CAMBRIDGE
UNIVERSITY PRESS

Exploring Elite Political Attitudes: Some Methodological Lessons

Author(s): Joel D. Aberbach, James D. Chesney and Bert A. Rockman

Source: *Political Methodology* , 1975, Vol. 2, No. 1 (1975), pp. 1-27

Published by: Cambridge University Press on behalf of the Society for Political Methodology

Stable URL: https://www.jstor.org/stable/25791403

Exploring Elite Political Attitudes: Some Methodological Lessons

by Joel D. Aberbach, James D. Chesney
and Bert A. Rockman

Studying elite political attitudes, values, and beliefs is a difficult undertaking. This article reports on the methods we used to examine the political thinking of American administrators and congressmen and discusses the implications of these techniques. Before turning to this subject, however, we should say a bit more about the context, assumptions and content of our work.

The material discussed here is drawn from the American segment of a comparative study of bureaucratic and political elites in nine countries.[1] Interviews were conducted in 1970 with 126 administrators from eighteen Federal agencies within the Washington, D.C. metropolitan area. Officials were stratified by position and samples were drawn of those formally designated as "political" executives and of career civil servants under civil service jurisdiction with supergrade status (GS 16-18). The response rate was 87.5 percent, i.e., we interviewed 126 of the 144 administrators contacted. In 1971 we conducted interviews with members of the U.S. Congress. Freshmen congressmen were excluded from our sampling universe and the sample was stratified by age. Seventy-seven congressional interviews were completed, a 70 percent response.

Our basic assumption is that the political attitudes, values and beliefs of bureaucrats and politicians are important determinants of the ways in which governments respond to social change and to the pressures brought to bear on them by groups in society. A study of these attitudes can tell us something about the nature of relationships between members of the governmental elite, about certain aspects of the decision-making process, about how elites analyze policy problems, and about the preferences, hopes and plans of those in key positions. We recognize, of course, that without knowledge of the "environmental situations in which actors find themselves, (knowledge of) the psychological predispositions they bring to those situations" (Greenstein, 1969:7) is not enough to predict discrete behaviors, especially choices of particular policies. Since our research stresses the

1

views of a cross-section of bureaucrats and politicians, our
familiarity with the environmental factors affecting behavior
is rather sketchy, at best, i.e., our knowledge of particular
factors and situations is limited.[2] We see elite political
beliefs as important parameters in the behavioral equation--
setting limits, defining the legitimate and the illegitimate,
directing inquiry and thought, influencing the interpretation
of events, guiding the definition of problems and the re-
sponse to them.

 The interview covered the respondents' social back-
grounds and career patterns in great detail, but the major
emphasis was subjective. We assessed administrators' and
congressmen's perceptions of how the policy-making process
operates and their evaluations of their own and other actors'
roles in the policy process. We especially wanted to obtain
an understanding of how administrators and elected leaders
view one another; whether the interactions among them occur
in an atmosphere of trust and cooperation or of hostility and
suspicion. We were interested in elite responses to issues
relating to political representation, interest articulation,
and citizen involvement. Special attention was given to the
actual and proper roles of constituencies in policy making
and the necessity, prospects and problems of increasing pop-
ular control over the activities of government. Respondents
were asked to discuss the character of the nation's political
parties, the proper role of the government in social and eco-
nomic affairs and, more generally, the nature of conflict and
consensus in politics and society. Finally, we were inter-
ested in respondents' aspirations for the country's future
and what their notions were about how best to achieve their
goals for the future.

 While the substantive findings of our research efforts
promise to be extremely rich, the purpose of this article is
to describe some of the interesting issues we faced in devel-
oping our findings. We hope to illustrate the gains and
losses in gathering and interpreting data that flow from the
use of open-ended elite interviews. The choice of instru-
ments and procedures applied in collecting data are the con-
sequence of prior theoretical concerns. The instruments also
shape the parameters of later research decisions. Research
decisions, in short, tend to have cumulative consequences.

 This paper covers the following topics: 1) open-ended
interviews: advantages and disadvantages; 2) the question-
aire; 3) the interview situation; 4) gaining access; 5) cod-
ing procedures and methods; and 6) the limits on data anal-
ysis.

II. OPEN-ENDED INTERVIEWS: ADVANTAGES AND DISADVANTAGES

　　　While we employed a standard interview schedule in this study, the questions were open-ended and the discussions were often wide-ranging. We wished to conduct the interviews in a conversational style in order to put the respondent at ease and to elicit thoughtful and, where appropriate, complex answers to our questions. Open-ended procedures emphasize the contextual richness of response and allow for the exploration of subtlety and nuance; they enable an investigator to assess not just the surface content of a response but also the reasoning and premises underlying it. For some, this is a sufficient virtue for employing open-ended materials in studying elite attitudes. In contrasting Robert Lane's (1962) open-ended depth interviews of the "common men" of Eastport with the close-ended interviews of French and German elites conducted by Karl Deutsch and others (1967), Bernard Brown (1969:421) comments that "Professor Lane interviewed bricklayers, and treated their views as if they were opinions of an elite. The Yale Project interviewed elites and treated them like bricklayers."

　　　To some extent, pursuing contextual richness through open-ended, semi-structured interviews precludes the use of statistically powerful analytic techniques on the resulting data. In Section VI we discuss the elaborate coding scheme used to record the political beliefs, attitudes and values of the respondents so that the interviews could be analyzed with precision. But while the open-ended, quasi-conversational technique maximizes the interviewer's ability to clarify, illuminate, or probe more deeply into the responses of the subject, it also means that questions are not asked always in exactly the same way (Dexter, 1970:132-133; see also Madge, 1965:163). The absence of closure limits one's capacity to impose powerful analytic or structuring techniques on the data. Unless responses are carefully coded to minimize information loss, the transformation of highly verbalized information into quantitatively useable data will lead to a sizeable missing data problem since even with a uniform stimulus people often talk about different things or use different frameworks. The missing data problem, of course, plagues multivariate analysis, particularly when there are small samples to begin with, as is almost inevitably the case in interview studies with elites.

　　　Our purpose here is not to celebrate the virtues or declaim the pitfalls of either close-ended or open-ended approaches in the abstract but to clarify the nature of the trade-offs in pursuing one or the other approach. We will discuss some of the reasons for choosing the open-ended

approach and point out, as well, some of the constraints
which this choice places on our data analysis.

