily covert (26, pp. 11 & 18). Allport contends,

> What is unnecessary and inadmissable [as data in the scientific study of personality] is a self (or soul) that is said to perform acts, to solve problems, to steer conduct, in a trans-psychological manner, inaccessible to psychological analysis (1, p. 55).

In addition to the views of self put forth by Allport and Cooley, several changes have crept into Meadian theory through interpretation. The self has come to be thought of not as the locus for an internal conversation, nor as an entity carrying on a conversation with itself and within itself, but, rather, as embodied in that conversation; or carried to an extreme form, the self is thought of merely as a reflected image of significant or generalized others (14, p. 629).

Kuhn, in pointing to this trend on the part of others (14, pp. 156 & 198) (11, p. 316), makes no mention of his own like tendency in the development of the TST, the instrument dealt with in this study. The TST, is purported to elicit communication from an individual about his self. In operation it uses a cue verbalization of a form approximating what one might imagine the "I" asking of "Me." (Who am I? - as if I were asking and answering myself, with no other involved.) This elicits communication of a "stream-of-consciousness" nature, directed toward the concept "I".

Now Kuhn & McPartland found that responses to this question could be divided into two gross categories which seem to fit the needs of Meadian theory (1) consensual responses which "refer to groups and classes whose limits and conditions of membership are matters of common

knowledge, (16, p. 69) " and (2) subconsensual responses which "refer to groups, classes, attributes, traits or any other matters which would require interpretation by the respondent to be precise or to place him relative to other people." (16, p. 69) It seems to me that the former category might be characterized as statements one could expect to be made by the "Me," or internalized other phase of the self, while the latter category might be characterized as statements one could expect to be made by the "I," in disagreeing retort to the "Me," or in simple interjection.

Indeed the finding that "respondents tended to exhaust all of the consensual references they would make before they made any subconsensual ones," (16, p. 70) points to an order in which the "Me" responds to the cue question with consensual responses, followed by a rejoinder from the "I" with subconsensual statements. Perhaps Waisanen also viewed subconsensual responses as related to the "I," for having accounted for some sub-consensual responses as evaluative consensual (responses of a consensual nature, but which included modifiers) he termed all remaining sub-consensual responses "idiosyncratic."* Dictionary definitions of idiosyncracy characterize it as a queer, peculiar, unusual or individual way of acting or expressing oneself (34, p. 310) (2, p. 599), while Mead characterizes the "I" as saying "something that was novel to himself." (25, p. 197)

Now it is apparent that the data elicited by the TST is data from the internal conversation which Mead refers to, rather than data

*This category system is attributed to Waisanen by Dyer (10, p. 39)

from the self. Furthermore, because the subsequent analysis of data obtained using the TST has largely neglected the sub-consensual responses, the "I" portion of the data has lain dormant. In essence then, Kuhn has, with good intention, created an instrument which, effectively, treats the self only as a reflected image of significant or generalized others.

It is indeed very likely, that what Allport saw as a self inadmissable to scientific study was as much a matter of convenience as anything. I see no useful way of getting at data which is internalized below the level of language. However, more rigorous treatment of the TST's subconsensual data may bring greater insight into the nature of the "I" phase of the self.

As we shall see shortly, there are a great many instruments designed as SIP's. Most of them have been used to measure the self in the manner of the psychologically oriented social psychologist, as a concept deduced from experimental data, rather than a self in the manner of the sociologically oriented social psychologist, as a concept induced from Meadian theory. (14, p. 629)

The very fact that so many self identification instruments are extant, documents a great need for an adequate SIP. The TST with its direct approach to the internal conversation of gestures comes closer than perhaps any other current instrument to allowing operationism of Meadian theory. Furthermore, the direct approach to data overcomes the bias of inferential research which attempts to deduce the nature of the self from non-verbal behavior or from tangential verbal behavior. (16, p. 68) Kuhn also strove to minimize the effect of directive or structural bias. (12, p. 243) There is evidence that even when the

TST is administered in a structured context respondents are not led to make responses reflecting that context. (10, p. 22)

However, the TST entails other biases; (1) a self-selection bias on the part of those respondents who choose to (a) respond volubly or (b) respond minimally, and (2) a post factum structural bias, dependent upon what form of content-analytic code the data is forced into. The former bias produces non-equivalent numbers of responses between protocols, making any rigorous statistical treatment impossible. However, if the TST incorporated any manner of intensity measurement, which it does not, this lack of equivalence would be less serious. The TST is an SIP designed to overcome experimental bias, and as such it has sacrificed a great deal of experimental utility.

Review of types of self-identification instruments

Wylie (35) has reviewed self-identification instruments at great length, including measures of both phenomenological and non-phenomenological self. The latter type of measurement includes the TAT and Rorschach. Here we are mainly concerned with reviewing those instruments which measure phenomenological self, however.

Wylie divides these instruments into three general types, (1) Q-Sorts, (2) rating scales, questionnaires, adjective check lists, and (3) coding plans for interview materials. Although she does not honor the TST with inclusion in her list of measurements, it could be placed logically in the third class. This omission is undoubtedly due to Wylie's psychological rather than sociological orientation.

Her list (35, pp. 61-64) includes 23 different Q-sorts, of which she deals most extensively with the Q-sorts of (a) Butler and Haigh and (b) Hilden. Wylie, writing in 1961 found that the second major category of instrument accounted for the most use. However, it was this author's experience, in surveying dissertations (1954-present) which dealt with "self" in a sociological orientation to social psychology, that Q-sort technique was used in a majority of cases.

Wylie lists (35, pp. 87-97) 83 instruments of the type gathered together under the heading "rating scales, questionnaires, and adjective check lists." For only one-third of these is reliability information available and it is usually of the split-half coefficient type.

In addition, Wylie lists seven coding plans for interview materials. (35, p. 107-110) Although the TST is essentially a coding plan type, it differs from all those listed by Wylie in that it is used to code data from a pen and pencil or oral interview which could be administered in either an experimental or field setting. The seven code plans listed by Wylie are for coding material sampled from a population of clinical clients. Reliability between judges is reported in six of the seven code plans.

Raimy (28) reported, "three out of four judges coded 356 client responses with percentages of agreement ranging from 50.5 percent for 'ambivalent' to 81 percent for informational questions. Bugental (36) reported five raters rated five protocols with reliabilities for "unit determination, 87.4 percent; for categorization, 59.1 percent; and for evaluation, 75.0 percent."

Vargas (33) reported a 96 percent agreement between two judges on which statements constituted self-description, in general. For Stock (37) two judges categorized each statement from three interviews. The percentages of exact agreement for each interview ranged from 68.5 percent to 82.2 percent.

Scheerer's (29) scheme used a 5 point scale for each variable. "To explore reliability, three judges rated all units in six interviews. At least two out of the three judges agreed 93.8 percent of the time. On only four out of 178 Self units did any one judge's rating deviate by two scale points from both other judges ratings." Raskin (38) using a similar technique with a four-point scale reports an "interjudge reliability on 59 items" of .91.

For Lipkin (18), "interjudge agreements were obtained ranging between 81.9 percent and 100 percent" using a "complicated reliability formula."

It is interesting to note that Wylie includes Bugenthal and Zelen's W-A-Y Test or "Who are you" as a questionnaire (5), even though it has great similarity to the "Who am I" of the TST, and employs a 17 category code plan for the coding of three responses.

Historical Development of the Twenty Statements Test

The TST was developed by Kuhn and McPartland and first reported by them in 1954. The instructions to respondents which they used in their initial applications were:

> There are twenty numbered blanks on the page below. Please write twenty answers to the simple question, 'Who am I?' in the blanks. Just give twenty different answers to this question. Answer as if you were giving the answers to yourself, not to somebody else. Write the answers in the order that they occur to you. Don't worry about logic or 'importance.' Go along fairly fast, for time is limited." (16, p. 69)

These instructions were given in writing to some groups, and orally to other groups in a college class setting. The time was limited to twelve minutes. The number of responses obtained in this way varied from 20 to one or two, with a median of 17. Each response either began with "I am ..." or omitted that phrase and consisted of single words.

The responses were content analyzed into two gross categories, either consensual or subconsensual. A consensual statement refers to "groups and classes whose limits and conditions of membership are matters of common knowledge. Those which refer to "groups classes, attributes, traits or any other matters which would require interpretation by the respondent to be precise or to place him relative to other people" were considered subconsensual.

> Examples of the consensual variety are 'student,' 'girl,' 'husband,' 'Baptist,' 'from Chicago,' 'pre-med,' 'daughter,' 'oldest child,' 'studying engineering,' that is, statements referring to consensually defined statuses and classes. Examples of the subconsensual category are 'happy,' 'bored,' 'pretty good student,' 'too heavy,' 'good wife,' 'interesting'; that is, statements without positional reference, or with reference to consensual classes obscured by ambiguous modifiers. (16, pp. 69-70)

An intercoder reliability check between two coders was reported as resulting in less than three differences per 100 responses. The major score developed from the responses was a locus score, or the number of consensual statements made before any but erroneous (Guttmann scale "error") subconsensual statements are made in a TST SIP protocol. This locus score was developed after submission of the data to Guttmann scalogram analysis techniques indicated that subjects tended largely to exhaust all their consensual responses before beginning to make any subconsensual responses. Kuhn and McPartland equated the size of this

locus or consensuality with the respondent's amount of anchorage or self-identification.

It is not clear what utilitarian purpose was served by submitting the data to Guttmann analysis. A locus score could be obtained simply by counting the number of consensual items made as responses. Only as evidence that consensual responses in general are more salient is such a method useful.

The salience score was based upon the assumption that earlier statements were more salient for respondents than later statements. This assumption was made in spite of the specific instruction to the respondent to the contrary. Therefore, this writer assumes that the TST's developers had some type of subconscious salience in mind when they made this assumption. The salience score of any statement is obtained from its rank position in the protocol, beginning with #20 for the first statement.

This assumption that sequence is equivalent to salience is based on a statement by Newcomb, who says that salience "refers to a person's readiness to respond in a certain way. The more salient a person's attitude the more readily will it be expressed with a minimum of outer stimulation." (16, p. 72) They also advance an alternative explanation that consensual statements are made first because individuals are habituated to such responses in our culture by remembered external stimuli, such as applications, questionnaires and census takers. They refute this alternate explanation on the evidence that three and four-year-olds answer the question, "Who are you?" with the consensual statements of name, sex and age, though they have not yet any memory

of such habituating stimuli.

The research reported in connection with this introduction of the TST SIP was a correlational study between religious denomination of the respondent and amount of anchorage of the respondent. Wylie says of this research, "Their results are not meaningful psychologically for several reasons. First, as we have said above, there is no clear relationship between religious affiliation and psychologically relevant variables. In addition, there was no control over response total in obtaining scores, and there was no attempt to match groups of varying religious affiliations with respect to variables relevant to the "Who am I" responses." (35, p. 142)

Following this original report of the TST, several dissertations were done under Kuhn's direction at the University of Iowa employing the TST to gather data. Notable among these is that of McPartland (19) which was completed a year prior to the Kuhn & McPartland article, and the master's thesis of Robert Stewart (31).

Other of Kuhn's students who have employed the TST in their post-doctoral research are Fred Waisanen and Carl Couch, along with McPartland. Under the influence of the presence of Waisanen, Couch and Stewart at Michigan State University the TST was employed by Wilbur Brookover, in a study of self-concept of Junior High (4) students. The TST was also used by Charles Tucker (32), Clark McPhail (23), and Delwyn Dyer (10) in their dissertations, and by this author in the present research.

Writing in 1956 Kuhn alluded to the TST as non-specific, non-suggestive and non-structured. He adds that locus scores increase with age during the first 20 years, and that there are significant variations in locus by sex and occupation. He tells of a different category system used on the responses of "more than 200 respondents";

> (1) references to statuses in social categories and social groups; (2) references to ideological beliefs -- moral, philosophical, ethical; (3) statements of personal aspiration and achievement; (4) identification in terms of interests and aversions, including positively and negatively held social objects; and (5) self evaluations. (12, p. 245)

Unfortunately, he reports no inter-coder reliability for this coding plan. In another source he elaborates on this coding plan:

> These five categories (sufficient to order all the responses made) are the following: (1) social groups and classifications (such as age, sex, educational level, occupation, marital status, kin relations, socially defined physical characteristics, race, national origin, religious membership, political affiliation, formal and informal group memberships); (2) Ideological beliefs (including statements of a religious, philosophical or moral nature); (3) interests (including statements relating objects to the self with either positive or negative affect); (4) ambitions (and all anticipated success themata); (5) self-evaluations (such as evaluations of mental and physical and other abilities, physique and appearance, relatedness to others, aspirations, persistence, industriousness, emotional balance, material resources, past achievements, habits of neatness, orderliness, and the like, and more comprehensive self-typing in clinical or quasi-clinical terms).(15, pp. 40-41)

In viewing these two explanations of the same code plan by the same author, there is one glaring contradiction which might lead to coder error: In the first explanation "aspirations" is included in the type three, achievement, ambition, aspiration category. In the second explanation "aspirations" are included in the examples of type five, self-identification category. Furthermore, the type three and Type

four categories interchange content in the two explanations which could lead to further confusion.

There are other sources of possible confusion which can be stated in more operational terms. These all pertain to a difference between one of the last four categories, and the first category. That is, they are differences between the consensual category and the four new categories into which Kuhn has split the subconsensual category. (1) While physical description is category one, physical evaluation is category five; (2) while membership is category one, aspiration to membership is category three or four, interest or affect toward a group is category three or four, and beliefs associated with group membership (such as religious or political belief) is category two.

McPartland, meanwhile, over the years 1958 to 1961, set out another method by which the TST might be coded, while suggesting still other methods. In his Manual for the Twenty Statements Problem, (22, p. 2) he suggests category systems based on "literal content, referential frame or logical form". Kuhn's category system discussed above, as well as the code plan developed by McPartland are of the referential frame type. Drawing upon three separate explanations of the McPartland category system Table I was constructed to aid the reader in grasping all the distinctions drawn by McPartland.

McPartland reports a time limit of 15 minutes on his respondents, all of whom were patients in one or another facility for treatment of mental illness. His instructions were printed for reading and read aloud for listening for each experimental group, as follows: "There are 20 numbered spaces on this sheet. Just write 20 different things

Table I

Distinctions Between Categories for McPartland's Code Plan

Labeled	Concrete and* Constricted	Positional*	Situation-Free Style*	Comprehensive*
Content	Physical attributes and "identification card type" of information.*	Implies involvement in explicitly structured social situations*	Styles of behavior which the respondent attributes to himself#	No particular context, act, or attitude%
Relation to Society	None - too concrete to require social interaction#	Defined by society through interaction#	Loose involvement in social structure*	No interactive consequence or commitment%
Validation by - - -	by reference to records or measurement#	consensually validated by society#	not verifiable, but understood by others, having a common meaning, communicable%	Not consensually validated, not verifiable by communication%
Implications for prediction of behavior	None #	Implies norms and specific prescription of behavior, permits prediction#	Varied in varied situations but with consistent style. Permits prediction of "how" but not "when or where"	No reliable expectations about behavior#
Implication of "others"	None% --	A generalized other in an institutional pattern%	A generalized other, not in an institutional pattern%	Transcendental other%
Implications for "self"	A social object#	a social object#	a social subject#	a social subject#
Basis from Original Categories	Consensual%	Consensual%	Subconsensual%	Subconsensual%

*McPartland, Thomas S. and John H. Cummings, "Self-Conception, Social Class, and Mental Health," Human Organization, Vol. 17, #3, 1958, pp. 24-29.

#The Greater Kansas City Mental Health Foundation, Department of Research, Manual for the Twenty-Statements Problem, Revised, January, 1959, (author known to be McPartland).

%McPartland, Thomas S., John H. Cumming, and Wynona S. Garretson, "Self-Conception and Ward Behavior in Two Psychiatric Hospitals, Sociometry, June, 1961, pp. 111-124.

about yourself in the spaces. Don't worry about how important they are or the order you put them in. Just write the first 20 answers you think of to the question: 'WHO AM I?'"

McPartland reports that for a sample of 60 respondents, three independent coders agreed on more than 97 percent of responses. He also developed another type of score for the TST, the modal response. Each respondent's protocol was characterized as type A, B, C, or D, dependent upon which category of response was modal for the responses of the protocol. Of the mode he says, "about nine patients in ten write responses which show a clear mode in some one category, that is, they made at least one more statement in some one category than in any other. The remaining 10 percent either tied in two categories or gave so few responses that the mode was obviously unreliable. These few respondents, nevertheless, were categorized by mode when there was one, or in the more "distinctive category involved in a tie; in 'A' or 'D' if one of these was tied with'B' or 'C' and in 'D' if that category was tied with 'A!' The rationale is that more unusual responses should receive greater weight in analysis." Further, when a respondent made no responses he was classified as Type A on the assumption that he was responding in a "constricted manner." (22, p. 9) The three coders reached 100 percent agreement on the modal type of 60 respondents. (20, p. 118)

Couch seems to be the first investigator to use a category system stressing the "literal content" referred to by McPartland. Using the same instructions as the original Kuhn & McPartland study, with an eight minute time limit, he content analyzed for "references to a role in, or an attachment to, a major institution of our society...responses that had

a reference to religion, the family or education." (8, pp. 492-493) In a later study his code plan "noted whether or not respondents had identified themselves on the TST as male or female, boy or girl, or man or woman." (7, p. 118)

The code plan under consideration in this study (see Appendix B) is perhaps the most elaborate in terms of "literal content" analysis, separating Kuhn's original "consensual" category into 23 separate content areas (see Table II). Each of these categories was required as a relevant variable in at least one of the research interests bound together by cooperative data collection. Because "wastebasket" categories are included, these categories comprise an exhaustive code plan for the consensual responses.

