QUESTION WORDING AS AN INDEPENDENT VARIABLE IN SURVEY ANALYSIS: A FIRST REPORT

Stanley Presser & Howard Schuman, University of Michigan

During the 1940's and early 1950's a number of experiments on attitude question wording and format were carried out by both academic and commercial survey researchers, seeking to determine whether different ways of asking the "same" attitude item led to different results. These experiments became most widely known through Cantril's 1944 collection of papers on Gauging Public Opinion and Payne's little book, The Art of Asking Questions (1951). An example of such experiments is a pair of questions on freedom of speech reported by Rugg in 1941. One national sample was asked: "Do you think the United States should allow public speeches against democracy?" A "comparable" sample was asked: "Do you think the United States should forbid public speeches against democracy?" It turned out that 20% more people were willing to "not allow" such speeches than were willing to "forbid" them -- a difference suggesting that a seemingly innocuous word change can shift univariate item results noticeably.

By the late 1950's such question wording experiments had largely disappeared from major surveys. We think the reasons for their demise were several. First, the basic fact that item marginals are in part a function of question wording was by then recognized, at least in theory, by virtually all academic survey researchers. Beyond repeatedly demonstrating the phenomenon, further experiments seemed to serve no particular purpose, as McNemar had observed in his extended review of "Opinion Attitude Methodology" in 1946. Of even greater importance, we believe, was the fact that virtually all these early reports of experiments were restricted to univariate results, while survey analysis is obviously concerned with relationships among variables. The assumption developed among investigators that though marginals could not be trusted, due to question wording uncertainties, associations among variables were not subject to this same sort of instability. Exactly this assumption seems implied, for example, by Davis's (1971) admonition to students that: "You should always be suspicious of single-variable results... It is well known that the distribution of answers on attitude and opinion questions will vary by 15 or 20 percent with apparently slight changes in question wording, even though both versions may provide valid orderings " (emphasis added). What Davis states explicitly seems, in the absence of warnings to the contrary, to have been accepted by many other survey methodologists and practitioners.[1]

The assumption of "format resistant correlations," as we will call it, was further bolstered as academic survey analysts came to stress the use of attitude scales. On the one hand, this steered the analyst away from single-item percentage results, with their illusion of absolute proportions for and against specific social objects or positions. On the other hand, attitude scaling is intended to reduce idiosyncratic effects of individual items, though how this will necessarily eliminate systematic format effects is rarely spelled out. It should be noted that use of scales or indices is often incomplete in even the best surveys; major variables may be constructed in this way, but parts of analysis frequently continue to draw on single-item variables because lack of time or anticipation prevents detailed measurement of every theoretical construct of interest.

Another reason for the decline of question wording experiments lies in the ad hoc character of most of the early work. Even in terms of univariate results, the reports seldom addressed larger theoretical issues of question construction and typology, nor was there much concern to replicate findings or to estimate the frequency, magnitude, or underlying nature of question wording effects. For this reason, wording experiments have come to be treated anecdotally, reported as illustrative warnings in most survey methods books, but not further developed theoretically or empirically.[2]

Goals of the Present Research

Our present research attempts to return to the question wording experiments of three decades ago, but to do so with a different primary focus, a more systematic concern with question type, and some improvements in methodological procedure and analysis. The change in focus arose when Schuman and Duncan (1974), in the course of several different substantive analyses, came upon variations in question wording that seemed to affect bivariate as well as univariate distributions. These examples were at best suggestive, some being seriously defective from an experimental standpoint and others too ambiguous in terms of item wording to be representative of contemporary surveys. The present project was initiated as a more systematic attempt to test the hypothesis of "format-resistant correlations," an attempt which grew to include rudimentary development of a typology of question forms, as well as treatment of some related issues. Two sets of interrelated experiments using Survey Research Center national samples have thus far been carried out, and we report here some first results of these surveys.[3]

Underlying this research are two general hypotheses. First, where question wording alters marginals appreciably, it seems unlikely from a theoretical standpoint that those persons being affected are simply a random subsample of all respondents. The effects, after all, are a kind of "self-selection," and self-selection is rarely a randomizing procedure.