In choosing an open-ended approach we were guided by
three considerations. The first of these had to do with the
degree of prior research of the subject of concern. Where a
sizeable body of data has been mined and basic patterns ex-
plored, it is possible to define important questions with
clarity and to develop fairly precise research expectations.
When such questions and expectations can be formulated in
reasonably precise fashion there is every reason to opt for
response-constraining research instruments. In studying
attitudes this is likely to mean an emphasis on standardized,
close-ended instruments providing fairly specific stimuli.
Conversely, the more exploratory one's concerns and the less
precisely stated one's expectations, the greater the need to
choose less constraining interview instruments in order to
increase the ability to capture the richness, subtlety, and
nuance of the responses. Typically, this situation leads to
a research instrument oriented toward flexibility, probing,
and open-endedness. As Roll and Cantril (1972:103) observe
in the context of mass opinion research: "Open questions are
useful when the dimensions of the public's view are not read-
ily apparent."

Similarly, in justifying his use of an open-ended inter-
view format for his study of U.S. Senators, Donald Matthews
(1971:20-21) asserts that at the time he ". . . was also
aware that I still did not know what the relevant and inter-
esting questions about senatorial behavior were; to lock my-
self into a standard interview seemed most unwise under these
circumstances."[3] As Matthews (1971:21) further points out,
it is essential to discover what the broad and interesting
patterns are before specifying precise questions requiring
the application of precise research instruments. Where know-
ledge of these broader contours is ample, we are in a better
position to develop more close-ended questions and to employ
more structured tools.

Our contention is that in the absence of prior descrip-
tive theory, where it is difficult to formulate highly pre-
cise expectations, less tightly structured research proce-
dures allow for greater adaptation, and enable the investi-
gator to apply what he learns while still in the field. As
Bakan (1969:XIV) says:

> . . . the fact of the matter is that good
> research into the unknown cannot be well
> designed in the usual sense of the term.
> Truly good research means that one allows
> the investigation to be guided by the exper-
> iences of the investigation. And this can-
> not be predicted.

In our investigation we wanted to explore the relatively un-
charted views and styles of thought of American administra-
tors and congressmen on a series of rather abstract and com-
plex issues. Emphasizing tight structuring and close-ended
material at this stage of our understanding merely to enhance
analytic elegance at a later stage would not have served our
major purpose.

A second consideration leading to the use of open-ended
materials was the matter of <u>response</u> <u>validity</u>. This is
clearly where the "bricklayer" image cited by Bernard Brown
(1969) becomes pertinent. Inferences about the logical struc-
ture of beliefs are especially risky with close-ended data
(Brown, 1970). This risk may be particularly heightened when
working with elite attitudes. For, as Dexter (1970:6) con-
cludes, ". . . a good many well-informed or influential
people are unwilling to accept the assumptions with which the
investigator starts; they insist on explaining to him how
they see the situation, what the real problems are as they
view the matter." Open-ended questioning within a conversa-
tional setting provides a greater opportunity for the respon-
dent to organize his answers in terms of his own framework.
Whether this procedure is wise or not is clearly a matter of
theoretical objective. A researcher interested in evaluating
components of a well-defined model may regard the richness of
response induced by open-ended materials as extraneous. When
one seeks to explain the structure of relationships within a
clearly defined system of variables, one should constrain the
respondent; when investigation is directed to the discovery
of patterns, it is wiser to allow the respondent more initia-
tive. Here, one may attempt to incorporate the respondent's
organizing principles as relevant data rather than view them
as extraneous material. Again, procedure follows research
objectives and those objectives should be formulated in light
of the available knowledge. Thus, response compression proce-
dures are more appropriate for analyzing alternative explana-
tory theories with a relatively concise set of variables.
Open-ended or response expansion procedures seem more appro-
priate for descriptive and exploratory studies. Such work
often precedes our capacity to specify expectations necessary
for working within the response compression mode.

Although imposing structure on the respondent may lessen
the opportunity to capture his pattern of thinking and possi-
ble motivations for response, an unbounded interview situa-
tion presents dilemmas to any researcher seeking to code re-
sponses systematically. There is no easy way out of this
predicament. Still, the choice of procedure should follow
upon rather than dictate the character of one's research con-
cerns. In interviewing elites, highly structured questions
are best for measuring choices between well specified behav-
ioral alternatives. They are inappropriate where the range

of responses is either unknown in advance or highly complex,
i.e., multidimensional. Structured questions are least at-
tuned to exploring and mapping basic beliefs, attitudes and
values. Similarly, Dexter (1970:55) concludes that:

> In most political interviews . . . where the
> important thing is the discovery of a social
> pattern or value of any sort, it is important
> to start off with comments or ask questions
> where the key words are quite vague and ambig-
> uous, so the interviewee can interpret them in
> his own terms and out of his own experience.

Finally, a third consideration guiding our choice of an
open-ended instrument rested upon the potential <u>receptivity
of the respondents</u>. While this may seem to be a theoretic-
ally trivial matter, its importance in gaining the coopera-
tion of elite interviewees and in carrying out the research
effectively should not be underestimated. Discussing her
experiences in interviewing an intellectual elite, Harriet
Zuckerman (1972:167) observes:

> They soon detect whether questions are
> standardized or tailored to their own
> interests and histories. They resent
> being encased in the straightjacket of
> standardized questions.

Zuckerman's experience accords well with our own. On a num-
ber of occasions, the administrators and congressmen remarked
informally after the taped interview that they had found the
interview experience to be an enjoyable one. Some commented
that they were able to be more frank in their responses than
they expected. When they discovered the broad-ranging char-
acter of our questions the vast majority appeared to relax
and open-up. It was a rare opportunity for many of them to
draw upon their experiences as administrators and congressmen
and to philosophize about society and politics. But because
our concerns were standardized, even though we did not fully
specify the response options, a few of our respondents found
some questions overly confining. For example, one focus of
this cross-national study was on whether elites in different
countries viewed the nature of political and social life as
fundamentally consensual or conflictual. Our phrasing was
abstract, but its purpose was to allow the respondent some
leeway to define the terms of discussion. The following dia-
logue with one of our administrative respondents is instruc-
tive:

    1. Let me move to another one of those sweeping
        questions again. Sociologists have often

> described society in terms of two kinds of
> models, one of which emphasizes conflict and
> talks largely about conflict among groups, ah,
> sort of a conflict model of society; that is,
> there's always going to be a great deal of
> conflict among the groups that constitute
> society. On the other hand, there's also
> something that I guess you could call the
> consensus model which essentially argues that
> ultimately there's really a great deal of
> consensus and agreement among these groups.
> Ah, how would you evaluate these alternative
> models in the light of American society? In
> other words, which one do you think is basic-
> ally more accurate?