In a suggestion for a code plan based on "logical structure," McPartland lists categories of responses (1) in the form of the verb "to be", (2) in the form of the verb "to have", (3) in the form of action verbs, and (4) in a form in which the "self drops entirely out of explicit attention." (22, pp. 9-10) Perhaps as a follow-up to this suggestion, Couch has studied the classes of verb forms used in response to a modified form of the TST. (9)

Probably the most useful innovation in the categories for TST codes was one dealing with logical or formal structure, made by Waisanen. His three major categories are (1) consensual, (2) evaluative-consensual, and (3) idiosyncratic or non-consensual. Essentially he has divided Kuhn's "sub-consensual" category into evaluative-consensual (positional references which are, in some way, modified) and idiosyncratic (all remaining sub-consensual references). Then he provides for 5 sub-

Table II

Growth and Development of Code Plans for the TST SIP

Author	Kuhn and McPartland	Kuhn	Couch	McPartland	Waisanen		Code Plan of Present Study	
Categories			Consensual Content	Concrete & Constricted	Consensual	Evaluative Consensual	Consensual	Evaluative Consensual
			1. Sex		1. Personal		1. Personal	
		Social Groups & Classifications	2. Religion				a. age b. name c. address d. physical characteristics e. sex f. health	
	Consensual		3. Family	Positional	2. Primary		2. Primary	
			4. Education				a. kinship b. marital c. friends d. residual	
					3. Secondary		3. Secondary	
		Beliefs		Situation-Free Style			a. community b. occupation c. work d. formal organization e. education f. religiosity	
		Aspiration, Achievement			4. Categorical		4. Categorical	
	Sub-Consensual	Interests & Aversions	Sub-Consensual	Comprehensive & Non Predictive			a. nation, race b. religious organization c. political organization d. socio-economic e. domestic f. residual	
					5. Residual		5. Residual	
		Self-Identification			Idiosyncratic or Non-Consensual		Idiosyncratic	

categories to analyze the "literal content" of both consensual and evaluative-consensual categories. The utility of this innovation lies in its ability to extract content data from a great number of TST responses which previously were classified sub-consensual and, therefore, paid little heed. His five content categories are; (1) Personal (physical characteristics), (2) Primary (continuing intimate relationships), (3) Secondary (associational, segmentalized, specialized interaction), (4) Categorical ("man," "citizen," "white," "American," are examples), and (5) Residual (all responses that do not fit the first 4 content categories). This is essentially the basic category system studied here, with the addition of the 23 sub-sub-categories already referred to in the form of specific roles or physical attributes.*

The present study also involves an innovation in the administration of the TST. The instructions were read to respondents to

> Ask this question of yourself, 'Who am I?' Think of as many answers as you can in answer to the question, 'Who am I?' In a moment I would like you to give me the answers as if you were giving them to yourself, not to me or anyone else. Take a little time to think about it. (pause) Now, please make what you consider to be the most important statement about yourself first. (pause) Now, make what you consider to be the next most important statement about yourself. (pause) Now, make what you consider the next most important statement about yourself. (pause) Are there any other statements you could make about yourself in answer to the question, 'Who am I?' (see Appendix A)

The interview schedule contained only 10 spaces for responses and interviewers were instructed to write down only as many as ten responses and to spend no more than 3 minutes on the SIP. (13, p. 7-8) We see,

*Attributed to Waisanen by Dyer, (10, pp. 39-40).

by comparison with prior uses of the TST that, (1) fewer responses were sought, (2) less time was allowed, and (3) respondents were directed into a format of most important statements first. These innovations were the result of a pre-test of the interview schedule.

The main aim of this pre-test was to test the schedule on persons who were functionally illiterate, to find what impairments to communication this would present to an interviewer. The pre-test was conducted with persons living in a rural, Michigan, migrant farm-workers enclave, and with persons living in an urban, Michigan, Negro and Puerto Rican slum ghetto. Such a pre-test was as essential for the TST as for any other phase of the schedule, for until now the TST had been employed only on extremely homogeneous groups, and often on middle-class college-age persons. Now we were pre-testing for its utility with an extremely heterogeneous sample, by testing the opposite extreme type of homogeneity.

We found respondents on the pre-test made many fewer statements, many non-directive probes were used, and a great deal of time was spent on administration of the TST in relation to other items, although there was only one hour available for the administration of the entire schedule. For the sake of expedience, then, the TST was administered with built-in probes which provided more structure than previously, and in a foreshortened format. Brookover had previously used a "Ten Statements Test", (4, p. 28) and Dyer had questioned the need for obtaining as many as twenty statements. (10, p. 41)

A further innovation in administration of the TST was developed and tested by McPhail. After obtaining TST protocols from respondents he presented them with three additional problems:

> (1) rank these statements in the order of their importance to you...
>
> (2) assign a plus (+), minus (-) or zero (0) to each statement, depending on whether you feel positive, negative, or neutral toward that statement... and
>
> (3) attempt to estimate (in units of five percentage points) what percentage of those persons who are in a position to know, would agree with you on each separate statement you have made. (24)

This was an attempt to (1) test the correlation between chronological order of response and salience for the respondent, (2) test for any correlation between negative form of response with negative feeling about the response. For example, had a response "I am not good-looking" been given, a coder might assume this to be a negative self-evaluation, but unless we know whether his looks makes the respondent sad or glad, we cannot tell if it is a negative or positive evaluation. Finally, (3) this was an attempt to tap the respondent's perception of consensuality or sub-consensuality for each statement he had made. Conceivably a respondent might estimate that 60% of the persons in a position to know would agree with his statement "I am stubborn", thus changing it from an idiosyncratic response to a consensual response. Again, a forerunner to the McPhail approach to salience was used by Brookover, when he asked respondents to circle the most important (no specified number) of their statements. (4, p. 28) Unfortunately, no report was made of the association between this self-coding and the ordering of the statements.

Looking toward the future for the TST SIP, Dyer has suggested that

responses to it be submitted to factor analysis, rather than content analysis or scaling techniques. (10, p. 41) Mielke has suggested a more specific plan for factor analysis of TST responses.

> A factor analytic method could also be used to devise a value adherence instrument, using a two stage pretest. The first stage would be exploratory, going to the field with open-end questions...Everytime a respondent would mention...something that sounded like a general value, this would be recorded for the second pretest stage. The entire pool of general values thus recorded would then be administered to a second pretest sample for "agreement" or "applicability to me" ratings. These ratings would then be submitted to factor analysis. Here the factors would represent a parsimonious set of independent value clusters. By discerning an underlying commonality of content, each value "cluster" could be called a single general value. The items most highly loaded on each factor would become the basis for the value adherence instrument in the main study. (27)

In summary, Table II represents a history of the growth and development of code plans for the TST SIP.

Review of Previous Investigations Relevant to this Study:

Berelson, (3) in his article on "Content Analysis" in the Handbook of Social Psychology presents a summary of content analysis reliability studies. He finds that reliability is reported in 15-20 percent of the content analytic studies, and that "the reports on reliability which do appear are uniformly high". In a population of "some 30 studies and experiments" he found "a range of correlation coefficients between about .78 and .99 with a concentration at about .90 and a range of percentage agreements between 66 percent and 96 percent with a concentration over 90 percent." He cautions that "many attempts at content analysis may have been abandoned because of low reliability in the preliminary results... published content analysis studies which do not report on reliability

may be presumed to have had less satisfactory results...most of the reported reliability results apply to relatively simple versions of content analysis." (3, p. 514)

He summarizes several reliability experiments as follows: "Reliability is higher under these conditions: the simpler the categories and the unit, the more experienced and better trained the coders, the more precise and complete the set of coding rules, the fuller the illustrations." (3, pp. 514-515)

Schutz (30) compared a code system using "one psychological operation" on the part of the coder, with a code system using a step-by-step series of decisions related to a complex category system and found the latter system produced a "little less" reliability.

Summary

In this chapter we have attempted to (1) present a general statement of the aims of this study; (2) show the need for a minimally biased SIP in tests of Meadian theory; (3) point to a number of existing SIP's; (4) outline the historical development of the Twenty Statements SIP regarding (a) code plans, (b) derived scores, (c) reliability tests, (d) administration procedures, and (e) future development; and (5) review the findings of other checks on the reliability of content analysis.

The remaining chapters deal with (1) the procedures employed in this study, including the hypotheses being tested, (2) the results obtained by testing these hypotheses, and (3) the conclusions arrived at, based on the results of the hypothesis testing procedure.

CHAPTER II

PROCEDURE

The Experimental Context of this Study

The present reliability test was conducted in conjunction with a cooperative research project which collected data in five nations: Mexico, Costa Rica, Japan, Finland and the U.S.A. This very extensive study was under the direction of Hideya Kumata, Charles Loomis and Frederick Waisanen in cooperation with Yrjo Littunen and Robert Stewart.

The United States sample was constructed and drawn by the Gallup Organization, which also supervised and conducted all U. S. interviews. It is "a (modified) area probability sample of the United States general public (outside of institutions), age 21 or older." Size of the sample was 1,528. The interviews were conducted September 2 to October 6, 1963. (39, p. 1) The data collection instrument was an approximately hour-long scheduled interview.

Using the 1,528 TST protocols thus obtained, a random sample of 150, or approximately 10 percent of the population of protocols was drawn (see Appendix C). This sample was the data assigned to each coder for coding. The data sampled contained 858 responses. This amounts to an average number of responses per protocol of 5.72.

It is assumed, by sampling randomly from a sample designed to represent the U. S. adult population, age 21 and over, that the range of coding difficulty found in these protocols will be closely representative

of the range of coding difficulty which would be found in protocols generated by the U.S. adult population, 21 and over.

The Coders and Data Collection Procedure

A coder employed by the Gallup Organization had coded the 150 protocols prior to the experiment. Future references shall refer to the Gallup coder. In addition, five non-Gallup coders operating independently, coded each of the 150 TST protocols. In Chapter IV I will deal with the possibility that these five coders were not truly independent. However to safeguard the independence of the five coding operations, each coder did the physical coding of protocols in a separate booklet of protocols. To further safeguard independence, the protocols in each booklet were ordered in a random fashion and differently from each of the four other booklets. We should note that this procedure may facilitate the contribution of varying levels of coder fatigue to differences in coding, thus biasing the results in the direction of lower reliability.

The booklets were divided into two sections of 75 protocols each. One and a half hours coding time was estimated for each section. Coders were instructed to "code each section on a separate day, and code all sections within a two week period. Please code each section during one continuous work period. Try to do this work at a time of day when you are mentally alert and relatively free from external distraction." Unfortunately not all coders were able to follow these instructions. While all coding was accomplished within a two week period, some coders found it necessary to use more than two separate days and shorter than one and a half hour work sessions. A change in the procedure of this sort,

however, would lead to expectations of lowered fatigue for the task.

On each page of the booklet in typewritten, dittoed form were the responses of a TST protocol. Only one protocol appeared on each page, and no protocol appeared divided between two pages, regardless of the number of responses it contained. The number of responses per protocol ranged from one to ten. A respondent number in four digits appeared in association with each protocol. We should note that the Gallup coder performed the coding task on data handwritten by interviewers under field conditions, and that this was the original data from which the experimental booklets were transcribed.

The code categories used were in the form of two digit numbers representing punches in two columns of an IBM data card. Each coder was instructed to: "Code in red pencil....Indicate by red slash marks within the statement the division of a respondent's statement into more than one statement for coding purposes. Write your codes in the left margin in the order the statements appear on each page."

At the beginning of each experimental booklet the following information was given the coders regarding the process of transcribing the protocols from the original handwriting:

> Following this page you will find a reproduction of the questionnaire section which asked and recorded the TST. (see Appendix A) Reference to this format may be helpful in coding some of the responses. In general, each statement which was written next to a number 1-10 has been reproduced as a separate response for you to code, by: (a) Triple spacing it from the other statements, (b) Capitalizing the first letter, if it were not capitalized by the interviewer (except in some cases where capitalization could affect meaning), and (c) Ending the statement with a period (.). Since in some cases, it appeared that though a response was written on several lines, it was really one statement, we have shown this by ending the line with a dash (--) and beginning the next line in lower case. Every effort was made to reproduce the statement as recorded by the interviewers. Therefore, grammar and spelling errors have been reproduced.

It is assumed that the Gallup coder had become highly aware of the 15 page coding instructions and code for the TST developed by Stewart and others. (see Appendix B) Four of the five non-Gallup coders had helped develop the coding instructions and code book, and were therefore very familiar with the code book. All five coders were instructed to "Read through the instructions in the code book for coding the TST before you begin this work. Follow the instructions in the code book during the coding and refer to the code book as frequently as necessary." All five coders were at the graduate or post-graduate level in the social sciences.

(See Appendix A for the format of the TST in the interview schedule, Appendix B for the TST coding instructions and code book, and Appendix C for the 150 TST protocols.)

The Hypotheses and Scoring Procedures

A basic hypothesis of this study is that there will be a high level of reliability among the six coders on the coding of the TST protocols. I will define "high level of reliability among six coders" as a high percentage of protocols with a high mean reliability score. The mean reliability socre of a protocol is an average across the reliability scores of all response in the protocol. The reliability score of a response is based upon a scale of agreement-consistency among the six coders as to how they coded the response. The scale of agreement-consistency is based upon two values (1) Agreement, or the number of paired agreements obtained, considering that a total of 15 paired agreements are possible among six coders, and (2) Consistency, or the size of the modal category assigned to the item by six coders, which mode may range from one to six. The scores thus

obtained from this scale range from 0-10. Each score represents a discrete structural pattern of category combinations, and, in each case, knowing nothing but the score we can determine the number of paired agreements and size of mode which it represents, but not the code category. Table III will make this relationship more apparent.

Table III

Reliability Score- Based on a Scale of Agreement-Consistency

Score	Combination of Categories (a structural pattern)	Number of Agreements	Size of Mode
10	AAAAAA	15	6
9	AAAAAB	10	5
8	AAAABB	7	4
7	AAAABC	6	4
6	AAABBB	6	3
5	AAABBC	4	3
4	AAABCD	3	3
3	AABBCC	3	2
2	AABBCD	2	2
1	AABCDE	1	2
0	ABCDEF	0	1

To understand the formal patterns one must realize that the letters A,B,C,D,E,F represent the order in which the analyzer notes different code category assignments by different coders. There is no relationship between these letter notations and the different numbers assigned to code categories in the code book. For example, if the analyzer notes that all the coders coded an item with the same code category, that item would be assigned an AAAAAA structure, whether it was coded a 21-category or a 53-category by all six coders. At the other end of the scale, if the analyzer notes that each coder assigned a different code category

to the same item, that item would be assigned an ABCDEF structure, whether it was coded as 72, 56, 12, 58, 60, 61, or any other combination of six different categories. Thus, the difference between a score of seven and eight on this scale is not found in the mode, for in each situation there are four A's. The difference between the reliability scores of seven and eight is found in the number of paired agreements, for an eight score represents one additional paired agreement between the two coders who assigned non-modal categories, i.e., "BB" represents a paired agreement which is missing in "BC".

For purposes of operationalizing this hypothesis a high reliability score was defined arbitrarily as any score of 8 or over, thus representing roughly the top quarter of the scale. "A high percentage of protocols with a high mean reliability score" was defined as 75 percent or more of all protocols have a mean reliability score of 8 or better. The percentage figure was chosen as an amalgam of (1) previously attained percentages of inter-coder reliability for the TST, and (2) indications from related research in content analysis that factors specific to this study indicate a lower percentage of inter-coder reliability can be expected.

The history of TST reliability checks has been discussed in Chapter I. We should recall that in their original study, Kuhn and McPartland reported 97 percent reliability between only two coders, coding material into only two categories. McPartland reports 97 percent reliability between only three coders, coding into only four categories. Berelson reports percentages of reliability for content analyses range between 66 percent and 96 percent with a concentration over 90 percent,

but he adds that "the reported reliability results apply to relatively simple versions of content analysis."

Berelson concludes that the simpler the categories the higher the reliability, and the reliability experiment reported by Schutz indicated that a simple decision by the coder would produce higher reliability than a step-by-step series of decisions by the coder. Reference to Table II and Appendix B would indicate that there are 45 discrete code categories used in the present study, requiring the coder to make decisions step-by-step as to (1) whether a response is a single or multiple statement, (2) whether a response is or is not modified in some way, (3) whether response contains material which may be classified by content (consensual) or does not contain material which may be classified by content (idiosyncratic) and (4) what content category (from among 23) should be assigned to this response.

Furthermore, although the difference between the use of three coders in McPartland's reliability check and the six coders in the present reliability check may seem inconsequential, a deeper consideration is that with three coders there can be only three paired agreements but with six coders there are 15 paired agreements. Adding only three coders to the reliability test multiplies the chances for disagreement five times, and we would expect such a procedure to lower the obtained reliability.

Thus three main factors differentiate this study from other reliability checks of TST coding; (1) many more categories, (2) many more decisions to be made by the coder, (3) many more paired agreements among coders. Add the final consideration that in the present study the data being coded, obtained, as it was, for the first time from a

heterogeneous sample rather than a homogeneous sample, might be expected to extend the range of expression thus making coding more difficult and therefore less reliable. For these reasons the expected percentage of reliability was lowered to 75 percent from the 97 percent obtained by McPartland and Kuhn & McPartland.

Now our original basic hypothesis, that there will be a high level of reliability among the six coders on the coding of the TST protocols has been operationalized. Its meaning has become, we expect to obtain a situation in which 75 percent or more of all protocols have a mean reliability score of eight or better, on the coding of 150 TST protocols by six coders. An item reliability of eight or better means the item will have at least seven of 15 pairs in agreement and a modal category no smaller than four of six codes. The mean reliability score of a protocol is an arithmetic average to the nearest tenth, of the scores obtained on all the items in the protocol.

A somewhat similar basic hypothesis is that there will be a high level of reliability among the six coders on the coding of the TST responses. In this case, "high level of reliability" will be operationalized as (1) 75 percent of all responses will be scored as eight or better in reliability and (2) the average reliability of all responses will be a score of eight or better.

The following sub-hypotheses will be operationalized in parallel fashion to the basic hypothesis:

H1 When the difference between consensual responses and evaluative-consensual responses is overlooked, reliability is expected to appear as greater than under realistic conditions.

H2 When differences between coders in the decision of whether to divide a statement into more than one statement for coding purposes are adjusted so that only the immediate statement is affected rather than all subsequent statements in the protocol, reliability is expected to appear as greater than under realistic conditions.

It can be seen from the discussion of step-by-step decisions by the coders that each of these hypotheses reflects an attempt to isolate one of the first two decision steps. The effect of the decision "whether a response is or is not modified in some way," is removed from consideration which allows us to hypothesize that this will make the coding appear more reliable. The accidental effect (not the main effect) of the decision "whether a response is a single or multiple statement" is removed from consideration, allowing us to hypothesize that this will make the coding appear more reliable. Note that neither of these hypotheses expect a change in reliability. Rather they expect a change in the appearance of reliability under a hypothesized condition which is not a condition of reality.

The actual data record on data cards or data computer tape, the realistic condition, is one which records discrete numbers for two statements which are alike except that one is modified and the other is not modified. Thus, "I am a man" would receive the code punch, 16, but "I am a fairly honest man," would receive the code punch 17, denoting what would have been a "16" category, except for the modification, "fairly honest". For our purposes "modified" is used in both the strict, grammatical sense, as well as in the less strict sense of overall meaning. Note also, that the terms "evaluated" and "qualified" will be used as if they were equivalent to the term "modified", since no

practical distinctions between them are necessary in this study.

The actual data record, the realistic condition, will also allow accidental effect to accumulate for all items following an item which was or was not split into two statements in an unreliable manner. This accumulation relates to the fact that each statement's category code number is punched in a set of two discrete data card columns.

Thus if the first statement of seven statements is split into two statements in an unreliable manner, each of the following six statements will be affected accidentally in the direction of unreliability by necessarily placing what may be reliable code category decisions in an unreliable set of columns. Under the condition of reliability the second statement's code category number would have been in the second set of two columns, and instead it is in the third set of two columns; the seventh statement's code category number would have been in the seventh set of two columns, and instead it is in the eighth set of two columns. Thus the unreliability of a split statement decision is compounded accidentally, and especially so where it occurs near the beginning of a protocol, and the number of statements in the protocol is relatively large.