Second, and more specifically those affected by format should usually be respondents less involved in an issue, as well as those who are in general cognitively less sophisticated -- the less involved on the assumption that individuals lacking interest or strong feelings on an issue are least likely to have an enduring attitude; the less sophisticated on the assumption that such persons are more likely to misinterpret meanings of particular words or miss unexpressed

implications in a question. Although we have a few indicators of involvement in the specific issues used in our experiments, for the present we have focused mainly on general cognitive sophistication. This is partly because of the availability of one variable, education, as an indicator of cognitive sophistication over all the items, and partly because of the near universal usage of this same variable in survey analyses. Thus, we have used years of schooling as the key variable in this analysis of question form effects.

Types of Experiments and Results

We decided as part of our first set of experiments to replicate one of the question variations of earlier years, in part to gain some assurance that chance factors had not misled investigators about even univariate effects. For this purpose we selected the forbid-allow example described earlier, using random divisions of the 1974 Survey Research Center Omnibus national sample.[4] The univariate results from 1940 and from 1974 are shown in Table 1. There is remarkably close replication in direction and degree of the wording effect after thirty-four years. The 21% difference in 1940 is paralleled by 16% in 1974, and thus the change over time is similar for both forms.[5] While we have no direct evidence as to the cause of the form effect, we assume that "forbid" is simply a more forbidding term than "not allow," and that it is this difference in bluntness of language that makes some people less willing to deny freedom of speech when that form is used.

Figure 1 shows the relations between response and education for the two question forms. In line with hypothesis, form seems to make little difference (6%) for those above the high school level, but has a substantial impact (26%) on those with zero to eleven years of school; high school graduates fall in between. Using Goodman's procedure (1970) for multiway contingency tables, the likelihood-ratio chi square for the three-way interaction is 5.75, $p < .06$.[6] While the borderline significance level makes replication essential, we conclude tentatively that this is a case where tone of word makes a difference not only in marginals but in a fundamental bivariate relationship. In both forms, to be sure, there is a positive relation of education to civil libertarian sentiment; however, the relation is not only stronger for one form, but by having both forms our understanding of the difference in crystallization of these sentiments by education is enhanced. Moreover, it is conceivable that a slightly larger gap in connotation or tone of word could remove entirely the association of education in one form, thus changing conclusions about type as well as degree of relationship to education.

These findings, though they remain tentative because of the borderline reliability of the interaction, point up the danger of survey analysis with single item opinion variables, even where one is interested entirely in associations, not in marginals. Furthermore, while one might reasonably assume that index construction would remove just this type of idiosyncratic effect, we shall also raise some questions about that strategy at a later point. For now we must admit that apart from its cautionary message, the forbid-allow example is not a very constructive one, since it would be difficult to predict in advance the effects of other variations in verbal connotation, and impossible to generalize from it to other types of items.

With this latter problem in mind, we constructed most of our other experiments to test types of question formats, rather than using purely idiosyncratic examples. Taking the point of view of the survey investigator faced with the need to create or select a pool of attitude items, we noted certain decisions that typically need to be made - and that typically are made on the basis of rule-of-thumb, personal preference, convenience or simply chance. This led us to formulate four sets of question types, though they are clearly not completely exclusive of one another in either conception or operation. We shall briefly describe and illustrate each of the four types. Note that in operationalizing each question type, we started wherever possible from items used in national surveys in order to assure realism in results.

1. Agree-Disagree vs. Forced Choice. Speed and convenience of administration frequently recommend agree-disagree items in questionnaires and interviews. But some past research,[7] as well as a certain intuitive regard for "fairness" in presentation of issues, suggests the desirability of providing forced-choice formats, rather than single propositions to be accepted or rejected. We have thus far tested this type of form variation in five experiments. Two use items dealing with the causes of crime and with the political role of women, respectively, and the other three employ questions concerning foreign policy issues. All five show univariate differences significant beyond the .05 level, and two of the five provide significant three-way interactions with education, as illustrated in Table 2. The agree-disagree form of the item on women in politics, which was included in both our NORC and SRC experiments, shows a bivariate relation to education, with the contrast found mainly between the least educated and the two higher educated groups. The forced-choice version, on the other hand, produces no clear relation to education. Thus, one would draw different conclusions about the existence of a relation to education in this case, not only about its magnitude. The response-by-education-by-form interaction is significant beyond the .01 level.