R: Ah, gee, I'm being caught on the horns of a
     choice here that I really don't care for.

I: Expand your alternatives as you will.

Subsequently, this individual responded in a way that
suited him and provided the basis for a highly informative
discussion of both his perceptions and views on social con-
flict and consensus. In many instances, the frameworks of
discussion chosen by respondents provided information that we
would probably not have been able to anticipate. Thus, the
ways individuals responded to open-ended questions allowed us
to see and to code for various organizing dimensions that
would have been lost in all likelihood had we proceeded with
greater closure. Measurement error, of course, stubbornly
crops back into the coding process with data such as these,
but, to some extent, it can be better estimated there than at
the source. Close-ended questioning is less adaptive from
the standpoint of capturing organizing principles and re-
sponse frames. Furthermore, a battery of close-ended items
is more likely to be seen as a grilling than an interview by
public officials. Elites not only tend to hold strong views,
but they are able to articulate them clearly. It is our im-
pression, sustained in post-interview informalities, that a
good many of our interviewees would have been less frank if
confronted with a less flexible interview schedule. There is
no way of telling how many interviews might have been termin-
ated prematurely without this flexiblity.
    Nevertheless, open-ended questions are not an unmixed
blessing. From his experiences in directing interviews with
a broad assortment of Yugoslav elites, Bogdan Denitch (1972:
150) asserts that "The study reaffirmed my bias against ex-
cessive use of open-ended questions: Not only were they
clumsy to apply and time-consuming to code, but at the end

they were the least used in analysis." There is certainly
little doubt that open-ended questions are far more difficult
to code and far more time-consuming to administer, but these
reasons alone are not sufficient to limit their use. The
critical questions have to do with research objectives.
Specific informational, recall, and opinion data are probably
better handled with tighter structuring. Broader predisposi-
tional data are captured better with less structure, despite
the drawbacks of time and coding difficulties. There is high
analytic risk in imposing structure when expectations are not
well defined.

In sum, we found an interview strategy oriented largely
to open-ended questions within a conversational setting to be
preferable on at least three grounds. First, this strategy
appeared to be most compatible with our research objectives,
which were geared mainly to the exploration of elite value
patterns and perceptions. Second, we felt this strategy to
be one that would maximize response validity. Finally, we
thought that open-ended questions would evoke more coopera-
tion from elite respondents than a close-ended format.[4]

III. THE QUESTIONNAIRE

Initially, in constructing the questionnaire we were
concerned about the possibility of encountering response set.
While this is ordinarily not a problem when an open-ended in-
strument is used, it was anticipated that an elite sample
might be sensitive to possible inconsistencies in responding
to questions they thought were tapping similar or related
matters. Therefore, we separated some questions to minimize
this problem, even at the expense of a more logical pattern.
In practice, however, limitations on available interview time
(which quite often are not known by the interviewer in ad-
vance) frequently forced us to roll with the direction of the
interviewee's responses. In short, a number of questions had
to be slipped in more obliquely and casually than the inter-
view schedule would indicate.

Similarly, constraints stemming from the interview con-
text forced some slight shifts in phrasing questions from
time to time, though the substantive thrust of the question
was always maintained. In other words, having begun with the
usual textbook concerns for pristine instrument and measure-
ment control, it became apparent that the more we adhered to
these formulations the more we would end up treating our
elite sample like "bricklayers." Moreover, we believe that
to have adhered rigidly to the formal sequencing and phrasing
of the questionnaire would have reduced substantially the
number of completed interviews.

In open-ended elite interviewing, it is very difficult

to get full answers to all questions. As the range of time
taken in these interviews attests, it is nearly impossible
for the interviewer to know in advance how garrulous a re-
spondent will be or how literally he will take the time esti-
mates for the interview communicated to him in the cover
letter. Therefore, some priorities had to be developed be-
forehand with respect to the most important questions. Also,
informal rules of thumb stemming from early interviewing ex-
periences had to be conveyed to all members of the interview-
ing staff to let them know when they might have to drop some
of the questions. As a result, the rate of response to some
questions is considerably lower than for others. Questions
which were especially pertinent to the cross-national facet
of the study were given the highest priority. During the
course of the interviewing we raised the priority of some
questions to which responses were particularly interesting.

        In a substantial number of cases, extensive post-inter-
view discussions took place between the interviewer and the
respondent. While these were especially frequent for those
interviews that the political scientists on the staff con-
ducted, all of the interviewers were instructed to take notes
of these discussions after they had left the respondent's
office. In many instances, these discussions elaborated
specific facets of the formal interview or added new inform-
ation. Sometimes the interviewer felt more rushed during the
formal interview session than he or she needed to be, so
these informal discussions provided an opportunity for the
interviewer to raise questions that had been skipped in the
formal portion of the interview. With an elite population,
as Denitch (1972:157) points out, one is forced to take the
interviews when the opportunity arises. In open-ended elite
interviewing one frequently must ask the questions when they
can be asked even though this is inconsistent with the con-
ventional wisdom of adhering to a tightly knit interview
schedule.

IV.   THE INTERVIEW SITUATION

        All but two of the administrator interviews were con-
ducted in the interviewee's government office. One was con-
ducted at the Washington office of a study director; another
was conducted in the private office of a respondent who,
only days prior to the interview, had left his government
post for a position as a university administrator. The con-
gressmen were interviewed either in their offices or in the
Rayburn Reception Room just off the House floor. Part of
one interview was conducted in the congressman's home dis-
trict office.

        The interviews were taped to facilitate the use of a

conversational style in the questioning and to minimize in-
formation loss. Only five administrators objected to the use
of the tape recorder. Four of these objectors were career
civil servants. Interestingly, only one of the political
officials refused to be taped. Among the congressmen, three
respondents refused us permission to use the recorder and two
interviews were not taped because of technical difficulties.
While mechanical troubles, usually the interviewer's fault,
plagued some of the recordings in spots, these cases were
exceedingly few and recovery of information was generally
quite complete. Only in two cases were these difficulties so
substantial and the interviewers' recall insufficiently com-
plete that the interviews had to be eliminated from the final
data file.