Three factors are apparent in the coding situation, which might be expected to affect reliability: (1) The data being coded, (2) The code categories being used, and (3) The coders who do the coding. Each of the following sub-hypotheses relates to one of these factors:

H3 The more statements made on a TST protocol the lower will be the mean reliability score for that protocol.

H4 As content categories increase in complexity response reliability will lessen.

H5 There will be no relationship in the pattern of coder agreements which contributes to unreliability.

H5a There will be no relationship in the pattern of paired agreements which contributes to unreliability.

H5b There will be no relationship in the pattern of triple agreements which contributes to unreliability.

Hypothesis three has three bases. (1) As mentioned above, a greater number of protocol statements is expected to lower reliability in the specific cases where an early protocol statement is split unreliably. (2) A greater number of statements in a protocol increases the appearance of the protocols' complexity. (3) A greater number of statements in a protocol increases the number of step-by-step decisions within the protocol.

Hypothesis four refers to the 23 specific content categories. "Complexity of reference " refers to a four place scale of complexity, in which personal statements are defined as less complex than primary statements, primary statements are defined as less complex than secondary statements, secondary statements are defined as less complex than categorical statements, and idiosyncratic statements (because this category has a residual dimension) are defined as equal in complexity with the personal statement, at the low end of the complexity scale.

Only those statements to which four, five or six coders assigned categories of the same complexity level can be analyzed within this hypothesis. This does not mean that at least four coders will have assigned exactly the same numerical category code to these statements. For example four of six code categories assigned to a statement are 40, 41, 42, 43. Such a statement is at the secondary level of complexity, since all four of these categories are secondary consensual statements.

If fewer than four coders coded the item at the same level of complexity, it will not be clear what level of complexity we are dealing with. Therefore, such items will not be analyzed under this hypothesis.

Hypothesis five is the expectation when all coders are independent. This experiment assumes coder independence. Therefore, this is a relevant hypothesis. To operationalize this hypothesis a Chi Square procedure will be utilized. The obtained value in each cell is the number of unreliable agreements between two coders, while the expected value in each cell is the number of agreements expected between two coders if unreliable agreement is equalized across the five other coders. We will assume the hypothesis is supported if the Chi-Square value is not significant. The same analysis will operationalize Hypothesis five-a and Hypothesis five-b.

Physical Analysis Procedure

The first step in analysis was to reorder the pages in each of the five experimental booklets from their random and different orders to the same order.

Next the raw data of six codes for each item was associated with the protocol number and item number in six two-digit columns of typewriter notation. Typewriter notation proved most efficient because the consistency of form in typewritten characters led to easier identification of differences in categories. The visual discrimination was more time-consuming with hand-written character notation. The column of two digit codes for the Gallup coder was transcribed from a print-out sheet of the Control Data 3600 Computer at Michigan State University. The column of two-digit codes for each experimental coder was transcribed from red

pencil notations in the left hand margin of each of the five experimental booklets. Examples using the typewriter notation method may be found in Appendix C, Section two, three and four, as well as on pages 41 and 43 of this chapter.

In association with this raw data for the reliability experiment, notations were made of (1) the score of each item, (2) the average score of the protocol, (3) whether qualification-modification contributed to less than perfect agreement, (4) whether splitting a statement contributed to less than perfect agreement, (5) what levels of complexity of reference were involved in less than perfect agreement, (6) the number of items in the protocol, and (7) within the raw data columns, all code categories which were non-modal were encircled, in the manner that scale errors are encircled in typewriter notation of the sort developed by Waisanen. The end product of this procedure was 22 single spaced pages of raw data, associated with all measures of the variables necessary to operationalize the hypotheses.

Finally, additional typewriter notations were used to rescore protocols in which either (or both) qualification and modification operated as a factor contributing to unreliability. You may wish to refer again to the examples cited above.

Data Analysis Procedure

Protocols were rank ordered by mean reliability score. From this Table IV was constructed, representing the number of protocols receiving each score (in tenths) the percent of protocols by scores, and the cumulative percent of protocols above or at each score. This operation

was necessary for testing and illustrating the basic hypothesis and hypothesis one and two. To further test the general reliability of the coding procedures under the second basic hypothesis, a determination was made of (1) the percent of items scoring eight or better, and (2) the average reliability score figured across all statements. From this data Table V was constructed.

To test hypothesis one the statements were rescored, deleting qualification-modification as a source of unreliability and similar manipulations of these new scores were made. In this process the following discrete code categories were lumped together and considered to be equivalent categories for obtaining the reliability score: (10,11) (14,15) (16,17) (20,21) (22,23) (24,25) (26,27) (30,31) (40,41,42,43) (44,45) (46,47) (48,52,53) (50,51) (54,55) (56,57) (58,59) (60,61) (62,63) (70,71) Please refer to Appendix B for the denotation of these codes.

To make more explicit the procedure followed in rescoring a protocol for a test of hypothesis one, let us refer to an example from the data; a protocol which reads as follows:

I'm a human being.
I'm able to work.
I'm able to carry my home.
I am proud to have children.
I am a citizen.

The reliability score for this protocol, under the realistic condition, was 8.6, with five scores of ten, nine, five, nine and ten. Rescored for testing hypothesis one, it obtained a reliability score of 9.8, with five scores of ten, ten, nine, ten, ten. To realize how this apparent change in reliability was obtained, it is necessary to compare the actual code categories assigned to each response by six coders with the same data

after adjustments have been made to overlook, or remove the effect of, qualification.

ASSIGNED CATEGORY UNDER REAL CONDITIONS:

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1	60	60	60	60	60	60	10	I'm a human being.
2	43	43	43	42	43	43	9	I'm able to work.
3	62	63	63	63	72	62	5	I'm able to carry my home.
4	20	21	21	21	21	21	9	I am proud to have children.
5	50	50	50	50	50	50	10	I am a citizen.
							43	

8.6 mean score of protocol

ADJUSTED CATEGORY UNDER HYPOTHESIZED CONDITION:

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1	60	60	60	60	60	60	10	I'm a human being.
2	43	43	43	43	43	43	10	I'm able to work.
3	63	63	63	63	72	63	9	I'm able to carry my home.
4	21	21	21	21	21	21	10	I am proud to have children.
5	50	50	50	50	50	50	10	I am a citizen.
							49	

9.8 mean score of protocol

Note that only the four encircled "errors" due to qualification were changed under the hypothesized condition; and that the one encircled "error" due to some other cause has remained unchanged. From inspection of the three responses involved in rescoring, we can see that the "errors" which were changed were evidently due to a differential treatment of the three qualifying words, "able", "able", and "proud".

To test hypothesis two those protocols affected by splitting were reordered to delete the accidental affects of disagreements in splitting, then the statements were rescored. In this process blank code categories became meaningful. In the case where a coder had not split a statement while others had split the same statement, that coder was shown as

assigning a blank code category to the second row of codes devoted to the second part of the split statement.

To clarify this rescoring procedure, we refer to a protocol from the data, an example which reads as follows:

I'm an American Citizen.
I believe in fairness.
I try to deal fair with everybody.
I can get credit anywhere I've been very fair.
I'm just a common-ordinary fellow.
I never had more than a country school education.

The reliability score for this protocol, under the realistic condition, was 6.5, with eight scores of ten, eight, ten, nine, two, three, two and eight. Rescored for testing hypothesis two, it obtained a reliability score of 8.8 with eight scores of ten, eight, ten, nine, seven, nine, seven, and ten. Note that while there are only six verbal responses in the protocol, the coders have split some responses, thus creating eight response scores. This fact makes it difficult to determine which scores were assigned to which verbal response, especially under realistic conditions, before accidental effects of splitting are removed or overlooked. Therefore, in the example below, we must view all six responses as a whole. We may rely only on the fact that the first category assigned codes the first verbal response, the last category codes the last verbal response, and all other categories are in the same order as the verbal responses which they refer to. The column headed "Response Number" goes unused in this example to avoid confusion.

ASSIGNED CATEGORY PLACEMENT UNDER REAL CONDITIONS:

Response Number:	Coder G	A	B	C	D	E	Score	Verbal Response:
	50	50	50	50	50	50	10	I'm an American Citizen.
	50	50	(72)	50	50	(72)	8	I believe in fairness.
	72	72	72	72	72	72	10	I try to deal fair with everybody.
	72	72	72	72	(61)	72	9	I can get credit anywhere I've
	72	(59)	61	72	(57)	61	(2)	been very fair.
	61	72	47	72	61	47	(3)	I'm just a common-ordinary fellow.
	47	(17)		47	(61)		(2)	I never had more than a country
		(47)			(47)		8	school education
							52	

6.5 = mean score of protocol

After accidental effect of splitting is removed, it is necessary to assign two lines to each verbal response which was split by any coder. In general the categories and scores which are associated with one verbal response will appear on the same line as the response number and all following lines, until the next verbal response is begun. Note that some verbal responses could be split into more than two codable responses. Below is the example of rescoring which goes with the above example of original scoring. Response number refers to each verbal response.

ADJUSTED CATEGORY PLACEMENT UNDER HYPOTHESIZED CONDITION:

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1	50	50	50	50	50	50	10	I'm an American citizen.
	50	50	( )	50	50	( )	8	
2	72	72	72	72	72	72	10	I believe in fairness.
3	72	72	72	72	(61)	72	9	I try to deal fair with everybody.
4		(59)			(57)		7	I can get credit anywhere I've
	72	72	72	72	(61)	72	9	been fair.
5	61	(17)	61	(72)	61	61	7	I'm just a common-ordinary fellow.
6	47	47	47	47	47	47	10	I never had more than a country
							70	school education.

8.8 = mean score of protocol

Now what was really happening among the coders becomes a bit clearer. There was a four to two disagreement on the splitting of response one and four. Notice that even though removing the accidental effect of splitting changed the category content assigned to the second codable response (second line) it did not change the obtained score of eight. Under the realistic condition we have tried to circle all "errors", while under the hypothesized condition we have circled all "errors" which remain and may be attributed to some other effect.

Specifically, in verbal response number 1. evidently some coders split this into the two responses, "I am an American, and I am a citizen." In verbal response number 4. evidently some coders split this into the two responses, "I can get credit anywhere, and I've been fair." Finally, in verbal response number five it appears that only coder A noticed the word "fellow" and coded this response as a qualified sex reference.

To test hypothesis four it was first necessary to select those items which could be analyzed under this hypothesis (see discussion of hypothesis) and then each of these items was assigned a score (1-4) of "complexity". In the next section the ensuing statistical analysis is explained.

To test hypothesis five and its sub hypotheses it was necessary to note all the encircled errors which were in agreement and all the sets of tripletons in agreement in the even split situation (Reliability score = 6, structure = AAABBB). Counting all occurrences of 15 sets of agreement pairs resulted in 15 obtained values for the Chi Square cells.

Statistical Analysis Procedure

To test hypothesis three it is necessary to correlate the mean reliability score of 150 protocols with the number of items in each protocol. Data analysis has produced a total of three sets of scores (under three hypothetical conditions). It is necessary to do three correlations to test this hypothesis, since there was no designation of which set of scores would be used in this test. A significant negative correlation will support the hypothesis.

To test hypothesis four it is necessary to correlate the complexity score (1-4) with the reliability score of all statements testable under the hypothesis. Again three correlations must be done to reflect the three score sets developed in data analysis. Again, a significant negative correlation is needed to support the hypothesis.

To test hypothesis five and its two sub hypotheses it is necessary to do three Chi Square analyses for each of the three score sets developed. See page 35 for a description of obtained and expected values in these analyses. The hypothesis is supported if the Chi Square value is not significant.

TABLE IV
PART A

Mean Reliability Score of 150 TST Protocols
By N and % for Three Conditions
(Mean Scores of Eight or Over)

Mean Scores*		10.0	9.8	9.6	9.4	9.2	9.0	8.8	8.6	8.4	8.2	8.0
No Coding Effects Removed	N	14	8	3	11	12	13	7	9	9	11	11
	%	9.3	5.3	2.0	7.3	8.0	8.7	4.7	6.0	6.0	7.3	7.3
Cumulative	%	9.3	14.6	16.6	23.9	31.9	40.6	45.3	51.3	57.3	64.6	71.9
Qualification Effect Removed	N	23	13	7	10	13	17	8	9	6	7	4
	%	15.3	8.7	4.7	6.7	8.7	11.3	5.3	6.0	4.0	4.7	2.7
Cumulative	%	15.3	24.0	28.7	35.4	44.1	55.4	60.7	66.7	70.7	75.4	78.1
Accidental Effect of Split Removed	N	14	8	4	14	14	12	9	14	11	13	7
	%	9.3	5.3	2.7	9.3	9.3	8.0	6.0	9.3	7.3	8.7	4.7
Cumulative	%	9.3	14.6	17.3	26.6	35.9	43.9	49.9	59.2	66.5	75.2	79.9

*Since the process of rounding back to the nearest tenth may have increased the N of cases accounted for by scores with even tenths, we have tabled these results by grouping together the results accounted for by each even tenth with the results accounted for by each odd tenth immediately higher in the reliability scale. Thus, in the column headed 9.8 will be found all the cases scores as 9.8 or 9.9, while in the column headed 7.2 will be found all cases scored as 7.2 or 7.3.

Table IV
Part B

Mean Reliability Score for 150 TST Protocols by N and Percent for Three Conditions (Mean Scores of Less Than Eight)

	7.8	7.6	7.4	7.2	7.0	6.8	6.6	6.4	6.2	6.0	5.8	5.6	5.4	5.2	5.0	4.8	4.6	4.4	Total
N	4	4	3	4	4	4	3	4	2	1	1	2	1	1	1	2	-	1	150
%	2.7	2.7	2.0	2.7	2.7	2.7	2.0	2.7	1.3	.7	.7	1.3	.7	.7	.7	1.3	-	.7	100.2
c%	74.6	77.3	79.3	82.0	84.7	87.4	89.4	92.1	93.4	94.1	94.8	96.1	96.8	97.5	98.2	99.5	99.5	100.2	
N	3	4	1	5	6	4	4	1	3	-	-	1	-	1					150
%	2.0	2.7	.7	3.3	4.0	2.7	2.7	.7	2.0	-	-	.7	-	.7					100.3
C%	80.1	82.8	83.5	86.8	90.8	93.5	96.2	96.9	98.9	98.9	98.9	99.6	99.6	100.3					
N	3	5	9	2	4	1	3	2	-	-	-	-	-	1					150
%	2.0	3.3	6.0	1.3	2.7	.7	2.0	1.3	-	-	-	-	-	.7					99.9
C%	81.9	85.2	91.2	92.5	95.2	95.9	97.9	99.2	99.2	99.2	99.2	99.2	99.2	99.9					

Table V

Reliability Scores for all TST Responses
By N and % for Three Conditions

Reliability Score	10	9	8	7	6	5	4	3	2	1	0	Total	$\bar{X}$
No Coding Effect Removed													8.19
N	330	210	75	76	39	63	20	5	26	14	-	858	
%	38.5	24.5	8.7	8.9	4.5	7.3	2.3	.6	3.0	1.6	-	99.9	
Cumulative %	38.5	63.0	71.7	80.6	85.1	92.4	94.7	95.3	98.3	99.9	-		
Qualification Effect Removed													8.60
N	402	200	63	66	34	53	16	6	13	5	-	858	
%	46.9	23.3	7.3	7.7	4.0	6.2	1.9	.7	1.5	.6	-	100.1	
Cumulative %	46.9	70.2	77.5	85.2	89.2	95.4	97.3	98.0	99.5	100.1			
Accidental Effect of Split Removed													8.59
N	384	219	84	53	37	54	12	2	12	7	1	*865	
%	44.4	25.3	9.7	6.1	4.3	6.2	1.3	.2	1.4	.8	.1	99.8	
Cumulative %	44.4	69.7	79.4	85.5	89.8	96.0	97.3	97.5	98.9	99.7	99.8		

*When more than one response is split in a protocol, it may necessitate an increase in number of responses to remove the accidental effects of these splits. In this way we account for an increase of seven responses in the total N, under this condition.

CHAPTER III

RESULTS

General

Using the procedures outlined in Chapter II, it was found that 72 percent of all protocols were assigned mean reliability scores of eight or better under the realistic condition. In line with this finding, 72 percent of 858 responses obtained reliability scores of eight, nine or ten, and the mean reliability score across all responses was 8.19, while the modal reliability score was a perfect score of 10. 38.5 percent of all items were coded with perfect reliability across 15 paired comparisons. As was noted earlier, comparison of six coders entails comparisons of 15 pairs.

Table IV gives the N and percent of all protocols assigned discrete reliability scores (to the nearest tenth) under conditions in which no coding effects have been removed from consideration. Table V gives the N and percent of all responses assigned discrete reliability scores (0-10) under conditions in which no coding effects have been removed from consideration.

These results relate to the basic hypothesis that there will be a high level of reliability among the six coders on the coding of the TST protocols and on the coding of the TST responses. Although the results did not reach the expected level of 75 percent either for protocols or responses, the mean reliability score per response did exceed the minimum expected value of eight. In addition, the large percentage of responses

which obtained a modal reliability score on the extremely high side of the reliability range indicates the excessively skewed distribution in the direction of high reliability. This skewed effect is made more apparent in Table IV and Table V.

When the Effect of Qualifiers is Removed

Hypothesis one stated that when the difference between consensual responses and evaluative-consensual responses is overlooked, reliability is expected to appear as greater than under realistic conditions. The basic hypothesis was tested under conditions that removed none of the effects which would operate in the reality of the coding procedure. Hypothesis one must be tested under the condition which removes the effect of qualification from the coding of protocols and responses. For examples of this procedure see page 41 and Appendix C, Section II. Under this condition, 77 percent of all protocols obtained a mean reliability score of eight or better, while 78 percent of all responses were assigned scores of eight, nine or ten. The mean reliability score across responses was 8.60, and while the modal reliability score persisted at the value of 10, there was an increase in responses accounted for by this mode of 8.4 percent (46.9 percent) These results support hypothesis one since they represent (1) an apparent increase of five percent in the number of protocols having a mean reliability of 8 or better, (2) an apparent increase of seven percent in the number of responses scored reliable at the eight, nine or ten level, and an increase of .4 in the average reliability of all responses. More complete results for the test of this hypothesis may also be found in Table IV and Table V.

When the Accidental Effect of Splits is Removed

Hypothesis two stated that when differences between coders, in the decision of whether to divide a statement into more than one statement for coding purposes, are adjusted so that only the immediate statement is affected, rather than all subsequent statements in the protocol, reliability is expected to appear as greater than under realistic conditions. Hypothesis two also must be tested under a hypothetical condition which removes a real effect from consideration. The results obtained under that condition represent an apparent increase of eight percent (80 percent) in the number of protocols with mean reliability of eight or better. An increase of eight percent (79 percent) in the number of responses scored as reliable at the eight, nine or ten level was found. The mean reliability of all responses increased .40 to 8.59 and the responses accounted for by the modal score of 10 increased 5.9 percent to 44.4 percent. Clearly the apparent increase of reliability which was hypothesized has been obtained. These results are made more explicit in Table IV and Table V.