The second clear-cut interaction actually involved three forms of an item on crime: two separate agree-disagree statements that are logical contraries and a third forced-choice version. The two contraries show a significant three-way interaction ($p < .05$) in relation to education, and one of these also shows a borderline interaction ($p < .10$) with the forced-choice version. In the latter of these comparisons, the three-way interaction with education occurs despite almost identical item marginals, so that one would not have had even the initial clue of a

TABLE 1

Comparison of Forbid and Allow Marginals in 1940 and 1974

Allow Form: Do you think the United States should allow public speeches against democracy?	1940	1974
1. Yes (Allow)	25%	56%
2 No (Not Allow)	75	44
	100	100
N	(*)	(494)

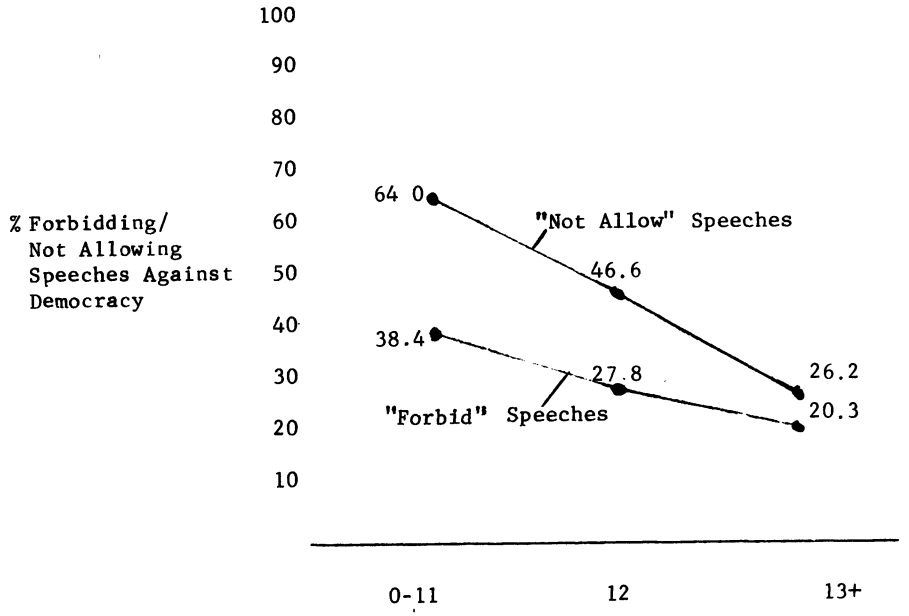
Forbid Form: Do you think the United States should forbid public speeches against democracy?	1940	1974
2. No (Not Forbid)	46%	72%
1 Yes (Forbid)	54	28
	100	100
	(*)	(936)

Response by Form, 1974 data only: $X_1^2 = 35.75$, $p < .001$

*N's for 1940 are not given in Rugg (1941) from which the earlier percentages are taken, but these were large national samples, presumably quota in design Percentages for all distributions have been recomputed omitting DK responses; their inclusion does not change results appreciably

FIGURE 1

Percent Opposing Free Speech by Education and Form (1974)



	0-11	12	13+
N (Allow Form)	(136)	(174)	(183)
N (Forbid Form)	(276)	(313)	(340)

two-way difference (i.e., response by form) to alert one to the possibility of this further important complication.