The average interview with the administrators lasted 69
minutes. The spread around that mean is fairly large, how-
ever, with a standard deviation of 26.1 minutes. The range
is also quite substantial with 30 minutes at the lower limit
and 2 hours and 45 minutes at the upper limit. We anticipat-
ed that the higher the administrator's status, the less ex-
tended would be the interview, and the data in Table I con-
firm this expectation. The mean interview time for the con-
gressional interviews was 42.9 minutes. (The interview
schedule for the congressmen was shorter than the administra-
tor questionnaire.) The standard deviation is 24.1 minutes
and the interviews ranged in length from 15 minutes for a
respondent who was in a great hurry to 3 hours and 30 minutes
for a very loquacious congressman. Surprisingly, there were
no significant differences in the mean times of interviews
with those outside of the leadership and those with leader-
ship positions (chairman or ranking minority members of com-
mittees or sub-committees).

TABLE I
MEAN INTERVIEW TIME FOR ADMINISTRATOR AND
CONGRESSIONAL INTERVIEWS

Sample	Mean Interview Time (in minutes)
Administrators	66.0
Political	66.0
Career Service	71.2
Congressmen	42.9

Our judgement, shared by our coders, is that the major-
ity of the respondents answered the questions frankly. More
than three-fourths of the interviews (78.6 percent of those
with administrators and 80.5 percent of those with congress-
men) were coded as either frank or very frank. (See Table
2.) These figures are comparable to those reported by Wahlke,
Eulau, Buchanan and Ferguson (1962:37) in their study of leg-
islators in four states, and to those reported by Putnam
(1973:21) in his study of British and Italian parliamentar-
ians. The questions were designed to minimize the respon-
dents' temptation to answer in an evasive or dishonest manner.
They elicited general views and no questions were asked seek-
ing information that might jeopardize their personal inter-
ests.

TABLE 2
CODERS' JUDGEMENTS OF INTERVIEW FRANKNESS

% of Interviews Judged to Be*:	Administrators Sample	Congressional Sample
Very Frank	49.2%	57.1%
Frank	29.4	18.2
Mixed	9.5	5.2
Somewhat Reserved	7.1	10.4
Very Reserved	.8	2.6
Not Recorded	4.0	6.5
	100.0%	100.0%
	(N=126)	(N=77)

*The inter-coder reliability measure for this item is,
T_b = .312 for the administrators and .710 for the congress-
men. A discussion of the inter-coder measure of agreement is
found in section VI.

Four interviewers were involved at one time or another
in the study of administrators and three in the congressional
study. All of the congressional interviews were done by pol-
itical scientists, but 40 percent of the interviews with ad-
ministrators were not. Considerable effort was expended in
training members of the interviewing staff, especially those
with experience limited to more standardized formats. This
training focused not only on the formal content of the ques-
tionnaire but also on the underlying objectives of the ques-
tions and the possibilities for probing. Obviously, stan-
dardizing the questionnaire would have yielded greater con-

trol, but at the cost of our interviewing philosophy. In-
stead, we opted for intensive instruction, constant feedback
and correction, and prayer.

After each day's interviewing a study director met with
other interviewers, singly or collectively, listened to their
tapes and discussed the problems they had encountered. Pos-
sible changes and corrections in interview prodedure were
suggested at each of these sessions. The study directors
also met and reviewed the procedures as often as possible.

Yet, in a relatively loose interview situation, tight control
is always elusive. When unanticipated circumstances arise
during the interview it is difficult to make proper adjust-
ments without a genuine understanding of the research objec-
tives. Despite the training sessions, those who are not
social scientists do not always understand those objectives.
The virtue of providing early feedback from the recordings
was to standardize certain types of adjustments to be made
during the interviews. On the other hand, interviews of this
nature provide a certain amount of idiosyncratic response and
interaction for which it is most difficult to standardize
interviewer reaction.

V.   GAINING ACCESS

All administrators and congressmen in the sample receiv-
ed a letter broadly describing the outlines of the project
approximately a week to two weeks prior to a telephone fol-
low-up to arrange an appointment. The letter emphasized our
interest in studying the kinds of challenges that public ad-
ministrators and legislators in a variety of societies would
face in the future. Because the letter was not specific, the
interviewer was in a position to elaborate some of the con-
crete details of the study at the onset of the interview.
Some respondents interviewed the interviewers, probing for
underlying theoretical objectives of the study. We were re-
luctant to be drawn into such detail prior to the interview.
Consequently, we responded as generally as possible to ques-
tions of this sort, indicating a preference for elaboration
after completion of the interview.

It was much more difficult to arrange interviews with
the congressmen than with the administrators. For example,
it took an average of 3.97 contacts to arrange an interview
with the typical congressman, but only 1.52 contacts to gain
entry to an administrator's office. Congressmen broke many
more appointments, not only because of their hectic work envi-
ronment, but also because frequent requests for interviews
have stiffened their resistance.

Interviewers who were not social scientists were given
detailed instructions for introducing the study and for re-

sponding to particular questions that we expected to arise
with some frequency. The assistants were further directed,
however, to refer additional questions pertaining to research
objectives to the study director. The very small number of
inquiries seemed to be motivated more out of an intellectual
interest in the project than by suspicion. The interview it-
self provided the respondents a rare opportunity to express
their thoughts in a reflective manner removed from their
daily routines.