Factors Present in Low Reliability

At this point, to visualize what sorts of responses produce certain levels of reliability, you may wish to turn to Appendix C, which lists each item and protocol, in conjunction with its score under three conditions. The most unreliable protocol was #736, which reads:

> I am an individual common everyday person.
> Try to live up to the laws of America & pay my bills.
> Try to bring up my family & get as meddui (sic) as you I can.

The most unreliable statements when accidental effects of splits had been removed were:

1. My family.
2. I stay home on the week ends.
3. I'm a Negro and proud of it because its Gods (sic) intention.
4. I sing in the choir.
5. I am a person who believes a happy marriage is the basis for success.
6. I live alone.
7. My occupation.
8. I am a person who would like to become head of large business firm.
9. I love my family and want to provide for them any way I can.*

For one of these there were no agreements, and for eight of these statements there was only one pair of coders in agreement. It is interesting to notice that the categories involved in these very unreliable responses were repeaters, as can be seen in Table VI.

The only non-existent category (73) used by any coder was used in coding the one response which was scored as zero in reliability.

Reliability Related to Complexity Level of Code Category

In order to test hypothesis four as content categories increase in complexity, reliability will lessen, each item was scored on the following scale of complexity, if there were at least four of six coders in agreement on complexity level:

Complexity Score	Name	Categories
1	Personal Categories	10-17, 70-71,
2	Primary Categories	20-27
3	Secondary Categories	40-48, 52-55
4	Categorical Categories	30-31, 50-51, 56-59, 62-63
5	Catch-all Categories	60-61, 72

*These responses are from several protocols, not a single protocol.

Table VI

Code Categories by Frequency of Use
For Nine Responses Coded Least Reliably

Category	Description	Frequency
72	Other idiosyncratic (residual)	8
Non-split	(Not a category)	6
43	Qualified general reference to work	3
48	General religiosity references, (qualified and unqualified)	3
51	Qualified reference to nation, race, ethnic or language group	3
59	Qualified other categorical referents	3
62	Domestic references (unqualified)	3
63	Domestic references (qualified)	3
22	Marital terms (unqualified)	2
41	Qualified occupational title	2
58	Other categorical referents, (unqualified)	2
61	Qualified other consensual	2
21	Qualified kinship, excluding marital terms	1
23	Qualified marital terms	1
42	General reference to work, (unqualified)	1
44	Formal organizations, (unqualified)	1
45	Formal organizations, (qualified)	1
53	Qualified reference to religion, religious organizations	1
73	(A non-existent code category)	1

It was expected that a complexity score of five would produce about the same reliability score as a complexity score of one. The decision to treat these two scores separately and merge them later was made because they were unlike groups in many ways. The analysis for this hypothesis was done under only two conditions (1) the realistic condition in which no coding effects are removed from consideration, and (2) the hypothetical condition in which the accidental effects of splits were removed from consideration. It was inappropriate to test the hypothesis under the condition in which the effect of qualification was removed from consideration, since that made up a large part of the effect this analysis was concerned with. The results are reported here in tabular form. There seems to be no relationship of the type hypothesized,

Table VII

Modal Complexity Score of Responses
By N and $\overline{X}$ Reliability for two Conditions

Complexity Score	1	2	3	4	5	Total
No Coding Effects Removed						
N	43	204	166	102	231	*746
$\overline{X}$ Reliability Score	9.1	9.2	8.9	8.5	9.0	
Accidental Effect of Split Removed						
N	39	197	158	83	217	#694
$\overline{X}$ Reliability Score	9.2	8.8	8.6	9.2	8.9	

*Only responses judged by a majority of all coders to be on a single complexity level were analyzed, thus accounting for a decrease of 112 responses from the total N of responses.
#Removal of the accidental effect of split statements created enough redistribution of categories to lower the N of responses analyzed by an additional 59, while the total N increased by seven, thus the net loss of 52.

at least not when complexity is defined as it was here. It did not seem appropriate to run a statistical correlation of complexity score with reliability score as originally planned.

Reliability Related to Number of Statements in Protocol

Hypothesis three, the more statements made on a TST protocol, the lower will be the mean reliability score for that protocol, was tested by running a correlation between number of responses in the protocol and mean reliability score of the protocol. Since no specific condition was specified in the hypothesis, all three conditions were tested. The results of the three correlations are:

No Coding Effects Removed - $r = -.2177$, $(p < .01)$

Qualification Effects Removed - $r = -.04117$ (NS)

Accidental Effect of Splits Removed - $r = +.000590$ (NS)

The hypothesis is supported under the realistic condition. However, since the correlation is very nearly zero when the accidental effect of splits is removed, probably the negative correlation can be accounted for entirely by that accidental effect, rather than by any effect of "more responses appearing more complex to the coder," on which this hypothesis was also based.

Results of Testing for Independence of Coders

Since the reliability results reported here depend to a great extent on the assumption that the coders were independent, it seemed wise to test this assumption, in the form of a hypothesis. This test may be

thought of as a validity check of the reliability test which is being reported. To put it another way, if the level of reliability which we have measured and reported is actually a valid measure it must be measuring no contaminating variables along with "true" reliability. If we find that the coders are related in their agreements when they are contributing to unreliability, we could no longer assume that they were not related, by some contaminating variable, in their agreements which contribute to reliability. In such a situation, the level of measured reliability would be higher than the level of true reliability by some unknown amount of contribution from the contaminating variable.

Therefore, we translated our assumption of coder independence into the hypothesis that there will be no relationship in the pattern of coder agreements which contribute to unreliability. This hypothesis was tested by a Chi Square analysis of the N of paired agreements for 15 pairs of coders. It was tested under only two conditions and not tested under the condition in which qualification effect is removed. Three separate tests were run for paired agreements, triple agreements and a combination of these when they contribute to unreliability. "No relationship" was operationalized by using as the expected value for the Chi Square analysis an amount of agreements equal for each of the 15 pairs of coders, which is equivalent to the average obtained N of agreements across all 15 pairs.

Table VIII

X^2 Values for Coder Pair Agreements
Under Two Conditions
(df = 14)

	N	Expected Value	X^2	Significance
No Coding Effects Removed				
Paired Agreement	168	11	15.9	NS
Triple Agreement	239	16	6.1	NS
Paired + Triple	407	27	10.8	NS
Accidental Effect of Split Removed				
Paired Agreement	150	10	33.7	$p<.05$
Triple Agreement	210	14	3.0	NS
Paired + Triple	360	24	15.8	NS

It appears that there may be a contaminating variable present, which only becomes apparent after removal of the accidental effect of splits from consideration. In order to make more explicit the relationships between coders which may, unwittingly, be contributing to measured reliability, the relational model in Figure 1. was constructed. Here each solid line represents a positive relation and each broken line represents a negative relation between two coders, A,B,C,D,E, and G (Gallup). The values associated with each line indicate the amount and direction of difference between the obtained number of agreements and the expected (average) number of agreements.

Note that the shortest line represents the most positive relationship (+12), while the longest line represents the most negative relationship (-6). It was not possible to reflect the value of each relationship in the length of each line. To achieve this effect, it would be necessary to go to three, rather than two dimensions for the model. The most notable change would be the positioning of G at a point above this page.

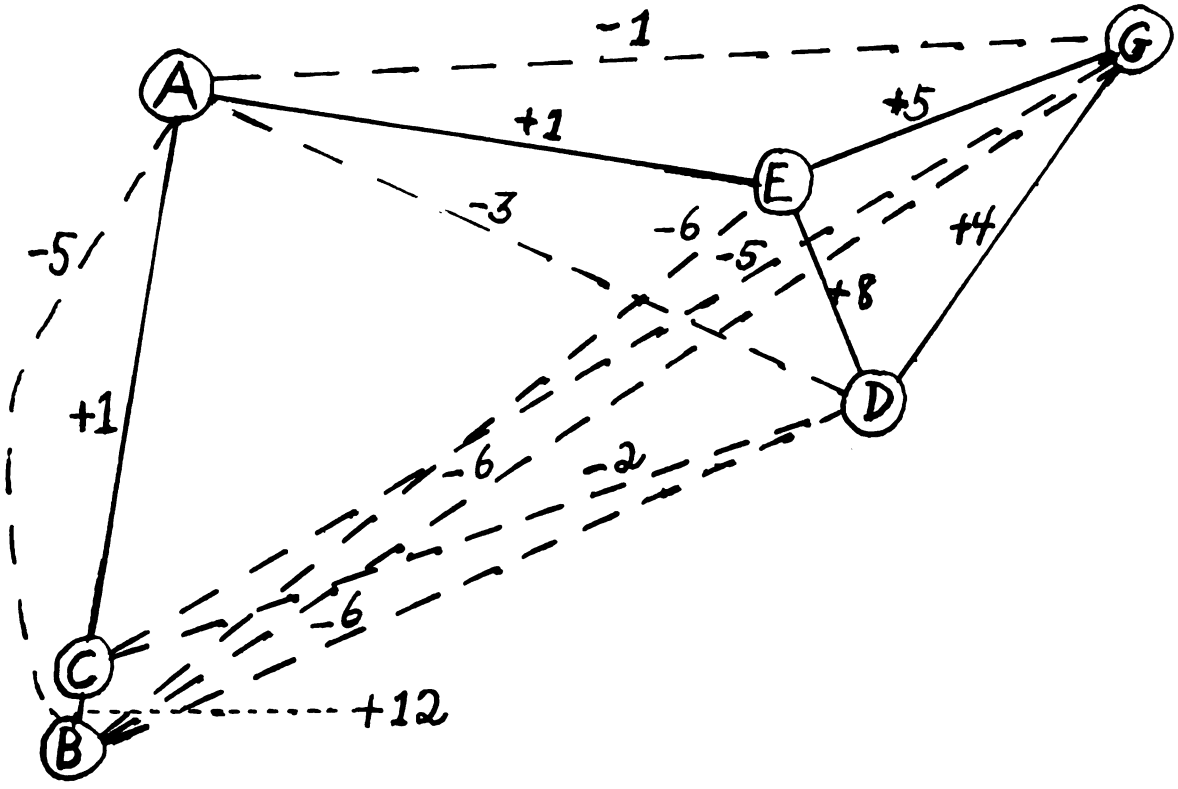

Figure 1. Model of Coder Relationships

This figure is based only upon the one Chi Square analysis which proved significant. In explanation of the strong positive relationship between C and B, and between E and D, I offer that C and B were two male

graduate students who had shared office space for several years, and that E and D are a married couple. Probably the contaminating variable unearthed here, is that these two pairs of the six coders tend to "think alike" due to habitual association and interaction. To this extent, then, the coders can not be thought of as independent. The somewhat strong positive relation between G and both E and D is without rationale, since these two pairs had no symbolic contact of any kind, while a number of written communiques regarding the TST coding procedures had passed between A and G.

CHAPTER IV

CONCLUSIONS

The Basic Hypothesis of Reliability

The basic hypothesis of high reliability, under the realistic condition in which no coding effects were removed from consideration was not supported at the level arbitrarily selected, i.e., 75 percent or more of all protocols, received a score of eight or better. The obtained results it will be recalled were that 72 percent of the protocols received a reliability score of eight or better.

Let's return for a second look at the basis for this hypothesis. In the first place, the decision to set the score of eight as an acceptance level, rather than a score of seven, while reasoned out in some detail, makes the assumption that a difference of one in paired agreements (between the score of eight and the score of seven) is somehow more important than a difference of one in the size of the mode (between the score of seven and the score of six). Therefore, we could also rationalize, after the fact, that the acceptance level could have been set at a score of seven, as easily as at a score of eight, and reference to Part B of Table IV shows that such an alteration would have produced high reliability in 85 percent of all protocols, well within the acceptance level of 75 percent or over.

In the second place, the hypothesis was based upon (1) greater complexity of code categories than had been used in the past to code the TST, plus (2) possibly a greater range in complexity of the TST response

protocols, due to a unique use of the TST on a heterogeneous sample rather than a homogeneous group, plus (3) the fact that testing reliability across six coders and 15 pairs of coders, rather than two or three coders as in past studies, was a more rigorous test of reliability, and finally (4) upon the obtained reliability results (97 percent) of Kuhn and McPartland.

From this line of reasoning it is possible to decide, as we did, to lower the expectation of reliability from the 97 percent level obtained in the past. However, there were no guidelines as to how much lower was low enough to account for the three differences in treatment. Therefore, we feel the obtained reliability figure supports our contention that the reliability figure should be lower than 97 percent. Beyond that one could rationalize that either (1) three percent is a very small difference from accurate prediction, or (2) we were mistaken in our arbitrary choice of a 75 percent acceptance level. We will not indulge in either line of rationalization, however, and the fact remains that reliability did not reach the level we thought it would.

The problem remains for other behavioral researchers to assess the processes which operate in arbitrary decisions regarding numerical levels. While there may be individual differences between researchers, I suspect that there are also cultural factors involved in such matters as (1) odd number versus even number, (2) how much is big, (3) how much is majority, (4) how much is little, (5) how much or how little is significant, etc. Obviously the most important cultural variables are the numerical base used and the habitual numerical operations reflected in monetary matters and pricing practices. I suspect in this instance

that a numberical base of ten influenced the arbitrary decision to set the acceptance level at 75 percent. Furthermore, had the researcher been raised in a culture in which the monetary system used pennies (.01), nickels (.05), dimes (.10) and tomens (.30), the arbitrary choice may well have been set at 70 percent. An experimental study of the psychological and social processes involved in setting numberical acceptance levels would be of importance to the study of philosophy of science.

Knowledge of Grammar related to Coder Reliability

Both of the hypothetical conditions studied are related to the coder's knowledge of grammar. The hypothesis that Removing the effect of qualification will operate to raise the appearance of reliability can be translated back to the belief that, at least for the coders involved in this reliability test, the coder competence was not at the highest possible level regarding ability to (1) identify adjectives, or (2) decide what sorts of clauses, phrases or allusions serve to modify the main clause. The hypothesis that removing the accidental effect, caused by non agreement between coders in the decision to split a response into two responses for coding purposes, will operate to raise the appearance of reliability was based upon the belief that, at least for the coders involved in this study, coder competence was not at the highest possible level regarding (1) ability to identify a compound sentence, and (2) ability to identify separate clauses associated with compound objects.

As pointed out before, the removal of these two effects was only hypothetical, and in each case the hypothesis was supported by the results.

From this we can conclude that if coders were employed whose levels of competence, at these four grammatical problems, were higher, then reliability would increase in a real, rather than an apparent fashion. While it is probably impossible to remove these effects entirely, due to the difficulties of the language structure, it should be possible to lessen these effects.

It seems an appropriate suggestion, that a replication of this study, using the same 150 protocols, but employing six high school English teachers, or six professional editors as coders would produce an even higher level of reliability than that obtained here, assuming adequate training in the code's content categories.

Independence Related to Coder Reliability

The hypothesis that there will be no relationship between coders in their agreements which contribute to unreliability, was based upon the assumption that the coders in the present study were independent. Recall that, at least after the accidental effect of splits were removed, this hypothesis was not supported in the test of paired agreements which contributed to unreliability.

A legitimate conclusion is that a replication of this study using coders who are selected in some manner so as to make them more independent would measure reliability at a lower level than the level measured in this study. For example, using the same TST protocols one might ask six graduate students in social psychology to act as coders, with the limitation that each graduate student be studying at a discrete university, and never have met any of the other coders.

In line with the suggested replication using coders of increased competence in grammar, one would expect reliability to increase if the six high school English teachers all knew each other and taught at the same high school, while one would expect little change in reliability if the six high school English teachers taught in separate cities or had never met; one would expect increased reliability if the six editors all worked in the same publisher's office, while one would expect little change in reliability if the six editors worked for discrete publishing companies and had never met. Because an increase in the grammatical competence of the coder is expected to increase real reliability, while an increase in coder independence is expected to decrease measured reliability, it is probable that the two factors would cancel each other out. Therefore, an increase in coder grammatical competence coupled with an increase in coder independence might well create little effect on the level of measured reliability.

Factors in the Present Code which Limit Reliability

The casual finding that certain code categories contribute to extreme low reliability in a disproportionately greater amount than other code categories, while still other code categories do not contribute at all to extreme low reliability suggests that there are real differences between code categories as to the reliability one can impute to their assignment to a response by an individual coder. Originally it was our intent to additionally study these differences in the several code categories, since, on the face of it, some code categories seem less clear-cut than others. However, our own bias places higher value on a

study which tests hypotheses than on a study which does exploratory counting, and there seemed no basis upon which to hypothesize this variable, the definitiveness of the code category. An attempt was made by testing the hypothesis that as complexity level of the category increased, reliability would decrease, but this hypothesis was not born out. Since the scale used to operationalize "complexity level" does not appear to be measuring what I referred to above as definitiveness of the code category, it was not really surprising that the hypothesis was not supported.

If some scale of definitiveness can be developed, then a hypothesis and corresponding test of the correlation of definitiveness level with reliability level could be run in a future reliability study. To point up levels of definitiveness, let us examine the actual code, as it appears in Appendix B. If as Berelson generalizes, the "more precise and complete the set of coding rules, the fuller the illustrations," (3, p. 514) the more reliable will be the coding; then this code is lacking only precision. Certainly the code approaches completeness and fuller illustrations would be beyond comprehension. Actually grave attempts at precision were made in the writing of this code. However, its very length, completeness, and fullness of illustration operate to nullify the attempt at precision.

Following are some of the inconsistencies in the code which lead to imprecision. Reference will be made by page number and row on the page, numbering rows from top to bottom.

1. On page 73 #9, row 3, compared with row 6 -- "I'd rather not answer that question." is labeled a refusal (99), while "I would rather keep that to myself" is labeled as idiosyncratic (72).

2. On page 73 #9, row 6, compared with 83 #61, last row--"I would rather keep that to myself," is labeled as idiosyncratic (72), while "I would rather keep this to myself." is labeled as qualified other consensual (61).

3. On page 75 #11, row 3, compared with 83 #61, row 9--"I am a young single girl" of which the portion "I am young" is labeled as qualified age (11), while "I am my plain old self" is labeled qualified other consensual (61).

4. Page 75 #14, row 7-- "I am very tanned" is inaccurately labeled other physical characteristic (unqualified) (14)

5. Page 76 #17, row 4, compared with 78 #40, row 8, and compared with page #42, row 6--"I'm a chore woman" is labeled a qualified sex reference (17), while "I'm a charwoman" is labeled occupational title (unqualified) (40), and "I am a working man" is labeled general reference to work (unqualified) (42).

6. Page 76 #20, row 9, 10 and 12, compared with #21, row 3--"I have eleven grandchildren" and "I am a mother of four daughters" and "I have four children -- three in school and one baby", are labeled kinship (unqualified) excluding marital terms (20), while "I raised a family of three" is labeled qualified kinship, excluding marital terms. (21).