TABLE 2

Agree-Disagree Example

Percent Against Women in Politics, by Education and Form*

		Total**	Education		
			0-11	12	13+
Do you agree or disagree with this statement? "Most men are better suited emotionally for politics than are most women."	% Agreeing	46.2	55.4	43.6	40.3
	N	(1664)	(520)	(557)	(583)
			X_2^2=27.28, p < .001		
Would you say that most men are better suited emotionally for politics than are most women, that men and women are equally suited, or that women are better suited than men in this area?	% Men Better Suited	35.6	35.6	39.3	31.5
	N	(1199)	(399)	(397)	(397)
			X_2^2= 5.31, p < .10		

Response by Form: X_1^2 = 32.04, p < .001

Response by Education by Form: X_2^2 = 11.31, p < .01

*The original agree-disagree item was already planned for the 1974 NORC General Social Survey, and we were allowed to add to that same survey the parallel forced-choice item shown here. Random subsamples were employed. Both forms were replicated in our 1974 SRC survey. Similar trends appear in the two surveys and we have combined the two sets of results here to increase sample size.

**Total N's differ slightly from sum of N's for educational categories due to missing data on education. The same applies to subsequent tables.

All three foreign policy items show modest but non-significant differences between forms in their relation to education -- the variation in response being greatest for the least educated. Other than to note the difference by subject matter, for now we leave open the apparent lesser susceptibility of these items to higher-order interactive effects with education.

2. Formal vs. Substantive Balance. A second and related type of issue that has concerned attitude survey investigators is that of balancing interrogative items. For example, the first question in Table 3 asks about gun control in the briefest way possible, following a form that was often used in surveys in earlier years. Probably in response to criticism that this type of format discouraged negative answers, the second form, in one wording or another, has tended to replace it.[8] We call this an example of "formal balancing," but our hypothesis was that it would have little effect, since the original form is a question rather than an assertion and negative answers are quite clearly implied as legitimate. An example of what we call "substantive balancing" is shown in the third version, where another side of the issue is not only stated, but justified. We hypothesized that this type of an opposing argument would lead to changes in response distribution.

We tested four different items using both

TABLE 3

Balancing Example

Percent Favoring Gun Control by Education and Form*

	Total	Education		
		0-11	12	13+
A. Would you favor a law which would require a person to obtain a police permit before he could buy a gun?	71.0	68.9	70.8	72.7
	(455)	(119)	(168)	(165)
		$X_2^2 = 0.49$, n.s.		
B. Would you favor or oppose a law which would require a person to obtain a police permit before he could buy a gun?	71.7	69.9	71.6	72.6
	(445)	(103)	(162)	(179)
		$X_2^2 = 0.24$, n.s.		
C Would you favor a law which would require a person to obtain a police permit before he could buy a gun, or do you think such a law would interfere too much with the right of citizens to own guns?	67.3	62.5	63.9	73.4
	(431)	(96)	(166)	(169)
		$X_2^2 = 4.80$, $p < .10$		

Response by Form (A and B): $X_1^2 = 0.05$, n.s.

Response by Form (B and C): $X_1^2 = 2.00$, n.s.

None of the response by education by form interactions is significant.

*Forms A and B were taken with slight modification from Gallup questions. See pages 2027 and 2077 of Volume 3 of The Gallup Poll, New York: Random House, 1972.

kinds of balance -- that is, with three form variations for each test. As predicted, the addition of a formal alternative produces virtually no change in univariate percentages in three instances. In the fourth case, for reasons we do not yet understand, it does show a significant change. Not surprisingly, in all four cases the substantive variation produces the larger difference from the unbalanced form -- reliably different from it in three of the four experiments, the non-significant exception being the gun control item. However, in none of the four experiments does the variation, either formal or substantive, significantly affect the response-by-education relationship. Unlike the agree-disagree vs. forced-choice problem, we cannot reject the null hypothesis for three-way form effects, at least in relation to education. We are not yet ready to draw final conclusions here, however, since there are non-significant trends (the lowest educated being the most affected by the addition of a substantive alternative) and there is also a significant interaction with race on the gun control item.