VI.   CODING PROCEDURES

        In surveys where most questions are standardized and
closed, coding procedures are often of secondary importance.
Even for the open-ended questions asked in standardized sur-
vey studies, the choice of coding dimensions and categories
often seems reasonably clear because many of the items have
been used frequently and are relatively straight-forward. As
a result, the coding process is not as complex or time-con-
suming as it is with a more conversational interview format.
Coding procedures assume paramount importance when, as in
this study, open-ended interviewing techniques are employed
to elicit subtlety and richness in responses and when one
wishes to make use of this information through quantitative
analysis.[5] The coding problems we encountered resemble those
facing content analysts (Holsti, 1969). The major difference
is that interviews rather than recorded documents constituted
the subject of a content analytic coding procedure.
        Data collected from semi-structured, open-ended inter-
views can be analyzed quantitatively. A semi-structured in-
terview which is coded systematically using clear code cate-
gories and competent coders can yield data comparable in form
and, for our purposes, superior in quality to data from very
structured, close-ended questionnaires. Even if the inter-
view situation and instrument are not highly structured,
structure may be gained during the coding process. The fun-
damental methodological issue is not whether the questions
should be structured, but when in the research procedure the
investigator should impose a structure on, or derive a struc-
ture from, the respondent's opinions. In elite interviewing
we think it safer to allow the respondents to answer in their
own ways and then to build a codification system which main-
tains the richness of individual responses but is sufficient-
ly structured that the interviews can be analyzed using quan-
titative techniques. There are, of course, problems in the
coding stage because of the semi-structured nature of our in-
terview instrument and because of errors which coders inevi-
tably make in classifying respondents' answers. However,

our coding procedures, which are described below, minimized
the errors and yielded reasonably reliable data. The ap-
proach we followed with these elite respondents not only en-
abled us to enjoy the benefits of semi-structured interviews
but also yielded data coded in precise and standardized cate-
gories whose form often resembled data from close-ended inter-
views (i.e., ordinal scales).

Simple, manifest coding procedures alone are inadequate
to capture the richness and nuance of these interviews. By
taping and transcribing the interviews, we were in a position
to employ a highly complex coding scheme which applied very
specific codes for answers to explicit questions, codes gaug-
ing latent features of the respondents' answers, and codes
requiring coders to make "global judgements about certain
traits of the respondents, based on the interview as a whole."
(Putnam, 1973:242) Our coding scheme created intriguing,
perplexing, and extremely time-consuming problems. Full re-
liance upon codes tapping only the most explicit, manifest
responses, however, would have defeated the purpose of the
interview style employed in this study.

Coding latent and contextual characteristics strengthens
our capacity to represent meaning and subtlety; it also
leaves room for greater error. Since the utility of latent
and global coding items is highly dependent upon the clarity
of the coding criteria and the code categories, each inter-
view was coded independently by at least two observers so
that measures of inter-coder reliability could be obtained.[6]

To be precise, manifest coding items measure direct re-
sponses to particular questions. For instance, when reson-
dents were asked whether the differences between the politi-
cal parties are great, moderate, or few, the answers were
coded directly in these terms, though these simple distinc-
tions often had to be extracted from more elaborate re-
sponses.

A latent coding item, on the other hand, is one that
reflects our structuring of an individual's style of response
to a given question or set of questions. In other words, a
latent item defines characteristics of response not explic-
itly called for by the questions themselves. For example,
from a series of questions dealing with respondents' beliefs
about the character of conflict in society, the degree to
which conflict seems to be capable of reconciliation, and the
amount of conflict existing in American politics, we con-
structed coding items dealing with positive and negative ref-
erences toward the role of conflict. Moreover, we construct-
ed codes which specified a particular set of reasons for be-
lieving that conflict was functional as well as a particular
set of reasons underlying negative comments about conflict.
These reconstructed coding items can be characterized as
latent since they refer to a mode or style of response not

directly asked for in any one question.

Finally, we also used a large number of <u>global coding</u> <u>items</u> which permit the coders to form broad contextual judgments from the entire interview transcript, and which tap general traits and styles characterizing the entire interview. An example is our attempt to measure the extent to which respondents employed coherent conceptual schemes throughout the interview. Thus, the coders were asked to consider whether, and to what extent, an individual employed a coherent framework in responding to political questions. The coders, in other words, were asked to summarize, in a specific code, the way a respondent framed his responses to questions throughout the interview.

Coding based upon considerations of meaning was both necessary and risky. In a fairly fluid interview context, we had to permit the coders some discretion to read the interview record for its contextual meanings (i.e., meanings gleaned from more than a single response) and to make judgments about the traits and styles of thought of the respondents. Yet, such procedures invite ambiguity and potential coder bias. Though close-ended questions are more likely than open-ended questions to possess <u>stimulus error,</u> open-ended questions tend to possess a greater degree of <u>transla-</u> <u>tion error</u>. Both are measurement error problems.

Table 3 presents the mean inter-coder reliability coefficients (T_b) for the 112 precisely comparable coding items in the administrator and congressional studies. The reasons for the higher average coefficients calculated from the congressional data will be examined below. For now, let us note the differences in the mean inter-coder reliability coefficients for the manifest, latent and global coding items. As expected, the manifest items have the highest coefficients and the global items the lowest. The latent items are in the middle, but closer to the manifest than to the global items. The more the coders have to use their judgments and the larger the portion of the interview transcript they have to cover, the lower the mean inter-coder reliability coefficient tends to be.

TABLE 3
MEAN INTER-CODER RELIABILITY COEFFICIENTS (T_b)
FOR 112 COMPARABLE CODING ITEMS IN THE
ADMINISTRATOR AND CONGRESSIONAL STUDIES

Mean Coefficients*	Administrators	Congressmen	No. of Items
Total Items	.391	.596	112
Manifest Items	.469	.630	53
Latent Items	.367	.598	41
Global Items	.218	.491	18

*The statistic used to measure inter-coder reliability
(agreement) is Kendall's tau-beta (T_b). See footnote 6 for a
brief discussion of its characteristics.

The possible effects of a large degree of coder discre-
tion are considerable. Discretionary coding procedures can
affect the independence of items; when coders are permitted
latitude to examine context, i.e., more than a discrete re-
sponse, the independence of the coding items may be compro-
mised. Relaxing coding constraints to some degree is essen-
tial to bring forth informational richness, but it also in-
creases risk that coders may form a biased Gestalt about each
respondent, which may, in turn, create an assimilating effect
across coding items. Coders worried over their own sense of
consistency may then generate "halo-effects" across a full
interview record (Selltiz, et al., 1965:351). The general
point is that the looser the constraints upon coders, the
greater the potential for systematic error. This was one
reason why we felt that independent double-coding procedures
were necessary. The danger of "halo-effects" and other forms
of item contamination may be especially prominent when public
figures, some of whom are well-known even to the coders, are
the subjects of a discretionary coding procedure.
To minimize these problems, coder meetings were held on
a frequent and regular basis to discuss difficulties. These
discussions helped to clarify boundaries of responses code-
able for a given item. Even here, variations across inter-
view records forced us to pay more attention to "case jus-
tice" than to rigidly held universal rules. The measures of
inter-coder reliability provide an approximation of the clar-
ity of the instructions followed by the coders in searching
for relevant materials to be coded under a given item. The
higher the inter-coder reliability coefficient, the more rea-
sonable it is to assume that two independent coders were per-

ceiving the same thing. Because this whole coding procedure
requires a very active input from the study director, his in-
fluence on the process must be regarded as substantial. De-
spite various control and precautionary measures, our coding
was influenced both by the coders and by the study director.