7. Page 80 #48, row 6 and 7, compared to Page 81, #53, row 7--"I am a person who believes in God" and "I am a religious person," are labeled Religiosity, (qualified or unqualified) (48), while "I live by my religious beliefs" is labeled Qualified reference to religion and religious organizations (53).

8. Page 77 #25, row 3, compared to page 83 #61, row 14 and compared to 84 row 6--"I have lots of friends" is labeled Qualified reference to a friend (25), while "I am a friendly person," is labeled qualified other consensual (61) and "I am friendly" is labeled idiosyncratic (72).

9. Page 83 #61, row 2 and row 6, compared with page 85 row 16 -- "I'm not much of a talker" and "I keep my mouth shut" are labeled as qualified other consensual (61), while "I do not like to gossip," is labeled idiosyncratic (72).

10. Page 83 #61, row 5, compared with page 84, row 9 and 15-- "I am a happy person" is labeled qualified other consensual (61), while "I'm happy" and "I have always been pretty happy" are labeled idiosyncratic, (72).

11. Page 82 # 59, row 2 compared with Page 85 row 18 -- "I am interested in sports" is labeled qualified other categorical referents (59), while "I am interested in recreation" is labeled idiosyncratic (72).

12. Page 82 # 59, row 2, compared with page 84 row 22-- "I like music" is labeled qualified other categorical (59), while "I am creative in art" is labeled isiosyncratic (72).

These illustrations of some of the specific difficulties in the present code may indicate some of the confusion caused in the individual coder's mind by the code. One would expect that such confusions would give rise to unreliability in coding. More careful attention to the code could operate to reduce unreliability.

Length of the TST as a Factor in Reliability

The hypothesis that the more responses there were in a TST protocol the lower the reliability of the protocol would be was upheld by a negative correlation at a significant level under the realistic condition. This evidence points to a value in using only ten statements, rather than 20, as has been done in the present study and in Brookover's study (4). Such foreshortening of the TST should act as an aid to reliability.

Future Utility of the Twenty Statements Test

The ease of administration of the TST makes it a natural choice for assessing self data in the field situation, and it has particular value in any type of cross cultural study, where the question can be rather easily translated into equivalent forms in several languages.

Perhaps the greatest utility for the TST in the future is to compile a body of statements about the self which can in turn be scaled by a large sample of persons. Finally by submitting these scale items to factor analysis, it should be possible to construct a standardized

scale of self-attitude with factor loadings for each item. Such a standardized scale would have the advantage of the TST in being self-defined by a large, representative sample of selves, plus the added advantage of structured and pre-coded responses, which would make unnecessary any future studies of inter-coder reliability.

Another future product arising from the TST, as a generator of items, would be a Q-sort of items designed to measure self-attitudes. Such a standardized instrument would introduce the additional advantage of allowing the individual respondent to add a good deal of self definition to the measurement in the style in which he rank orders the items, while reliability remains enhanced by holding the possible responses constant.

Another innovation which should enhance the reliability of the TST is the self-coding form under development by McPhail (24). By placing the coding under the direction of the respondent, coder reliability becomes a meaningless concept.

In conclusion, let us look at some important ways this study differs from the original reliability check by Kuhn and McPartland: (1) the data being coded in the present study represented a wider range of expression, generated as it was by a heterogeneous rather than homogeneous group of respondents, (2) code categories being used represented 22 times as many possibilities for unreliability as the original Kuhn and McPartland code (45 categories vs. two categories), (3) six coders represent 16 times as many possibilities for disagreement as the two coders used in the Kuhn and McPartland reliability check, and (4) the present study set a standard for acceptance of reliability (75 percent of all protocols and responses should have a reliability score of eight or better) rather than accepting

whatever reliability level was obtained.

All of these differences make a high level of reliability more difficult to attain. The 72 percent level obtained vs. the 97 percent level obtained by Kuhn and McPartland appears encouragingly good. In spite of the fact that the results did not quite reach up to the a priori acceptance level, we feel the TST can be counted upon for very good inter-coder reliability even when heterogeneous populations are sampled, a complex code category system is used and a relatively rigorous reliability test is applied.

REFERENCES

1. Allport, G.W., Becoming: Basic Considerations for a Psychology of Personality, Yale University Press, New Haven, 1955.

2. American College Dictionary, Random House, New York, 1947.

3. Berelson, Bernard, "Content Analysis," in Gardner Lindzey (Ed.), Handbook of Social Psychology, V. I., Addison-Wesley Publishing Co., Reading, Massachusetts, 1954, 488-522.

4. Brookover, Wilbur B., Ann Paterson & Shailer Thomas, Self Concept of Ability and School Achievement: The Relationship of Self Images to Achievement in Junior High School Subjects, Office of Research & Publications, Michigan State University, East Lansing, Michigan, 1962.

5. Bugental, J.F.T. and S. L. Zelen, "Investigations into the Self-Concept. I. The W-A-Y Technique," Journal of Personality, V. 18, 1950, 483-498.

6. Cooley, Charles Horton, Human Nature and the Social Order, Charles Scribner's Sons, New York, 1902.

7. Couch, Carl J., "Family Role Specialization and Self-Attitudes in Children", Sociological Quarterly, April, 1962, 115-121.

8. _____, "Self-Attitudes and Degree of Agreement with Immediate Others," American Journal of Sociology, 1958, 491-96.

9. _____, Unpublished manuscript dealing with classes of verb forms used in response to a modified form of the Twenty Statements Test.

10. Dyer, Del A., Self Concept, Role-Internalization and Saliency in Relation to 4-H Leader Effectiveness, Unpublished Ph.D. Thesis, Michigan State University, East Lansing, Michigan, 1962.

11. Hartley, E. L., and R. E. Hartley, Fundamentals of Social Psychology, Alfred Knopf, New York, 1952.

12. Hickman, C. Addison & Manford H. Kuhn, Individuals, Groups and Economic Behavior, Dryden Press, New York, 1956.

13. Interviewers' Bulletin, Gallup Organization, publication #GO/6368 and #GO6369, September 3, 1963.

14. Kuhn, Manford H., "Self," in Julius Gould and William L. Kolb (Eds.), A Dictionary of the Social Sciences, Free Press, Glencoe, Illinois, 1964, 628-30.

15. _____, "Self-Attitudes by Age, Sex and Professional Training," The Sociological Quarterly, V. I, #1, January, 1960, 39-55.

16. _____ and Thomas S. McPartland, "An Empirical Investigation of Self-Attitudes," American Sociological Revue, 1954, V. 19, 68-76.

17. Lindesmith, A. R. and A. L. Strauss, Social Psychology, Dryden Press, New York, 1949.

18. Lipkin, S. "Clients' Feelings and Attitudes in Relation to the Outcome of Client-Centered Therapy," Psychological Monographs, 1954, V. 68, 1-30.

19. McPartland, Thomas, S. D., The Self and Social Structure, Unpublished Ph.D. Thesis, University of Iowa, Iowa City, Iowa, 1953.

20. McPartland, _____, John H. Cumming, Wynona S. Garretson, "Self-Conception and Ward Behavior in Two Psychiatric Hospitals," Sociometry, June, 1961, 111-124.

21. McPartland, _____ and John H. Cumming, "Self-Conception, Social Class, and Mental Health," Human Organization, Vol. 17, #3, 1958, 24-29.

22. Manual for the Twenty Statements Problem, Department of Research, The Greater Kansas City Mental Health Foundation, Kansas City, Missouri, 1959.

23. McPhail, S. Clark, Self-Identification Within a Specific Context of Experience and Behavior, Unpublished Ph.D. Thesis, Michigan State University, East Lansing, Michigan, 1965.

24. McPhail, Clark, Unpublished manuscript on development of a self-coding form of the Twenty Statements Test.

25. Mead, George Herbert, Mind, Self & Society, University of Chicago Press, Chicago, 1943.

26. Meltzer, Bernard N., The Social Psychology of George Herbert Mead, Division of Field Services, Western Michigan University, Kalamazoo, Michigan, 1959.

27. Mielke, Keith W. Evaluation of Television as a Function of Self-Beliefs, Unpublished Ph.D. Thesis, Michigan State University, East Lansing, Michigan, 1965.

28. Raimy, V. C. "Self-Reference in Counseling Interviews," Journal of Consulting Psychology, 1948, V. 12, 153-163.

29. Scheerer, E. T., "An Analysis of the Relationship Between Acceptance of and Respect for Self and Acceptance of and Respect for Others in Ten Counseling Cases," Journal of Consulting Psychology, 1949, V. 13, 169-175.

30. Schutz, W., Theory and Methodology of Content Analysis, Unpublished Ph.D. thesis, University of California at Los Angeles, 1950.

31. Stewart, Robert L., An Empirical Test of the Reference Group Hypothesis, Unpublished M.A. thesis, University of Iowa, Iowa City, Iowa, 1952.

32. Tucker, Charles W., Occupational Evaluation & Self-Identification, Unpublished Ph.D. Thesis, Michigan State University, East Lansing, Michigan, 1966.

33. Vargas, M. J., "Changes in Self-Awareness During Client-Centered Therapy," in C. R. Rogers & Rosalind F. Dymond (Eds.), Psychotherapy and Personality Change, University of Chicago, Press, Chicago, 1954, 145-166.

34. Webster's Elementary Dictionary, American Book Company, New York, 1941.

35. Wylie, Ruth C., The Self Concept, A Critical Survey of Pertinent Research Literature, University of Nebraska Press, Lincoln, Nebraska, 1961.

36. Bugental, J. F. T., "A Method of Assessing Self and Not-Self Attitudes During the Therapeutic Series," Journal of Consulting Psychologists, 1952, V. 16, 435-439.

37. Stock, Dorothy, "An Investigation into the Intercorrelations Between the Self-Concept and Feelings Directed Toward Other Persons and Groups." Journal of Consulting Psychologists, 1949, V. 13, 176-180.

38. Raskin, N. J., "An Objective Study of the Locus-of-Evaluation Factor in Psychotherapy," in W. W. Wolff & J. A. Precker (Eds.), Success in Psychotherapy, 1952, New York, Grune & Stratton, 143-162.

39. A Manual for the Analyzed Data from United States and Mexico, Charles W. Tucker, principal editor, Michigan State University, Sociology Department, East Lansing, Michigan, September 18, 1964, mimeo.

APPENDIX A

Form of the TST in the Interview Schedule

APPENDIX A

Form of the TST in the Interview Schedule

INTERVIEWER: Do not spend more than three minutes on the next item. Probe until you get at least five statements from the respondents. When you must probe, indicate with a horizontal line below the number of his last statement.

Now, we have quite a different thing for you to do. Although probably new to you, it is easy and I think you will find it quite enjoyable. Everyone we have asked to do this has found it to be interesting. Now, let me tell you what we have in mind.

Ask this question of yourself, "Who am I?" Think of as many answers as you can in answer to the question, "Who am I?"

In a moment I would like you to give me the answers as if you were giving them to yourself, not to me or anyone else. Take a little time to think about it. (PAUSE HERE MOMENTARILY.)

a. Now, please make what you consider to be the most important statement about yourself first.
1. _____

b. Now, make what you consider to be the next most important statement about yourself.
2. _____

c. Now, make what you consider the next most important statement about yourself.
3. _____

d. (PROBE) Are there any other statements you could make about yourself in answer to the question, "Who am I?"

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

APPENDIX B

TST Code Book

READ CAREFULLY - GENERAL CODING INSTRUCTIONS FOR ITEM 17 - Who am I?

1. The object is to code all statements as separate items. A STATEMENT CAN BE A SINGLE WORD, A PHRASE, OR A COMPLETE SENTENCE. Thus, you may find more than one statement on a particular line in the questionnaire. Each such statement will be coded separately.

   Examples: a. carpenter; a happy fellow; I am a good father (This is three separate statements.)
   b. I am a father and husband. (This is two statements.)
   c. I am a very hardworking, unhappy father. (This is one statement.)
   d. I am handsome but weak. (This is two statements.)
   e. I am a young single girl. (This is three statements.)
   f. I am a American Negro citizen. (This is three statements.)

   GENERAL RULE: Any element which can be construed as a single thought about oneself will be coded as a statement. A noun and its accompanying modifiers is usually a single thought. Adjectives modifying nouns are to be included with the noun (as in example c. above). Adjectives which stand alone as self descriptions (as in example d. and e. and f. above) will be counted as statements.

2. Go through question 17 and mark off statements. Usually only one statement will be found on a line. In those cases where more than one statement is found on a line, code each of the statements in order, beginning with the first mentioned statement.

3. Each statement will be coded into a two column code described below. When you get to ten statements, ignore the rest. In case there are less than 10 statements, put code 88 into the unfilled columns.

4. Coding of this item can best be done by following two procedures:

   a. Matching statements on the protocols with statements provided as examples for each category (see following pages.)
   b. Following these General instructions with special attention to the Arbitrary Rules (see following pages.)

5. The titles of the categories are only suggestive of the range of statements to be included in each category. Basically, the category serves to collect statements which have some common referent. However the logic of placing certain statements in the same category may not always be apparent. Attention to such questions as "What did the respondent really mean to say?", or "How can this category placement make sense in the context of what the respondent said in another statement?", will lead to random coding, and low inter-coder reliability. Therefore, the best check on where a statement is to be coded is to try to find an example like the statement in the list of examples, or one clearly close to it. A final check should be made to make certain that the category assignment does not violate any of the arbitrary rules and exceptions.

GENERAL CODING INSTRUCTION - Who am I? (Continued)

6. Definition of a qualified and unqualified statement:

   a. A QUALIFIED statement is any statement which is modified by the inclusion of an adjective or adverb, along with the noun and verb. Sometimes a verb will modify a statement, as in the statement "I hate work."
   b. An UNQUALIFIED statement is any statement which is not modified by the inclusion of an adjective or adverb. This would be only the noun or verb, in cases where the verb has no qualifying effect.

   Examples: I am a poor student. (This is a qualified statement.)
   I am a father. (This is an unqualified statement.)

7. Each statement should be coded disregarding (a) negation or affirmation and (b) verb tense.

   Examples: Treat the statement "I am not a mother." as in the same class as "I am a mother."
   Treat the statement "I was a foreman." as in the same class as "I am a foreman."

8. The distinction between "consensual" and "idiosyncratic" statements can best be understood by reading through the list of examples provided. All of the references in the list of idiosyncratic statements are to events, states of mind, or processes which are private, lacking in specificity, or lacking in general agreement as to meaning. On the other hand, all consensual statements have at least one word which points to something most people would recognize, and have some agreement for the meaning of.

9. Use a 99 to code all columns in the case of a clear refusal. We define a refusal as one of the following:

   a. I'd rather not answer that question.
   b. I don't want to answer that question.
   c. Would you please go to the next question.

   "I don't know." and "I would rather keep that to myself" are not defined as refusals to the question but as responses to the question which should be coded as Idiosyncratic (72)

ARBITRARY RULES

A. In cases of statements which fit into two or more categories, assign the statement to the category closest to the top of the list numerically, with the following exceptions:
   1. Any reference to health, whether consensual or not, should be coded as a 70 or 71, depending on the matter of qualification.
   2. Any reference to an organized church or religion should be coded as 53 or 52 regardless of what other codable items are mentioned in the statement.

GENERAL CODING INSTRUCTIONS - Who am I? (Continued)

3. Any reference to race, irrespective of other references in the statement which are codable, should be coded 50 or 51.

4. General religiosity references frequently include words that refer to kinship, sex, work, etc. However, all religiosity references should be coded as 48. (see examples)

B. Code the statement as it is recorded on the schedule. Do not add or subtract words or impute meanings. The only exception is described in rule C below.

C. In compound sentences with a single stem, treat the clauses as if each were separate statements with the stem attached to each clause.

Examples 1. "I am interested in my house, in recreation, and in education." should be treated as the following three statements:
a. I am interested in my house.
b. I am interested in recreation.
c. I am interested in education.
2. "I am an American Negro citizen." should be treated as the following three statements:
a. I am an American.
b. I am a Negro.
c. I am a citizen.
3. "I am a young single girl." should be treated as the following three statements:
a. I am young.
b. I am single.
c. I am a girl.

D. In cases where titles are mentioned which include several words, treat the entire title as the object of the sentence.

Examples: 1. I am a U.S. citizen (Do not treat U.S. as qualifying citizen.)
2. I am an executive secretary (Do not treat executive as qualifying secretary)
3. I am a retired boilermaker (Do not treat retired as qualifying boilermaker.)
4. I am a successful medical doctor. (Treat successful as qualifying medical doctor)

E. When coding initials (e.g., B.P.O.E., G.O.P., I.O.O.F., etc.) which have not been defined by the interviewer on the schedule, <u>all initials</u>, whether they are recognized or not recognized by the coder, should be coded as 58 or 59 depending upon qualification.

Example: "I am a B.P.O.E." (This is an unqualified statement - 58)
"I am a good B.P.O.E." (This is a qualified statement - 59)

TST CODE

10 - Age (unqualified) Only include specific age such as "I am 25 years old." Statements referring to "old" or "young" are NOT to be included in this category but are coded as 11.

I am 25 years old

11 - Qualified age

I am almost 35.
I am (a) young (single) (girl)

*Note: This is the first example of a sentence in which there is more than 1 codable referent. In this case "I am young" is coded as 11; "I am single" is coded as 22; and, "I am a girl" is coded as 16.

12 - Name This would be the first, last or both names with or without the prefix of Mr., Mrs., or Miss.

I am Ardith Weaver
I am Willie Ella Barnell
My name is Joseph Smith

13 - Address This would include only street address and not the name of the community or nation.

I live at 712 Creek Road

14 - Other Physical Characteristics (Unqualified) This could include references to height, weight, skin color, eye color, etc. This does not include "colored", "negro", etc., which is coded under 50, 51.

I am bald-headed
I am 5' 2" tall.
I am very tanned.

15 - Qualified physical characteristics

I am nearly 5'2" tall.
My hair is turning gray.

16 - Sex (unqualified) This would include the words male or female. It would also include the words; man, woman, boy, girl, lady, and gentlemen.

I am a man
I am a female
I am a woman

17 - Qualified Sex Reference

I am a contented woman
I am a nice gentleman
I'm a chore woman
*I am a man who is loved by his friends
I am a good man
I am a self-made man
I am a fortunate man
I am a man able to help my children
I am a free man.
I'm the man of the house.
I am a man who does a lot of public service work free.
You are a nice gentle woman that paid me a visit.

*In the fourth statement there are two possibilities of coding: friends or sex ref. categories. Following Rule 1 the statement is pushed up to the first of the two possible categories.

18 - OPEN

19 - OPEN

20 - Kinship (unqualified) Excluding marital terms. Including such terms as family, child, parent, relative, daughter, son, father, mother, sister, brother, uncle, aunt, grandfather, grandmother, great aunt, great uncle, godmother, godfather and all in-law referents. Do not count number of children or relatives as a qualification.

I am a grandmother
I have children
I am a mother
I have eleven grandchildren
I am a mother of four daughters
I am a father to my children
I have four children -- three in school and one baby.
I am making a living for my family.

21 - Qualified Kinship. Excluding marital terms.