These comments make clear that the data are not insulat-
ed fully from subjective influences. To some extent, an in-
creased reliance in political science upon secondary data
analysis has fueled a tendency to ignore vital problem areas
involving the creation of data. This, of course, occurs be-
cause these problems are more easily ignored the more removed
one is from the development of the data. The general tenden-
cy of secondary analysis is to emphasize data analysis tech-
niques and to deemphasize questions of measurement validity.

With each item independently double-coded, decisions on
breaking any coding discrepancies were necessary before the
final data set could be built. Hard information such as
background material and explicit responses to specific ques-
tions were the easiest to code, created the fewest interpre-
tive problems, and were the easiest to check for error. But
the more latent or global (i.e., judgmental) the item, the
more the coding was susceptible to interpretive error.[7] It
was especially difficult in such cases to construct rules for
breaking discrepancies.

The procedures used to arrive at final coding decisions
differed between the congressional and administrative inter-
views. They differed in the type of coding personnel em-
ployed, in the circumstances under which these personnel
worked, in the nature and purpose of coder meetings, in the
methods used in order to arrive at the final code for each
interview, and in whether the interviews were coded serially
or simultaneously. The differential circumstances under
which coding for the two samples took place are important
because they apparently affect the reliability of the result-
ing data.

The congressional interviews were coded by virtually
full-time, highly educated coders, with the assistance of a
coding director. Each interview was coded simultaneously by
each coder. The final codes for each congressional inter-
view were the product of a consensus reached between the
coding director and each of the two coders. The congression-
al interviews were discussed in an atmosphere of collegial
decision-making. Both of the coders and the coding director
had to agree on the proper code for the item before the next
item could be considered.

The coding environment and the coding procedures for the
administrative interviews differed significantly from those
in the congressional coding process. The personnel coding
the administrators' interviews were undergraduate students

who were able to work only part-time. The turnover among the
administrative coders was substantially higher than among the
congressional coders. Inevitably, the administrative coders
were less involved with the project than were the congression-
al coders. Unlike the congressional coder meetings where
consensus on substantive codes was attained, the administra-
tive coder meetings were directed largely to enabling the
coders to learn rules and to clarifying ambiguities. Differ-
ences of opinion concerning coding items were voiced in the
meetings, but since the interviews were coded serially and no
two coders were working on the same interview at the same
time, it was not possible to obtain a substantive consensus.
Consequently, after each administrative interview had been
coded twice and the coders had left the project, the coding
director alone had to reconcile differences between the co-
ders. When there were great differences between the coders
on an item, the coding director was forced to make some
choices without direct consultation with the departed coders.

Because of the conditions under which the administrative
data were coded, the coding director sought to impose a uni-
form decision rule for breaking discrepancies between coders
based upon a "best coder" criterion for each dyad. (A "best
coder" is here defined as one who disagrees the fewest times
with other coders.) This proved to be unworkable because
measures that permit some evaluation of individual coders are
typically chi-square based (Cohen, 1960; Funkhouser and Par-
ker, 1968; Guetzkow, 1950; Krippendorf, 1971; Robinson, 1957;
Schutz, 1958-59; Scott, 1955). With small sample sizes, zero
or near zero cells, and abnormal distributions, it is not
feasible to operate on chi-square based assumptions.[8] On cer-
tain items the zero cell problem particularly affected our
data.

In order to break coding discrepancies, then, the coding
director scrutinized each administrator's interview record.
In the absence of an obvious coding error, ordered judgments
were broken randomly when discrepant. In cases where discre-
pant judgments were off by more than one rank, the differ-
ences were usually split. There are several ways of evalu-
ating such a procedure. The most important question is whe-
ther or not one prefers to substitute randomization for care-
ful, if subjective, judgments. The danger is that as the
coding director gets closer to the data, he may become at-
tached to his own judgments and, thus, less willing to allow
random processes to determine the data. There is a tendency
to label something an obvious coding error (i.e., a misinter-
pretation of the rule) and not a difference in judgment. In-
deed, the temptation is well nigh irresistible for the coding
director who has scrutinized the data to want to overrule
judgments discrepant with his own. Typically, this problem

arose only when judgments between the two coders were highly
discrepant. Exceptions to the random split rule occurred
most frequently when the coding director felt that he agreed
with one of the coders. In these cases, the coding director
acted as a third coder supporting one of the judgments. The
procedure for breaking discrepancies in the congressional in-
terviews, of course, occurred with the coders still present
until an agreement was reached.

The differences in the coding procedures employed were
not a matter of choice, they were dictated by administrative
necessity. The fact that the coding operations for the two
samples did differ, though, provides an opportunity to mea-
sure the overall effectiveness of the alternative coding
methods. The average inter-coder reliability coefficients
presented in Table 3 indicate that, on items comparable a-
cross both samples, the congressional data clearly possess
higher reliability.[9]

To conclude, the coding process is also a data creating
process. Typically, no more than a footnote covers the co-
ding process in study reports, but the subject deserves more
attention. With the development of different data bases in
political research, it is necessary to detail and to justify
as explicitly as possible the transformation procedures and
rules by which data are created for quantitative analysis
from qualitative observation. It has sometimes been con-
tended that such activity is a reflection of scientism and
false precision; the primacy of method over content. Nothing
could be further from the truth. The quality of our empiri-
cal statements can be only as good as our understanding of
the underlying measurement processes. Without an increased
sensitivity to data creation procedures it is all too likely
that false precision will be claimed, especially when data
such as these are used in secondary analysis. A science that
is properly self-conscious should be critically aware of this
transformation process. In Alker's words (1971), "...method-
ology becomes a kind of self-critical normative epistemology
..." In elaborating the data creating procedures involved
in this study, our purpose is to indicate the unusually large
number of subjective influences upon the data. For this rea-
son, great pains were taken to ascertain the reliability of
these data.

VII. THE LIMITS ON DATA ANALYSIS

Let us briefly consider some of the implications for
data analysis of the methods we used to collect and prepare
the data. There is, unfortunately, a kind of "Catch-22"
rule that is applicable here. Frequently, the more painstak-
ing the efforts expended in the data creation process, the
less manipulative capability there is for substantive data

analysis. That is, a greater effort is necessarily expended
in developing data which are weakly measured by the research
instrument. And, as we have noted already, the more open the
questionnaire, the more difficult it is to maintain indepen-
dence between items.

Because the link between theoretical construct and mea-
surement in the social sciences is so tenuous, it is gener-
ally wise procedure to develop <u>multiple indicators</u> for theo-
retical constructs. Multiple indicators, in turn, can be
evaluated for fit and dimensionality and are necessary for
analytic reduction of the data. In the study of values and
attitudes especially, reduction techniques such as factor
analysis and scaling procedures assist in structuring the
data more coherently. These techniques are, of course, de-
pendent upon multiple indicators, the absence of significant
amounts of missing data, and independence between items.
Without the use of such techniques it is often difficult to
assess the theoretical significance of any given item.

Most indexing procedures cannot be used readily with
these data. The costs of open-ended questions and contextu-
al coding are most evident here. Open-ended questioning
means that there will be fewer questions in an interview;
hence it is more difficult to build in multiple measures.
Secondly, all multivariate analytic techniques assume a <u>mini-
mum of missing data</u> (since missing data can accumulate across
variables) and <u>orthogonality</u> of the items (in a measurement
sense, not in an analytic sense). The interview procedure
and questionnaire used in this study have led to unusually
high proportions of missing data in both the usual sense of
non-response and in more unusual senses as well.[10] Coding
for latent features or global characteristics, of course,
tends to compromise orthogonality of items to some degree.
Finally, the relatively small sample size under these condi-
tions becomes a critical inhibiting factor when one wants to
employ multivariate analysis techniques.

Because of these difficulties, qualitative data must be
woven into and around quantitative data presentations.
Quotes, sadly, do not account for variance, and the express-
ion "for example" does not generally fit the rules of scien-
tific evidence. On the other hand, we are not positing a
specific explanatory model, since to do so would have in-
volved a different design and wholly different research me-
thods and techniques. Rather, we are exploring the content,
style and pattern of social and political thought of Ameri-
can administrators and congressmen. In this exploration,
resorting to qualitative data is not so much a regrettable
necessity as it is an avenue for <u>interpreting</u> and <u>informing</u>
the quantitative analysis.

VIII.   SUMMARY

This article reports on the methods we used to examine
the political attitudes, values, and beliefs of American ad-
ministrators and congressmen and it discusses the implica-
tions of these techniques. Open-ended, semi-structured
interviews were utilized in the study. Like any method, the
open-ended interview has some disadvantages, but it is espec-
ially well suited to exploratory studies of the type we con-
ducted. It maximizes response validity and it makes the
respondents more receptive to the interview experience. The
interviews were tape recorded, fully transcribed and each was
coded independently by at least two coders. Double coding
allowed us to measure the reliability of the data. Latent
and global as well as manifest codes were used. In this way
we took advantage of the subtlety, nuance and styles of
thought captured in the taped interviews and, as a result, in
our data analysis we will be able to explore, systematically
and in depth, the political views of the American govern-
mental elite.

FOOTNOTES

1. The countries involved, in addition to the United States are Britain, France, Italy, Morocco, The Netherlands, Sweden, West Germany and Jamaica. Our colleagues on the larger project are Samuel J. Eldersveld, Thomas J. Anton, Ronald F. Inglehart, Robert D. Putnam, Archibald Singham and John Waterbury. Professor Putnam's work (1973) has strongly influenced the design of the study and the thinking of his colleagues on the project.

2. We are fortunate in the American part of the study to have some meaningful data readily available on the behavior of our respondents. For the congressmen, roll call votes and records of committee hearings can be used both to measure behavior and to gain insights into environmental factors. The available data are not as good for the administrators, but we often have records of their testimony before Congress. Their statements on policy and administrative practices in oversight hearings should be especially valuable.

3. The study referred to, of course, is Matthews (1960)

4. As Galtung (1969) correctly points out, typical references to close-ended questions really refer to the closure imposed by the researcher upon the answers. This observation is important because there is a necessary distinction to be made between focused and unfocused questions. Neither type necessarily has to have close-ended answers. While question focus should be thought of as a matter of degree not a dichotomy, much of our questioning was closer to the focused than the unfocused end of the continuum. This is in contrast to Harriet Zuckerman's (1972) description of her interviews with Nobel-laureates, where she apparently tailored her line of questioning to the particular respondent. When a <u>very high</u> degree of focus can be attached to the question it usually makes sense to structure the range of answers as well. As has been argued repeatedly here, this structuring should occur when rather precise expectations exist as a result of considerable background knowledge. The difficulties with open-ended material expressed by Denitch (1972) could have had more to do with an absence of focus to questions than the absence of pre-fixed response alternatives.

5. Depth interviewing with a small number of subjects typically is not amenable to quantitative data manipulation.

Lane (1962) did employ some standard psychological tests, but these were peripheral at best to his analysis of the interview materials. Of course, employing quantitative techniques makes sense only where the sample size is sufficient to warrant quantitative data manipulation. This naturally depends upon (a) the unit of analysis, and (b) the form of investigation. If the unit of analysis is some form of interactive behavior, then the sample size of individuals may not be pertinent. Secondly, if the investigation assumes the form of a controlled experiment, then sample size becomes less pertinent since the variables one wants to randomize by generating large sample sizes often are controlled in experimental selection.

6. The statistic employed for measuring inter-coder agreement is Kendall's tau-beta (T_b), used for this purpose because exact agreement for ordered categories should be expected for a perfectly reliable measure. The tau-beta statistic is especially sensitive to linear monotonicity. It is a conservative measure of association since its computing formula treats all observations outside of the main diagonal of a matrix as the same.

Missing data categories were included in these matrices because in some of the subtle codes an important issue often arose as to whether a respondent's comments should be construed as belonging in a substantive code category or as missing data for that item. As a result the (T_b) coefficients reported here are an understated indicator of the actual inter-coder reliability of the items. Though not reported here, we also have other indicators of inter-coder agreement.

7. As Sussman and Haug (1967) point out, however, a sizeable amount of error crops up even in more routine coding schemes, though these can be checked more readily than judgmental items which compound known errors stemming from coder fatigue with unknown errors stemming from category or boundary ambiguity.