I'm a good relative.
I raised a family of three.
I come from a good family.
I have a very happy family.
My family is all gone.
I can educate my children.
I am a referee for my daughter.
I am proud of my kids.
I am interested in my children's education.
I hope to have made my folks happy in my life.
I am a farmer's daughter
I am not the same as I would be if my mother was living.
I enjoy being close to my family.
I am loyal to my family.
The future lies in the hands of mothers
I, as a mother, am responsible for the future generation.
My family is the most to me.

I am a family man.
I am a devoted family man.
As a mother, I can't be done without.
And my children whom I couldn't do without.
I am a person whose parents are both working.
I am not the person I would be if I didn't have a family.

22 - Marital terms (unqualified) This would include the terms husband and wife, married and single, and those referring to marital status including divorced and separated.

I am a wife.
I am single.
I am divorced.
I am married.
I am my husband's wife.

23 - Qualified Marital Terms

I'm a wife of a progressive businessman.
I lost my wife.
I am proud of my husband.
I'm going to be married.
I have a nice wife.
I am not the person I would be if I wasn't married.
I married a farmer.
I'm important to my husband.
Financially, I like to feel I am helping my husband.
Ask my wife, she would say I am a gun nut.
I help mate financially.

24 - Reference to Friend (unqulaified) This would include those statements with the word "friend" in them. A statement referring to the respondent's own friendly quality would not be included, e.g., "I am friendly" would be coded as 72.

I am a friend.
I have friends.

25 - Qualified Reference to a Friend

I have many friends.
I have lots of friends.
I am people's friend.
I am a friend to other men.
I'd like to make more friends than I have.

26 - Other (Residual) Primary Terms (unqualified) This would include neighbor, have neighbors, peers, bridge clubs, buddy, pal, colleague, associate, companion.

I am a neighbor.
I am a companion.

27 - Qualified Other (Residual) Primary Terms

I hope that I am a good neighbor
I am a good neighbor
I have never had much trouble with neighbors.

28 - OPEN

29 - OPEN

30 - Reference to Community (Unqualified) This would include any reference to the respondent's town, city, community, or state.

I live in New York.
I am a Californian.
I am a Michigander.

31 - Qualified reference to community

I am from a poor community.
I am a citizen of a nice community.
I like to mix with community groups.
I am an example in the community.
I am loyal to my community.
I am well-liked in the community.
I am a community service worker.
I am a community worker.
I am a good community supporter.
I am a civic minded person.

32 - OPEN

33 - OPEN
34 - OPEN

35 - OPEN

36 - OPEN

37 - OPEN

38 - OPEN

39 - OPEN

40 - Occupational Title (Unqualified). This would include all names of specific occupations including the term "laborer", housewife.

I'm a dish-washer
*I'm a retired housewife
I am a seamstress
I am a punch press operator
I am a beautician
I am a musician
I am a teacher
I am a chauffer
I'm a painter
I am a pastor
I am a housewife
I'm a char-woman

*"Retired" in conjunction with an occupational title is not, for our purposes, considered as a qualifier of occupation and should be coded as 40.

I am a post-master
I am a banker
I'm a filling station operator
I'm a carpenter
*I am a retired truck driver
I am a farmer
I am a nurse
I'm a counselor
I'm an entertainer
I am an educator
I'm a lawn mower
I am a gardner
I'm a knitter
I am a civil worker
I am a cook
I am a spiritual advisor
I am a nursemaid
I am a political worker
I am a foreman on my job
I am a religious instructor

41 - Qualified Occupational Title

I am an enthusiastic advertising man.
I am a happy housewife
I am a very average businessman
I am a good housekeeper
I hope to become a good author.
I ran a foster home for five years in New York state.
I'm a housewife, all inclusive.
I'm a farmer, still able to do a day's work.
I don't like housework.
I'm tops at housework.
I love housework.
I just finished some housework.

42 - General Reference to Work (Unqualified). Includes reference to work, worker, retired, unemployed.

I work.
I am a laborer.
I work 10 hours a day.
I am a working man.
I'm retired.
I am employed.
I am making a living.

43 - Qualified General Reference to Work

I am a worker by trade.
I am a good worker.
I am a poor working man.
I'm dedicated to my work.
When I work and am hired to work, I do the best of my ability.
I don't shirk my job.
I work everyday.
I'm a hard working man.
I'm a good employer.
I like my work.
I am fair and average in my work.
I like to work.
I am strong in buying and selling property.
I enjoy making a living
I am good at work.

*"Retired" in conjunction with an occupational title is not, for our purposes, considered as a qualifier of occupation and should be coded as 40.

I live by the sweat of my brow.
My job comes first.
I have pride in my work.
You have to (work) if you want to keep even.

44 - Formal Organizations (Unqualified) This would include all educational, professional, civic, social, farm and labor organizations. THIS WOULD NOT INCLUDE ANY RELIGIOUS OR POLITICAL ORGANIZATIONS. All initials should be coded as 58 or 59 depending on the matter of qualification.

I'm a Rotarian.
I'm a Mason.
I'm a class sponsor.
I am a Boy Scout Committeeman.
I am a member of the Eastern Star.
I am a member of organizations.

45 - Qualified Formal Organizations. Excluding Religious and Political Orgs.

I am a loyal unionist.

46 - Reference to Education (Unqualified). Excluding occupational titles such as teacher, instructor, sponsor, advisor, etc.

I am a student.
I would say I am a student.

47 - Qualified Reference to Education

I enjoy studying.
I am a tired student.
I am going to get my BA degree next winter.

48 - Religiosity (Qualified or Unqualified) Excluding references to church, formal religion or religious organizations.

I am one of God's workers put here to make this world better than when I came here.
I try to serve the Lord in every way possible.
I am a person who believes in God.
I am important to the Lord as an individual
I was blessed by the Lord in everything.
*I am a Christian lady.
*I am a child of God.
*I am one of God's children.
I lead a Christian life.
I am a creature created by God.

*Where kinship referents are used here, (child, children), and sex referents, (lady) disregard usual coding by kinship and sex and place all such referents to religiosity in 48.

49 - OPEN

50 - Reference to Nation, Race, Ethnic or Language Group (Unqualified). This would include reference to nationality and citizenship. American, English, Finnish, Negro, White, Spanish (speaking) etc. Any referrent to being Jewish should be coded as either 52 or 53 depending upon qualifications.

I'm English.
I am a Negro.
I am a white woman
*I am a American (Negro) (citizen)
*I am a (American) Negro (citizen)
*I am a (American) (Negro) citizen.
I am a (middle-class) citizen.

*Sentence containing more than 1 codable referent. In each case here, the code is 50.

51 - Qualified Reference to Nation, Race, Ethnic or Language Group

I am a good U.S. Citizen
I am a common everyday American
I am a privileged American
I'm glad that I am an American by choice
I am a person with opportunities unequaled in U.S.
I am as deep as anyone in the progress of our country
I am glad to be in America

I am a visitor to this country

52 - Reference to Religion and Religious Organizations (unqualified) This would include all church or church-related groups, membership and participation in same. Jewish would be included in this category. Initials that can not be recognized by the coder should be coded either 60 or 61 depending upon qualification.

I am a Catholic
I'm a church member
I am a church goer
I go to church.
I am Jewish

53 - Qualified Reference to Religion and Religious Organizations

I am a good church member
I am important to my church
I am loyal to my church
*I enjoy my church work.
I am free to choose my religion
I live by my religious beliefs.
I go to church every Sunday.
I go to church regularly.

*In the fourth item "work" is considered here as a modifier of church. Church is categorized as referring to Religion Organization and it is thus coded as 52 or 53 depending on qualification. It should not be pushed up to 42 or 43.

54 - Reference to Politics and Political Organizations (Unqualified) This would include reference to political, politics, all political parties, political action groups and affiliated clubs and associations. Initials that can not be recognized by the coder as falling in any category should be coded either as 60 or 61 depending upon qualification.

I am a voter.
I am a Democrat

55 - Qualified Reference to Politics and Political Organizations

I'm a poor politician
I am an unhappy Democrat
I'm interested in national and international concerns
I feel myself as much concerned about public life as any other human
I am able to vote

56 - Reference to Socio-Economic Status (Unqualified) This would include working class, lower class, middle class, upper class, rich, poor, wealthy or well-to-do. An evaluation by the respondent of working ability will not be included.

I am middle-class
I am in the upper-class
I am (a) middle-class (citizen)

57 - Qualified Reference to Socio-Economic Status

I am making an average income
I don't have much money

58 - Other Categorical Referents (Unqualified) These would include such referents as sports fan, athelete, lover, member, hobbies.

I belong to the AFL-CIO
I am a server.
I am a stamp collector

59 - Qualified Other Categorical Referents

I am interested in sports
My hobbies are gardening
My hobbies are bowling
I do serve a lot
I'm an outdoor person
I like music.
I love dogs.
I love German shepards.
I'm interested in sports.

60 - Other Consensual (unqualified) These would include such referents as: person, individual, human being, myself, me, human, homo sapien.

I am an individual
I am me

61 - Qualified Other Consensual

I'm not much of a talker
I am law abiding
I try to better myself
There is no one like me
I keep my mouth shut
I'm a very conscientious person
I am a mere individual
I am my plain old self
I am an average person
I am a person brought into the world to do something
I am a person who lives in the future
I am a friendly person
I am a very nice person
I'm a very unimportant person
I'm an individual with my own beliefs
I am a busy person
To me I'm just myself
I am a person who forgives easily
I am a good person
I would rather keep this to myself
I'm a person who enjoys living in the present
I am not the person I'd like to be
I am a happy person
I am a common person
I'm a good natured person
I am an individual who thinks for himself.
I am a good guy

62 - Domestic References (Unqualified) including such statements as I cook good, I can food, I wash the car, etc.

I am a provider
I am a homeowner
I am a breadwinner
I own my own property
I don't owe any debts
I am a home body
I have no car
I pay my bills
I am a tax payer
I work in the yard

63 - Qualified Domestic References

I don't like to cook
I keep my place up so I can be proud of it
My home is my haven
I love my home
I love canning
I love food
I am interested in my home
I like to sew
Around the house, I am actually lord and master.

64 - OPEN

65 - OPEN

66 - OPEN

67 - OPEN

68 - OPEN

69 - OPEN

70 - References to Health (Unqualified)

I am healthy

71 - Qualified References to Health

I am in good health
I am not in good health
*I am a sick person
It is important to help ill and handicapped people

*Any time there is a reference to health it is coded as 70 or 71 regardless of referents such as person, begin, individual, man, woman, etc.

72 - Idiosyncratic

I help those in need
I'm not a hypocrite
I am given the opportunity to choose
I'm sociable
I'm friendly
I am one who leads a life morally and financially successful
I'm happy
I am easily satisfied
I have responsibilities
I am not enthusiastic
I have been beat out of some money
I am fair
I have always been pretty happy
I am different
I am dissatisfied
I am contented
I am not too smart
I'm not too dumb
I am congenial
I am concerned about my position
I enjoy making other people lead a fuller life
I like people
I have everything else
I was lucky in every way
I'm not too much really
I like to be friendly with everybody
I try to be nice to other people
I try to treat everyone as I like to be treated
I like to see everyone happy
I am concerned for my fellow man
I have room to improve
I abide by peace
I like my comfort
I am a small amoeba
I am uncertain
I am confused
I am not very confident
I am honest
I am upright
I am square
I am creative in art
I have an optimistic attitude
I am urged to do things I can't and I become frustrated

I am honest
I don't steal
I am agreeable
I can carry on
I have things done on time
To be honest is always the first thing that comes to mind
I have always been helpful
I go out of my way to help others
I believe that I cast an influence for good
I am a very small ingredient of a vast mixture
I like to do things
I cooperate with everyone
I think
I am a nervous wreck
I am the same as everybody else
I am interested in recreation
I am good to animals
I think good thoughts
I have a clean mind
I'm morally clean
I think on clean and noble things
I do not live in the past
I am important to life
I am glad to be able to do what I am doing
I'm glad to be able to be in this apartment
I get emotional about little things
I have a temper
I am stubborn
I am not much of anybody right now
I don't pattern after anyone else
I appreciate it
You are bringing me something
I feel humble and honored
Actually we are so insignificant today
I think I have a place no one else can get to which gives me strength
I do not like to gossip
I am being of service to others
I like to be around other people
I can carry on
I have my own individuality
I love helping other children
Iam a nobody
I am a free soul
I am trying to get along with everybody
I love to take care of people

Column Number: 13, 14

Page Number: 4

Question Number: 17

Item Description: Now, we have quite a different thing for you to do. Although probably new to you, it is easy and I think you will find it quite enjoyable. Every one we have asked to do this has found it to be interesting. Now let me tell you what we have in mind.

Ask this question of yourself. "Who am I?" Think of as many answers as you can in answer to the question, "Who am I?"

In a moment I would like you to give me the answers as if you were giving them to yourself, not to me or anyone else. Take a little time to think about it.

Code for first statement (unit of thought).

Coder note: This item will hereafter be referred to as "Who am I?". There will be two columns for each unit of thought (see General Instructions).

Codes:

%	N	
_____	_____	10 - Age (unqualified)
_____	_____	11 - Qualified age
_____	_____	12 - Name
_____	_____	13 - Address
_____	_____	14 - Other physical characteristics (unqualified)
_____	_____	15 - Qualified other physical characteristics
_____	_____	16 - Sex (unqualified)
_____	_____	17 - Qualified sex
_____	_____	18 -
_____	_____	19 -
_____	_____	20 - Kinship (unqualified and excluding marital terms)
_____	_____	21 - Qualified kinship (excluding marital terms)
_____	_____	22 - Marital terms (unqualified)
_____	_____	23 - Qualified marital terms
_____	_____	24 - Reference to friends (unqualified)
_____	_____	25 - Qualified reference to friends
_____	_____	26 - Other (residual) primary terms (unqualified)
_____	_____	27 - Qualified other (residual) primary terms
_____	_____	28 -
_____	_____	29 -
_____	_____	30 - Reference to community (unqualified)
_____	_____	31 - Qualified reference to community
_____	_____	32 -
_____	_____	33 -
_____	_____	34 -
_____	_____	35 -

Column Number: 13, 14

Page Number: 4

Question Number: 17

Item Description: "Who am I?"

Code for first statement (unit of thought)

Codes:

%	N	
_____	_____	36 -
_____	_____	37 -
_____	_____	38 -
_____	_____	39 -
_____	_____	40 - Occupational title (unqualified)
_____	_____	41 - Qualified occupational title
_____	_____	42 - General references to work (unqualified)
_____	_____	43 - Qualified general references to work
_____	_____	44 - Formal organizations (unqualified)
_____	_____	45 - Qualified formal organizations
_____	_____	46 - References to education (unqualified)
_____	_____	47 - Qualified references to education
_____	_____	48 - General religiosity references (qualified and unqualified)
_____	_____	49 -
_____	_____	50 - Reference to nation, race, ethnic or language group (unqualified)
_____	_____	51 - Qualified reference to nation, race, ethnic or language group
_____	_____	52 - Reference to religion, religious organizations (unqualified)
_____	_____	53 - Qualified reference to religion, religious organizations
_____	_____	54 - Reference to politics, political organizations (unqualified)
_____	_____	55 - Qualified reference to politics, political organizations
_____	_____	56 - Reference to socio-economic status (unqualified)
_____	_____	57 - Qualified reference to socio-economic status
_____	_____	58 - Other categorical referents (unqualified)
_____	_____	59 - Qualified other categorical referents
_____	_____	60 - Other consensual (unqualified)
_____	_____	61 - Qualified other consensual
_____	_____	62 - Domestic references (unqualified)
_____	_____	63 - Qualified domestic reference
_____	_____	70 - References to health (unqualified)
_____	_____	71 - Qualified references to health
_____	_____	72 - Other Idiocyncratic (residual)
_____	_____	88 - No statement to be coded in these columns
_____	_____	99 - Refusal

APPENDIX C

150 TST Protocols

Section I

Protocols Scoring the Same Under All Three Conditions
N = 44

*Where no score is noted for a response it is understood to have a perfectly reliable score (10)

9.6 Mean Score

I am a mother.
I am a wife.
I am a woman.
9 I am an artist.
9 I am a friendly woman.

10.0
I am a religious man.
I do the best I can.

10.0
I am a beautician.
I am a housewife.
I am a mother.

9.2
I'm a worker.
I'm a good mother.
I'm a faithful mother.
I am the breadwinner here.
6 I am a clean housekeeper.

8.8
Disgruntled man.
Lost confidence in life.
Ambitious for son.
5 Bitter at stupidity of gov't.

9.8
I'm myself.
A housewife.
A mother.
I like fun.
9 I do the best I can, in all I do.
A cook.

10.0
I am a person.
I am a mother.
I am a wife.
Teacher.
Companion.
Friend.

9.3 Mean Score

I am a wife.
I am a mother.
7 I am a companion.
I am a teacher.
I am a nurse.
9 That's about all I can think of.

8.0

8 I am a good family man.
5 I am law abiding.
I am very healthy.
9 No, I have scraped the bottom of the barrel now.

9.1

8 I'm a business man.
I'm a public accountant.
8 I love tax work (income tax).
9 I'm a school board member.
9 I'm a school board member.
9 I love sports.
I think I'm a good father -- (3 children).
Like to play golf.

8.5

9 An honest man.
A religious man.
9 Member of Fraternity.
6 Family man.

10.0

I am me.
I am nothing.
I speak what I think.

9.5

Child of God.
Good husband.
Good father.
7 Good name.
Good friend.
Religious.

9.4

I am a man.
I am honest.
I work hard.
7 I don't drink.
I do unto others like I would have them do unto me.

9.2

I am a citizen.
I am an everyday person.
9 I am trying to make a decent home for my family.
9 I am a father.
8 I am a taxpayer.

10.0 Mean Score
Housewife.
Mother.

9.5
I am a wife.
A mother.
8 A home-maker.
It feels good to be important to those I have chosen to make a part of my life.

8.8
I'm a mother.
I'm a wife.
7 I'm a clean person.
8 I'm a nutritious cook.
My church is important.
6 It's important to me to have recreation outside the home.
9 It is necessary for me to have my friends in at least once a week.
9 I help my children with homework.
Another important thing is to have a reliable baby sitter.

9.2
10 I'm a housewife & mother.
10
9 I'm a cook.
9 I'm a seamstress.
8 I enjoy reading.

9
9 I'm a small gear in a big machine.
8 I'm wrangler - (my nickname).
I'm free as heck.

8.8
A mother.
A wife.
8 Home maker.
Waitress.
6 Gemocratic. (sic)

9.8
I am an American.
I am a Negro.
I am healthy.
9 I am honest.

9.8
I guess the most important my honor.
Concern for others.
Well I guess answered another in general form.
Difficult to think of another one.
9 I'm someone who tries to set an example for the youth.

8.8 Mean Score
I am an American.
I am a father.
I am a husband.
5 I am a leader.
9 I am well adjusted.
9 I am a thinker.

10.0
Human being.
Negro.
Husband.
American.

8.2
9 I am an individual.
I am talented person.
9 I am good housewife.
5 I am good dresser.
8 I am good cook.

9.0
I am a mother.
I am a wife.
I am a grandmother.
I am trying to help people.
9 I am making best of my life and ability.
5

10.0
I am a human being.
I am an individual.
I am a creature of God.
I am a husband.
I am a father.

10.0
Mother.
Individual.
Citizen.

10.0
I am a mother.
D.K.

10.0
Wife.
Mother.
Home.

10.0
I'm me.
A child of God.

9.5 Mean Score
I am a housewife.
8 A homemaker.
A mother.
I am a Baptist.

9.8
I am a father.
A Husband.
Teacher.
8 A Friend.
A Scholar.
A Citizen.
A human Being.
I'm a man.
A son.
A brother.
A writer.
A camera bug.

9.5
I am a Father.
I am a Husband.
I try to be a good Church worker.
7 I think I am a good provider.
I try to be a good citizen.
I am a Grand Father.

10.0
An average human being.
D.K.
D.K.
D.K.

9.6
9 Just a small cog in a big wheel.
9 I am not too important.
I am not ambitious.
I am a man.
I do my best.

10.0
Steady worker.
Dependable.
Reliable.
Young.
Fortunate.

8.2
I am a salesman.
I am in the middle class bracket.
2 I am a resident of a very nice neighborhood.
9 Own my home.
I am satisfied with what money I'm making.

7.5 Mean Score

Getting along very well with majority of people I know.
If I do a job I do it to the best of my ability.
5 No.
5 No.

10.0

I am mother.

9.5

Wife.
Mother.
9 Competent.
9 Loved.
9 Country I love.
Wonderful future.

9.5

I am a man.
I do my best at my job.
9 I try to help people.
9 None.

8.3

9 A mother.
An American.
I'm a Presbeterian. (sic)
9 I am easy going.
I am friendly.
7 I like kids.
5 I like to be by myself.
7 I'm a thinker, not a doer.
8

Section II

Protocols Scoring Differently When Qualification Effect* is Removed N = 53

---

*Qualification effect is the effect caused by disagreement between coders as to the interpretation of a response in regard to qualification, modification or evaluation vs non-qualification, non-modification or non-evaluation. Below is one example of what is actually happening when qualification effect is removed:

ASSIGNED CATEGORIES UNDER REALISTIC CONDITIONS:

Response Number:	Coder G	A	B	C	D	E	Score	Verbal Response:
1.	48	48	48	48	(52)	(52)	8	Christian.
2.	22	22	22	22	22	22	10	Wife.
3.	20	20	20	20	20	20	10	Mother.
4.	72	58	72	(60)	(40)	58	(2)	Leader.
5.	60	60	(61)	60	60	60	9	Person.
6.	60	60	(61)	60	60	60	9	Individual.
7.	40	40	40	40	40	(58)	9	Artist.
8.	40	40	62	62	40	62	(6)	Homemaker.
9.	12	12	12	12	12	12	10	Mrs. Edward E. Rockelt.
							73	

8.1 = mean score of protocol

ADJUSTED CATEGORIES UNDER HYPOTHESIZED CONDITIONS:

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1.	48	48	48	48	48	48	10	Christian.
2.	22	22	22	22	22	22	10	Wife.
3.	20	20	20	20	20	20	10	Mother.
4.	72	58	72	(60)	(40)	58	(2)	Leader.
5.	60	60	60	60	60	60	10	Person.
6.	60	60	60	60	60	60	10	Individual.
7.	40	40	40	40	40	(58)	9	Artist.
8.	40	40	62	62	40	62	(6)	Homemaker.
9.	12	12	12	12	12	12	10	Mrs. Edward E. Rockelt.
							77	

8.6 = mean score of protocol

Note that in response 5. and 6. the qualification effect may be attributed to one coder mistakenly identifying the response as qualified or modified. In response 1. the effect is due to a disagreement between coders as to whether the response is a "general religiosity reference" (category 48) or a "reference to religion, religious organizations" (category 52). The disagreement probably hinges upon the capitalization of the word "Christian." See page 41 for a second example.

The following discrete code categories were lumped together and considered to be equivalent categories for rescoring under the hypothetical condition of overlooking qualification or evaluation of the response. In all cases qualification and nonqualification categories have been lumped together. In addition, general and specific categories have also been lumped together for the subjects of "occupation-work" (categories 40,41,42,43) and "religion-religious" (categories 48,52,53).

(10,11) (14,15) (16,17) (20,21) (22,23) (24,25) (26,27) (30,31) (40,41,42, 43) (44,45) (46,47) (48,52,53) (50,51) (54,55) (56,57) (58,59) (60,61) (62,63) (70,71)

#NR = No Coding Effects Removed, QR = Qualification Effect Removed.

+No score notation, in QR denotes no change in score under this condition.

#NR	#QR	
8.4	9.2	Mean Score
9	+	I'm dependable.
10		My work is of excellent quality.
6	10	I have a good business.
7		I make a good living for my family.
		My family is happy.
8.7	9.0	
10		Mother.
10		Housewife.
8		Constant driver.
7	8	Help husband in newspaper work.
9	10	Good Christian, I hope, I try.
8		
9		
8.7	9.0	
10		Trucking contractor.
10		Interest's in family.
8	10	Father of married children.
8		Trying to do and make a success.
9		Not fulfilling ambition as liked.
7		Too much gone from income before personal use.
6.4	7.0	
7		I am supporter of my family.
10		A good father.
10		A good husband.
2	5	Give them an education.
3		A good helper.
8.0	10.0	
9	10	I am a religious leader.
9	10	Being a parent.
6	10	Maintaining a home.

NR	QR	
9.2	9.8	Mean Score
8	10	I am a usher of the church.
10		I am a grandmother.
10		I am a wife.
10		I am a day hand.
8.4	8.8	
9		I'm a mother.
10		Good person.
8	10	Good neighbor.
10		Interior decorator.
5		Beautifier.
8.4	8.6	
10		I am a husband and a father.
5	6	I am a provider of a family.
10		A citizen of a free country.
7		No.
9.3	9.5	
10		I'm a father.
10		I'm a husband.
10		I'm a religious man.
9	10	I'm an office worker.
8		I'm a helper to other people - I try to get them out of
9		trouble.
8.0	8.8	
9	10	I'm employed and I hate to have to work.
10	10	
10		I'm a good church member.
1	5	I sing in the choir.
10		I get along fine with everyone.
8.1	8.6	
8	10	Christian.
10		Wife.
10		Mother.
2		Leader.
9	10	Person.
9	10	Individual.
9		Artist.
6		Homemaker.
10		Mrs. Edward E. Rockelt.
8.3	9.0	
8	10	I am a healthy person.
10		I am a Negro.
7		I am an honest person.

#NR	#QR	
9.5	10.0	Mean Score
8	10	I am a Christian.
10	10	I am an honest Person.
10		I am a church goer.
10		I am a Negro.
8.2	8.5	
10		I am a husband.
9	10	An officer in the Navy.
9		That would about do it.
5		No.
9.0	9.8	
10		I'm Mrs. David Lambiotte.
6	10	I am Marcias Mother.
10		I am a woman.
10		I am an American.
9		I guess that is about all of me there is.
9.0	9.7	
9		I'm just an ordinary guy.
10	10	I'm an American, a Hoosier.
10	10	
9	10	I'm a good barber & I'm not conceited.
10	10	
10	10	I'm a Quaker, belong to Friends Church.
9	10	
4	8	I'm a family man & a child of God.
10	10	
9.1	9.6	
10		That I am a good mother.
9	10	That I keep my home up.
10		I have raised 5 fine children.
9		I am a happy person.
10		My health is good.
9	10	I am a grandmother.
7	8	I am a person that has the love of her family.
8.0	9.3	
6	10	I am just a retired person.
9		Ive lived my life.
9		That's all.
9.7	9.8	
10		I am an individual.
10		I am a housewife.
10		I am very optimistic
10		I am a pretty happy-go-lucky person.
9	9	I am a foreigner.
10		I am not too hard to get along with.

NR	QR	
9.8	10.0	Mean Score
10		I'm an average male.
9	10	Content being the head of the family.
10		I'm satisfied in type work I'm doing.
10		I'm a good husband.
10		I'm a good father.
9.0	10.0	
10		I am a housewife.
9	10	I am a mother of five children.
8	10	I am a Christian.
8.2	9.2	
10		I am free to do whatever I want.
9		I am happy.
8		We are financially solvent.
6	10	I am engaged to be married.
8.0	9.0	
8	10	Breadwinner for 3 people.
8	10	I am a middle income bracket person.
9		That's all.
7		None whatsoever.
8.4	9.6	
4	9	I am a retired professional man.
9		I was successful.
10		I am an individual.
9	10	I am a citizen of the United States.
10		I am a deep thinker.
6.7	7.3	
2	3	My name, of course.
9	10	My health.
9		Love.
7.3	8.2	
10		Tried to live right.
9		Nice.
9	10	Lived a Christian Life.
10		I am a mother.
1		I live alone.
5	9	Do my own housework.
9.8	10.0	
10		I have been a good mother.
10		I have worked hard.
10		I try to help others.
10		Try to Family (loyal).
9	10	To my Church loyal.

NR	QR	
7.5	9.8	Mean Score
2	9	I am a Church Deacon.
8	10	I raised 5 boys and 4 girls.
10		I can still do a good days work.
10		I goes to Church every Sunday.
9.3	10.0	
8	10	I am a father of 6 children.
10		I am a husband.
10		I am a plumber.
9.1	9.6	
10		A child of God.
9	10	A citizen of U.S.
7		A person with opportunities unequaled except in U.S.
10		An average American.
7	10	An average job.
10		An average family.
10		I live by the sweat of my brow.
10		I enjoy making a living.
7.5	8.0	
6	10	I work for a living.
9		I get along with others.
10		I am pretty well satisfied.
10		I am old.
9		I've raised 2 boys.
2		I'm raising another.
4		Use to live on a farm.
10		Got my education the hard way.
9.0	9.3	
10		Fathers.
9		Dishwasher.
9		Lawnmower.
10		Grandfather.
9		mason.
7	9	Good neighbor.
8.2	9.1	
10		I am a mother.
10		I am a housewife.
9		I have a lot of responsibility.
10		I don't care for sports.
5	8	I work too hard in house.
5	9	I don't work in yard.
10		I'm not lazy.
7		I love to sleep.

NR	QR	
9.3	9.5	Mean Score
9	10	I'm trying to live a Christian life.
9		I'm honest.
10		I'm a mother.
10		Housewife.
9		I work outside of home.
9		I'm not lazy.
8.8	9.0	
10		I am a mother.
10		I am a wife.
10		I am a church worker.
9		I am active in community.
5	6	P.T.A.
8.6	9.8	
10		I'm a human being.
9	10	I'm able to work.
5	9	I'm able to carry my home.
9	10	I am proud to have children.
10		I am a citizen.
9.2	10.0	
10		I am an individual.
10		An average person.
6	10	Try to be a Good Neighbor.
10		Cheerful person.
10		Cooperative Person.
9	10	Good wife.
8.0	8.5	
7		I am a helper of people.
8	10	I am head of a family.
10		I am a church member.
7		None.
7.0	8.2	
10		My name is Viola Gustafson.
7	9	I am raising 2 children.
4	7	I must make sure the children have enough to eat and
7		clothes on their back.
7.7	8.7	
5	8	Im an uncertain person.
10		Im a good mother.
8		But a very poor housekeeper.

NR	QR	
8.2	9.5	Mean Score
10		I guess I just live a clean life.
6	10	I like my neighbors.
9	10	And my family.
8		I like my car.
9.5	9.8	
10		I am a good husband.
9		A good worker.
10		Agreeable.
10		Not quarrelsome.
10		Not in good health.
8	10	Hope a good neighbor.
9.4	9.6	
9		I am a very contented person.
10		I am fortunate to have such a wonderful family.
9	10	I am not very tall.
9		I enjoy being myself.
10		I wouldn't want to change.
9.8	10.0	
10		Glad to be a citizen of the U.S.
10		I'm a carpenter.
9	10	Happy to have opportunity to have a home & work.
8.6	8.8	
10		I'm a woman.
10		I'm a housewife.
10		I'm a mother.
10		I'm an individual.
4		I am part of a group.
9	10	I am a sister.
9	10	I am a daughter.
7		I am a consumer.
8		I am a homemaker.
9		I am a cook.
8.5	9.1	
6	10	I've kept my home together.
10		I've kept my job.
9	10	I've kept my children together.
10		I've kept my family as a family.
10		I've always kept my temper.
8		I was never frivolous with a dollar.
5		I gave my boy the best education you could get.
10		I have a good wife.

NR	QR	
8.0	8.8	Mean Score
10		Good mother.
10		Good wife.
5	9	Overseeing the home.
5		Good living women - all my life I've tried to do what is right.
8.7	9.2	
10		A wife.
10		A mother.
10		A teacher.
9	10	Community leader.
7	9	Sportswoman.
6		Poor housekeeper.
9.2	9.8	
9		Idea of God.
9	10	Wife.
10		Painter.
9	10	Lover of Animals.
9	10	Lover of flowers.
9.0	10.0	
10		Mother.
10		Worker.
10		Chinese.
6	10	Going to be a Grandmother.
8.8	9.0	
10		I'm a worker.
10		I'm an American.
9	10	I'm a music lover.
10		I'm a husband.
5		No.
9.4	9.8	
10		A human being.
10		I try to be a nice person.
10		Generous to a fault.
10		Helpful.
7	9	There's nothing more important to me than whats between me and my husband.
6.7	7.0	
4	5	I live all alone.
8		I am an old man.
8		

Section III

Protocols Scoring Differently When Accidental Effect* of Splitting is Removed N = 17

*Accidental effect of splitting should not be confused with the main effect of splitting. If we were treating the main effect here, we would arbitrarily remove or add code categories to the real judgements of non-conforming coders. Thus, if four of six coders had split a verbal response into two or more codable responses, we would arbitrarily add some sort of code categories to the judgements of the two nonconforming coders. Likewise, if four of six coders had not split a verbal response into two or more codable responses, we would arbitrarily subtract code categories from the judgements of the two non-conforming coders. Such a procedure was not feasible, for at least three reasons; (1) often the disagreement on splitting is not modal, but rather an even split, artifact of an even number of coders; (2) There is little basis for deciding which of two code categories should be removed from a coders judgement, unless it is done on the basis of conforming to a modal category; and (3) There is even less basis for deciding which of many code categories should be added to a coders judgement, unless it is done on the basis of conforming to the single category already assigned to the verbal response, as when a single code of "12" is changed to two codes of the category "12" by the same coder.

Therefore, removal of the accidental effect of splitting does not necessitate adding or subtracting or changing the categories of different codes. The latter method was used when the qualification effect was removed in Appendix C, Section II. What is involved here is a rearrangement of the existing categories in the code matrix, so that reliability may be gauged upon a comparison of the categories assigned by all coders to the same response. We will, in effect, be straightening out the response rows across the code matrix by leaving some blank gaps within the code matrix where no split was made by five or fewer coders, and, therefore, no category was assigned by five or fewer coders. Following is an example of what happens. (see page 43 for another example)

ASSIGNED CATEGORY PLACEMENT UNDER REAL CONDITIONS:
(Response Number not used to avoid confusion)

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
	11	17	17	17	11	17	8	I am an old woman.
	16	25	25	25	16	25	8	I've got lots of friends.
	25	72	72	72	25	59	5	I like people.
	72	72	72	72	72	72	10	D.K. I'm awfully dumb.
	72		72	72	72		8	
							39	

7.8 = mean score of protocol

ADJUSTED CATEGORY PLACEMENT UNDER HYPOTHETICAL CONDITIONS:
(Response number refers to verbal response)

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
1	(11)				(11)		8	I am an old woman.
	(16)	17	17	17	(16)	17	8	
2	25	25	25	25	25	25	10	I've got lots of friends.
3	72	72	72	72	72	(59)	9	I like people.
4	72	72	72	72	72	72	10	D.K. I'm awfully dumb.
			(72)	(72)			8	
							8.8	= mean score of protocol

This example is particularly helpful in showing how the rescoring procedure to test this hypothesis can create a larger number of scores for the protocol. Here we see that the N of scores increases from five to six. The reason for this is that adjustment within the code matrix to straighten response rows necessitated expansion of the matrix. To be exact, two coders, (A and E) did not split any responses, two coders (G and D) split only the first response, and the remaining two coders (B and C) split the last response only. If any of the six coders had, under the realistic condition, split both the first and last response, there would have been six codes for the realistic condition, and therefore, no difference in the N of scores under the two differing conditions.
+NR = No coding effect removed, SR = Accidental effect of splitting removed, R# = response number of verbal response. For the NR scores it is not useful to match verbal response to scores, by number.
% = No score notation denotes no change in score.

R#+	NR+	SR+	
	8.3	9.0	Mean Score
1	10	%	I'm a good mother.
2	3	8	That I am able to hold a position at my age.
	9	5	
3	9	10	I am an efficient worker.
4	9	10	I am a grandmother.
5	9	10	I am a religious person.
6	9	10	I am a good citizen.
	7.2	8.6	
1	9		Let's pass that up. I'm not very cooperative.
	7	8	
2	7	8	Just a teacher.
3	5	9	The community thinks we're just a servant.
4	8	9	So you can write your own criticism.

R#	NR	SR	
	7.8	8.8	Mean Score
1	8		I am an old woman.
	8		
2	5	10	I've got lots of friends.
3	10	9	I like people.
4	8	10	D.K. I'm awfully dumb.
		8	
	9.0	9.3	
1	10		A good mother.
2	10		Important to my husband.
3	8		A good housekeeper, I hope.
	9		
4	8	10	A good Church member.
5	9		Important to my community.
	7.9	8.5	
1	10		I am a good daughter.
	7	9	
2	9	7	I am a reasonably successful secretary.
3	6	10	I am a good and loyal friend.
	7	6	
4	7	9	I could be more active in community life,
	9	8	but I have a busy schedule.
5	-	9	I am a little cog in the wheel.
	8.3	8.5	
1	10		I'm very happy person.
2	9		I'm healthy.
3	10		I'm young.
4	10		I have high hopes.
5	4	7	Hope to have a healthy & happy family.
	8	7	
6	6	10	Very lucky to have a wonderful husband.
7	9		Wonderful mother & father in law.
	8	5	
8	8	10	Wonderful mother and step-father.
	9	6	
	9.3	9.5	
1	9		I am a widow.
2	10		I have a great burden in life.
3	9		I have always sacrificed my life for my family.
4	10		I have raised my own children and now must raise
	9		grandchildren.
5	9	10	I have lived a clean life above reproach.
	9.8	9.8	
1	10		I'm a woman.
2	10		I'm a mother.
3	10		I hope I'm useful. I think I am useful.
	10	9	
4	10		I'm needed by others.
5	9	10	I help - or try to help others.