8. Funkhouser and Parker's RSE coefficients (1968) seemed most attractive for these purposes, at least until their chi-square base was noted. The RSE coefficients will yield measures assessing (1) the proportion of disagreements for each coder with respect to each code category, and (2) disagreements in individual cells of the matrix. Thus, where C = the number of categories and N = the number of disagreements with respect to each coder, the Funkhouser and Parker measure may be expressed as:

$$\text{RSE} = \sqrt{\frac{C\Sigma N^2 - (\Sigma N)^2}{(\Sigma N)^2 \quad (C-1)}} \text{ which reduces to } \frac{x^2}{x^2 \text{ max}}$$

Since the process of coding the administrators' inter-
views took approximately fifteen months to complete, the turn
over of coders was sizeable. Different coders did dispropor-
tionate amounts of coding, making it difficult to compare
coders directly. Formalizing "a best coder criterion" pre-
sented formidable difficulties.

9. It also should be noted, with respect to the inter-
coder reliability coefficients, that the larger the number of
coding categories for a given item, the greater the probabil-
ity of error (Holsti, 1969:138). To the extent, therefore,
that some items ultimately are collapsed in analysis, reli-
ability for these items often is strengthened. Upon close
inspection, many of the coder errors tend to be trivial
(coding in different missing data categories, for example)
and many of the coding discrepancies are slight when a fair
number of gradients are involved. Nonetheless, the effects
upon the inter-coder reliability measures can be substantial.
In addition, one can assume that the coefficients of inter-
coder reliability actually represent the lower limits of re-
liability because of the increment in the final data set re-
presented by the consensus of the coders or the judgment of
the study director.

10. For example, the multi-mention codes constitute an
unorthodox missing data problem. Since these derive from
open-ended questions, respondents can cite a varying number
of conditions or attributes. This non-uniformity makes it
difficult to employ data reduction techniques dependent upon
multiple correlations. As a corrective, attributes mentioned
in multi-mention codes were sometimes generalized, with each
generalized attribute dichotomized for its presence or ab-
sence allowing some further data structuring capability. At
the same time, it should be recognized that the generalized
dichotomies (sometimes trichotomies) are the product of a
priori decisions rather than decisions based upon empirical
structure.

REFERENCES

Alker, Jr., Hayward R.
     1971 "Research Paradigms and Mathematical Politics."
           Unpublished paper presented to the Round Table on
           Quantitative Methods and Political Substance --
           Toward Better Research Strategies. University of
           Mannheim, Germany.

Bakan, David
     1969 On Method: Toward a Reconstruction of Psychologi-
           cal Investigation. San Francisco: Jossey-Bass.

Brown, Bernard E.
     1969 "Elite attitudes and political legitimacy in
           France." Journal of Politics 31:420-442.

Brown, Steven R.
     1970 "Consistency and the persistence of ideology:
           some experimental results." Public Opinion Quar-
           terly 34:60-68.

Cohen, Jacob
     1960 "A coefficient of agreement for nominal scales."
           Educational and Psychology Measurement 20:37-46.

Denitch, Bogdan
     1972 "Elite interviewing and social structure: an
           example from Yugoslavia." Public Opinion Quarter-
           ly 36:143-158.

Deutsch, Karl, Roy C. Macridis, Lewis J. Edinger, and Richard
L. Merritt.
     1967 France, Germany and the Western Alliance: A Study
           of Elite Attitudes on European Integration and
           World Politics. New York: Scribner.

Dexter, Lewis A.
     1970 Elite and Specialized Interviewing. Evanston,
           Illinois: Northwestern University Press.

Funkhouser, G. Ray and Edwin B. Parker
     1968 "Analyzing coding reliability: the random-syste-
           matic-error coefficient." Public Opinion Quarter-
           ly 32:122-128.

Galtung, Johan
     1969   Theory and Methods of Social Research. New York:
            Columbia University Press.

Greenstein, Fred I.
     1969   Personality and Politics. Chicago: Markham.

Guetzkow, Harold
     1950   "Unitizing and categorizing problems in coding
            qualitative data." Journal of Clinical Psycholo-
            gy 6:47-58.

Holsti, Ole R.
     1969   Content Analysis for the Social Sciences and Hu-
            manities. Reading, Mass.: Addison-Wesley.

Krippendorf, Klaus
     1971   "Reliability of recording instructions: multivar-
            iate agreement for nominal data." Behavioral
            Science 16:228-235.

Lane, Robert E.
     1962   Political Ideology: Why the American Common Man
            Believes What He Does. Glencoe, Ill.: The Free
            Press.

Madge, John
     1965   The Tools of Social Science. Garden City, N. Y.:
            Anchor Books.

Matthews, Donald R.
     1960   U. S. Senators and Their World. Chapel Hill:
            University of North Carolina Press.

     1971   "From the Senate to simulation." In Oliver Wal-
            ter, ed., Political Scientists at Work, pp. 9-27.
            Belmont, Calif.: Duxbury Press.

Putnam, Robert D.
     1973   The Beliefs of Politicians: Ideology, Conflict,
            and Democracy in Britain and Italy. New Haven:
            Yale University Press.

Robinson, W. S.
     1957   "The statistical measure of agreement." American
            Sociological Review 22:17-25.

Roll, Charles W., Jr. and Albert H. Cantril
    1972 Polls: Their Use and Misuse in Politics. New
          York: Basic Books.

Scott, William A.
    1955 "Reliability of content analysis: the case of nom-
          inal scale coding." Public Opinion Quarterly 19:
          321-325.

Schutz, W. C.
    1958-59 "On categorizing qualitative data in content
             analysis." Public Opinion Quarterly 22:503-515.

Selltiz, Claire, Marie Jahoda, Morton Deutsch and Stuart W.
Cook.
    1965 Research Methods in Social Relations. New York:
          Holt, Rinehart and Winston.

Sussman, Marvin B. and Marie Haug
    1967 "Human and mechanical error -- an unknown quantity
          in research." American Behavioral Scientist 11:
          55-56.

Wahlke, John C., Heinz Eulau, William Buchanan and Leroy C.
Ferguson.
    1962 The Legislative System: Explorations in Legisla-
          tive Behavior. New York: John Wiley and Sons.

Zuckerman, Harriet
    1972 "Interviewing an ultra-elite." Public Opinion
          Quarterly 36:159-175.