R#	NR	SR	
	6.5	8.8	Mean Score
1	10		I'm an American citizen.
	8		
2	10		I believe in fairness.
3	9		I try to deal fair with everybody.
4	2	7	I can get credit anywhere I've been very fair.
	3	9	
5	2	7	I'm just a common-ordinary fellow.
6	8	10	I never had more than a country school education.
	7.0	7.5	
1	9		I like football, basketball.
	9	7	
2	7	8	I like to drink, smoke.
	7		
3	9	10	I am a father & mother.
	9	10	
4	4	5	I like women.
5	6		I like to go to parties.
6	5		I don't like onions.
7	5	8	I don't like the Chicago Bears & L.A. Dodgers.
	6.8	8.8	
1	10		I am an American citizen.
	6		
2	5	10	I am a Husband.
3	6	9	I am unemployed.
	8.7	9.0	
1	10		Father.
2	10		Citizen, a damn good one too.
	9		
3	9	10	Good Blacksmith.
4	5	9	Good husband.
5	9	6	Good gardner.
	8.8	9.5	
1	10		A mother.
2	10		I am a middle class citizen.
	10	8	
3	8	10	I'm a Negro.
4	7	10	I'm a nurse.
5	8	9	I'm concerned about my position.
	9.3	9.7	
1	9		Very small ingredient of a vast mixture.
2	10		That Speck ingredient is as small as a grain of salt
3	10	9	Of itself in itself its insignificient. (sic)
4	9	10	But in the whole matter it can influence the taste of the finished product.
5	9	10	Important to my family.
6	9	10	Important to my country.

R#	NR	SR	
	7.1	7.7	Mean Score
1	9		I'm a good human being.
2	10		I enjoy life.
3	6	5	I appreciate many things.
4	9		Art,
5	5	7	Music,
6	6	8	People and,
7	5	6	Nature.
	8.0	8.0	
1	9		I'm not what I'd like to be.
2	10		I'm not as intelligent as I'd like to be.
3	7		I don't persue educating myself to the degree I feel I should.
4	5		I neglect myself physically - in appearance.
5	6		I think of myself as a person with tremendous am't of love.
6	9		Emotional.
7	9		Concern for others & the world esp. children.
	9		
	9		
8	5	9	Someone who can't say no if it hurts others.
9	9	5	Sometimes I think I have a lot of knowledge
	9	8	& sometimes I think I'm a phoney.
	8.6	8.4	
1	10		I'm a relatively unimportant person who likes
	9	7	to do things for others.
2	10		I think it's important to watch out for others as well as ourselves.
3	7	10	I have a good sence (sic) of humor and that's
	7	5	important.

Section IV

Protocols Scoring Differently When Accidental Effect* of Splitting is Removed And When Qualification Effect* is Removed

N = 37

---

*See Appendix C, Section III for an explanation of the rescoring process in removing accidental effect of splitting. See Appendix C, Section II for an explanation of the rescoring process in removing qualification effect. Following are examples which should clarify the rescoring done in this section:

ASSIGNED CATEGORY UNDER REAL CONDITIONS:
(Response number is meaningless)

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
	(61)	72	72	72	(16)	72	7	I am a guy with a lot of problems.
	43	43	43	43	(72)	43	9	I am concerned about finding a good job.
	47	72	47	72	43	43	(3)	I have reasonable experience but do not
	41	46	41	46	(43)	(47)	(2)	have a college degree.
	22	40	22	40	(46)	(41)	(2)	I am ex claim insurance adjuster.
	22	22	23	22	40	22	7	I am single & I wish I was married.
	(72)	23	(72)	23	(22)	23	5	I like people.
	(43)	72	72	72	(22)	72	7	Enjoy working with people.
	43	(72)	43	43	(61)	43	7	I like work that is worth while.
	72	72	72	43	43	43	(6)	I like to do something that is
		72		(42)	(43)	72	(2)	beneficial to other people.
					(43)		9	
							66	

5.5 = mean score of protocol.

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
	50	50	50	50	50	50	10	I am an American citizen.
	50	50	72	50	72	72	(6)	I am honest.
	72	72	72	72	72	72	10	I am good.
	72	72	(43)	72	(43)	(42)	5	I work every day.
	(42)	43	(63)	43	43	(72)	4	I stay home on the week ends.
	(62)	(59)	53	(73)	(52)	53	1	I go to church on Sunday.
	(52)	(53)		(53)			5	
							41	

5.9 = mean score of protocol.

ADJUSTED CATEGORY UNDER HYPOTHESIZED CONDITION:
REMOVAL OF QUALIFICATION EFFECT
(Response number meaningless)

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
	(61)	72	72	72	(16)	72	7	I am a guy with a lot of problems.
	43	43	43	43	(72)	43	9	I am concerned about finding a good job.
	47	72	47	72	43	43	(3)	I have reasonable experience but do not
	(41)	46	(41)	46	(43)	46	5	have a college degree.
	(22)	40	(22)	40	(46)	40	5	I am ex claim insurance adjuster.
	22	22	22	22	(40)	22	9	I am single & I wish I was married.
	(72)	23	(72)	23	23	23	8	I like people.
	(43)	72	72	72	(22)	72	7	Enjoy working with people.
	43	(72)	43	43	(61)	43	7	I like work that is worth while.
	72	72	72	43	43	43	(6)	I like to do something that is
		72		43	43	72	(3)	beneficial to other people
					(43)		9	
							76	

6.5 = mean score of protocol

Response Number	Coder G	A	B	C	D	E	Verbal Score	Response:
	50	50	50	50	50	50	10	I am an American citizen.
	50	50	72	50	72	72	(6)	I am honest.
	72	72	72	72	72	72	10	I am good.
	72	72	43	72	43	43	(6)	I work every day.
	43	43	(63)	43	43	(72)	7	I stay home on the week ends.
	(62)	(59)	53	(73)	53	53	4	I go to church on Sunday.
	53	53		53			(6)	
							49	

7.0 = mean score of protocol

When qualification effect is removed, there seems to be disagreement between coders on the first protocol whether (1) "do not" in response three is a qualifier, (2) "ex" in response four is a qualifier, (3) "wish" in response five is a qualifier, (4) "worth while" in response eight is a qualifier. In the second protocol there seems to be disagreement between coders on whether (1) "every day" in response four is a qualifier, and (2) "on Sunday" in response six is a qualifier.

ADJUSTED CATEGORY PLACEMENT UNDER HYPOTHESIZED CONDITION:
REMOVAL OF ACCIDENTAL EFFECT OF SPLITTING
(Response number refers to verbal response)

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1.					16		9	I am a guy with a lot of problems.
	61	72	72	72	72	72	9	
2	43	43	43	43	43	43	10	I am concerned about finding a good job.
3.		72		72	43	43	3	I have reasonable experience but do not
	47	46	47	46	46	47	6	Have college degree.
4	41	40	41	40	40	41	6	I am ex claim insurance adjuster.
5	22	22	22	22	22	22	10	I am single & I wish I was married.
	22	23	23	23	22	23.	8	
6	72	72	72	72	61	72	9	I like people.
7	43	72	72	43	43	43	8	Enjoy working with people.
8	43	72	43	43	43	43	9	I like work that is worth while.
9.	72	72	72	42	43	72	7	I like to do something that is
							94	beneficial to other people
								7.8 = mean score of protocol

Response Number	Coder G	A	B	C	D	E	Score	Verbal Response:
1.	50	50	50	50	50	50	10	I am an American citizen.
	50	50		50			6	
2	72	72	72	72	72	72	10	I am honest.
3	72	72	72	72	72	72	10	I am good.
4	42	43	43	43	43	42	8	I work every day.
5	62	59	63	73	43	72	0	I stay home on the week ends.
6.	52	53	53	53	52	53	8	I go to church on Sunday.
							52	
								7.4 = mean score of protocol

When accidental effect of splitting is removed, we can see that there is, in the first protocol, disagreement between coders on whether to split (1) "I am a guy with a lot of problems" into the two responses, "I am a guy, and I have a lot of problems", and (2) "I have reasonable experience but do not have college degree" into two clauses; but there is no disagreement that "I am single & I wish I was married." should be split into two clauses. In the second protocol there is only disagreement on whether to split the response "I am an American citizen." into the two responses, "I am an American, and I am a citizen." However, this disagreement is so near the beginning of the protocol that it creates a good deal of accidental effect.

In the following material, NR = no coding effects removed, QR = qualification effect removed, SR = accidental effect of splitting removed. Notice that none of these results amounts to the removal of both qualification effect and accidental effect of splitting in the same operation. R# = response number, which refers to the verbal response and can be meaningfully linked to the score only in the condition in which accidental effect of splitting is removed (SR).

+When there is no score notation in column QR or SR, it denotes that no change in score occurred due to the rescoring.

R#	NR	QR	SR	
	6.3	7.2	7.7	Mean Score
1	10	+	+	I am dependable and steady.
	9			
2	5	9	8	I'm a family man.
3	7		9	I try to do the best for my community and country.
	5		4	
4	7		10	I do my job as well as I can.
5	2	5	5	I want my wife and children to respect--
	4	5	5	
6	8		9	one so I do try to have high standards of behavior.
	4.8	6.2	5.2	
1	5			I have a lot of duties to perform.
2	5			I like social life - lodge work.
	1	4		
3	5	9		My family.
4	8		9	Nothing else.
	5.5	6.5	7.9	
1	7		9	I am a guy with a lot of problems.
	9			
2	3		10	I am concerned about finding a good job.
3	2	5	3	I have reasonable experience but do not have
	2	5	6	college degree.
4	7	9	6	I am ex claim insurance adjuster.
5	5	8	10	I am single & I wish I was married.
	7		8	
6	7		9	I like people.
7	6		8	Enjoy working with people.
8	2	3	9	I like work that is worth while.
9	9		8	I like to do something that is beneficial to other people.
	5.9	7.0	7.4	
1	10			I am an American citizen.
	6			
2	10			I am honest.
3	5	6	10	I am good.
4	4	7	8	I work every day.
5	1	4	0	I stay home on the week ends.
6	5	6	8	I go to church on Sunday.
	5.0	6.6	6.6	
1	9			I am most important to my family.
2	8	10		I hold a place in my church, I feel important.
	4	5	5	
3	2	7	6	I love my neighbors.
4	2		5	I'm too much fat.

R#	NR	QR	SR	
	6.9	7.7	7.1	Mean Score
1	8			Outside of just an ordinary home maker.
2	10			Mother.
3	9			And citizen.
4	2	5		I try to work in school affairs.
5	6	10		And a few fraternal organizations.
6	9			Well, when I was working in the biz (sic) world I felt--
7	5		6	I should try to be sincere & make myself valuable--
	4		9	
8	9		5	& not loaf on the job.
	7.7	8.6	8.3	
1	9			I am a chauffer, bookkeeper (sic), laundress, cook, housekeeper.
	9			
	9			
	9			
	9			
2	5	9	6	I am just a wife and mother.
	5	9	6	
3	7		9	Don't know any more answers.
4	7		9	Just don't know.
	6.8	7.5	8.5	
1	9	10		A human being as everyone.
2	6			I am an honest man who never went to jail.
	7			
3	6		10	I know how to work.
4	2	7	8	I am not lazy.
5	8			I like my neighbors.
6	8		10	I think I am a friend to everyone.
7	8		10	I like to work.
	8.3	9.0	8.7	
1	5	8		I am a good provider for my family.
	10		9	
2	7	9	10	I am a father.
3	9		8	I am a christian (sic).
4	9			I like people.
5	9		10	I am a Negro.
6	9		10	I am a man.
	7.5	7.8	7.0	
1	5	9		I'm just me.
2	8		5	Absolutely.
3	8		9	Refused - even--
4	9			after probing.

R#	NR	QR	SR	
	6.3	6.8	6.8	Mean Score
1	10			Your attitude.
2	10			The way I carry myself thru life, my job, etc.
	7	9		
	7		9	
3	1	4	10	I love my family & want to provide for them
	7		1	any way I can.
4	1		10	I'm a Negro and proud of it because its Gods (sic)
	2		4	intention.
	9		2	
5	9		5	Refusal.
	4.5	5.2	8.2	
1	6	10		I am an individual common everyday person.
	6		10	
2	2		9	Try to live up to the laws of America & pay
	2		4	my bills.
3	5		10	Try to bring up my family & get as meddui (sic)
	6		10	as you I can.
	8.7	9.0	9.3	
1	8	10	9	I am a mother & happy to be one.
	9			
2	9		10	I like people.
3	8		9	D.K. -- I am a wife.
	9		9	
4	9		10	I can't think of anything more.
	6.5	7.0	8.3	
1	9			I'm a helper to my family.
	9			
2	2	5	10	I like people.
3	5			I'm a working mother.
	5		8	
4	9			I'm really not much.
	5.6	6.2	7.5	
1	8			I'm just a poor working gal.
	8			
	4	7	9	
2	1	4	9	I'm just an ordinary person.
3	4		9	I go to work, come home attend church.
	1		4	
	4		10	
4	8		4	I like to travel.
5	9		8	I enjoy reading when I have the time.
	9		6	

R#	NR	QR	SR	
	5.2	6.8	7.0	Mean Score
1	8	10		First of all a mother, and still needed by
	2	5		family & husband.
	3	8	5	
2	5		10	I'm a companion to my husband.
3	5		8	I try to be a good homemaker.
4	8		9	Nothing else.
	5.6	6.6	7.9	
1	7	9		If I was well I would like to work.
	7	9	8	
2	1	4	9	I'm retired was 2 weeks in hospital have
	4		7	ulserated (sic) foot.
	7		9	
3	6		9	If nobody bothers me I'm O.K.
4	7		9	I don't like to nag or fight.
	-		5	
	7.8	8.4	8.6	
1	10			I'm a farmer.
2	10			A father.
3	9	10		I'm a boater - sailor when I have time.
	1	4	10	
4	5	9	4	I'm a director of many organizations.
5	7		8	I'm a bridge player.
6	9			But I work late most nights.
	7.2	7.4	8.2	
1	10			I have 29 grandchildren so a Grandmother.
	10		8	
2	6		10	And mother.
3	2		9	A wife.
4	9		5	And the official telephone answerer here.
5	8		10	I'm quite active in church.
6	4	5	7	And I love music.
7	9		7	And sing in the choir.
	8.0	8.3	8.7	
1	9			I'm very congenial I think.
2	9			I'm understanding.
3	7			I'm very unselfish I think family comes before me.
	7	9	9	
4	7		9	I always try to look out for my fellow man.
5	10		9	I like to apply the Golden Rule.
6	7		9	I try to do things other people ask me to do.

R#	NR	QR	SR	
	6.1	6.3	6.4	Mean Score
1	1	2	9	I am a person who would like to become head
	1			of a large business firm.
2	7		8	I am a person who believes a happy marriage
	7		1	is the basis for success.
3	9		8	I am a person who believes one should have a
	9			Supreme Being in which to believe.
4	9			I like to enjoy life.
	9.1	9.3	9.3	
1	9	10		I am a citizen of the United States.
2	9			I am a good understanding honest woman.
	9			
	10		9	
	9			
3	9		10	I am religious.
4	9		10	I am respectable.
	6.8	7.2	7.5	
1	5	6		An old woman.
	5	6	8	
2	9		10	I live alone.
3	8		7	I like to be alone.
	7.2	7.7	8.5	
1	10			I'm a believer of God.
2	5	9		A wife that loves her family.
	8			
3	5		9	I give untireing (sic) of my service to others.
4	7		10	Good American.
5	8		9	Don't know.
	4.8	5.6	7.7	
1	6		5	I am responsible for my family's health & well-
	6		8	being.
2	6		10	I am a mother.
3	5		10	I am a wife.
4	1	4	5	I am able to sew and knit for my family.
	2		9	
5	4		8	I am said to be a good shopper as I know prices.
	8		5	
6	-		7	I participate in sports with other women.
	-		10	
	9.0	9.3	9.5	
1	10			I am a man.
2	8	10		A Christian man.
	9			
3	9		10	A husband.
4	9		10	A farmer.
5	9		10	Father.

R#	NR	QR	SR	
	8.2	8.4	8.2	Mean Score
1	10			I am a happy mother.
2	10			I am a good wife.
3	10			I am helpful to other people.
4	10			I am an eager student.
5	10			I am an eager outdoor sports woman.
	7	8		
6	7			I am a good house-keeper.
7	2	6	5	I enjoy gardening.
8	8	5	2	I am a good needleworker, sewing and knitting and
	7		8	crocheting.
	9			
9	9		10	I enjoy the arts, (painting in oils).
	7.7	8.0	8.7	
1	9			Old every day worker.
	9			
2	10			I have a family.
3	7	9	10	Father and Husband.
	2		9	
4	9		5	No.
	8.0	8.6	8.6	
1	9			I am nobody.
2	9			I intend to be somebody.
3	9			I would like to make other people happy.
4	9			I am a person of responsibility.
5	7	9		I would like to have a large family about
	7		9	12 kids.
6	9			I am sure of myself.
7	9		10	I am confident with respect to work.
8	7		10	I have respect from my friends.
9	5	9	9	I feel a little unsettled as far as my family
				is concerned.
10	-		5	I have a friendly relationship with neighbors.
	8.5	9.0	8.5	
1	8	10		I am another individual.
2	10			I have certain rights and opportunities.
	10		6	
3	6		10	I have been fortunate in life.
	8.5	9.0	9.2	
1	10			I'm just me - a person.
	5	8	6	
2	10		9	I'm satisfied with my way of life.
3	10			I get along ok.
4	10			D.K.
5	6		10	D.K.

R#	NR	QR	SR	
	7.1	7.7	7.7	Mean Score
1	7	8		I am the man of the house.
2	10			I made a change for better instead of worse.
3	5	6		I try to provide for my family.
4	7	9		I like to help my neighbors & friends.
	9			
5	5		8	I like to help myself last.
6	7		8	I am a Christian.
	6.4	7.7	7.2	
1	10			I am a housewife.
2	10			I am a mother.
3	8	10		I am a part-time worker.
4	2	9		I am a baby sitter sometimes for friends or relatives.
	7			
	1		9	
5	7		10	I am married.
6	-		2	N. F. I.
	7.9	8.4	8.6	
1	10			I'm Bernard Perry.
2	8	10		I'm Father of 5 children.
3	9	10		I'm a married man.
	4	5	9	
4	5		7	I'm average fellow.
5	9		8	Right now I've got a h--- of a cold.
6	9			Had Pneumonia last year &--
7	9			I'm afraid of this cough getting to be Pneumonia again.
	6.7	6.8	7.5	
1	10			Honesty, Truthful, Sincere.
	9			
	7		8	
2	5		8	Love for home & country.
	5		9	
3	4	5	1	My occupation.
	8.6	9.0	8.7	
1	10			I try to be a good mother.
2	10			I feel I am a good wife.
3	9			Fianancially (sic), I like to feel I help my husband.
4	7			I love helping other children. We ran a--
5	5	8		foster home for 5 years in New York State.
6	10			I believe when I work and am hired to work I-- do the best of my ability. In other words, I--
7	9		10	don't shirk my job.