3. Middle Alternatives. When forced-choice questions are employed, frequently there is a logical middle alternative, as in the example on Vietnam aid shown in Table 4. Investigators sometimes choose to omit the middle alternative in order to produce an easier-to-work-with dichotomous question, on the assumption that most respondents opting for the middle position do in fact lean one way or the other. For the three items that we varied in this way, all show significant differences in the middle category percentage, but in no case does the response-by-education-by-form interaction reach significance. It is possible, however, that here the decision of respondents to avoid extremes is associated not with education, but with other respondent characteristics, such as caution or perhaps a general belief in the status quo. One clue supporting this hypothesis is a significant association

between choosing the middle position when it is offered on the Vietnam item and also choosing the middle position on a similarly formatted item on a quite different topic. No such relation occurs for the forms without the middle alternative. Thus, the choice of the middle alternative, when it is explicitly offered, seems to reflect for some respondents a generalized set, although we have not yet identified its correlates. This particular analysis is of potentially great importance, for it points toward not only a difference by form but a possible basis for saying which form is more valid. If offering a middle alternative encourages an irrelevant "set," whereas omitting it does not, then the latter form may be the better one to use in question construction.

We also hypothesized that intensity of feeling might be involved in this type of format, and included a follow-up intensity measure immediately after the item shown in Table 4: "How strongly do you feel about this issue: quite strongly, or not so strongly?" If one assumes that the people who choose the middle alternative when it is

TABLE 4

Middle Alternative Example

Aid to Vietnam by Education and Form*

Looking back, do you think our government did too much to help the South Vietnamese government in the war, or not enough to help the South Vietnamese government?

	Total	Education		
		0-11	12	13+
Too Much	71.7%	65.2%	73.5%	75.2%
(If Volunteered) Right Amount	17.2	22 0	17.3	13.4
Not Enough	11.1	12.8	9.2	11.3
	100 (882)	100 (250)	100 (294)	100 (335)

$X_4^2 = 9.95, \; p < .05$

Looking back, do you think our government did too much to help the South Vietnamese government in the war, not enough to help the South Vietnamese, or was it about the right amount?

	Total	Education		
		0-11	12	13+
Too Much	62.0%	57.6%	55.4%	72.0%
Right Amount	28.8	34.5	34.5	18.2
Not Enough	9.2	7.9	10.1	9.8
	100 (434)	100 (139)	100 (148)	100 (143)

$X_4^2 = 13.65, \; p < .10$

Response by form (collapsing 'too much' and 'not enough'): $X_1^2 = 22.62, \; p < .001$

Response by form (excluding 'right amount'): $X_1^2 = 0.04$, n.s.

None of the response by education by form interactions is significant.

*Modeled after the Vietnam items analyzed in Schuman and Duncan (1974).

offered have less intense convictions on the issue than other people, then when the alternative is not offered and such people are forced toward one extreme or the other, the average level of intensity should drop for the extreme responses. This in fact seems to happen: on the form offering the middle alternative, 77% of the respondents giving non-middle responses feel very strongly about their positions, whereas on the other form the corresponding figure is only 71% (X^2=3.52, 1 df, p<.10). Thus, there is reason to conclude tentatively that this type of question format manipulation is especially likely to affect those who have low intensity of feeling on the issue being asked about. Whether this is true for each issue separately, or reflects a more generalized personality trait, we do not yet know.

4. Opinion Screening Filters. It is well recognized by now that on many issues a large fraction of the public has no opinion, for reasons of lack of information or interest in the issue. In asking survey questions, one can attempt first to screen out those who admit having no opinion, as have SRC election studies for many years. This step has the seeiming merit of reducing the nonsense component in survey data. However, the effect of such screening on substantive distributions appears not to have been carefully studied, and there is some uncertainty whether and how to compare items with and without such filters.

To study this problem we constructed three items dealing with foreign affairs, intentionally varying the presumed familiarity of the issue for

TABLE 5

No Opinion Filter Example: Russian Intentions by Education and Form*

Here are some questions about other countries. Not everyone has opinions on these questions. If you do not have an opinion, just say so. "The Russian leaders are basically trying to get along with America." Do you have an opinion on that? (IF YES) Do you agree or disagree?

	Total	Education		
		0-11	12	13+
No opinion	37.6%	58.5%	37.1%	21.9%
Agree	39.2	25.9	34.9	53.5
Disagree	23.1	15.6	28.0	24.6
	100 (510)	100 (147)	100 (175)	100 (187)

$X^2_4 = 52.05, \quad p < .001$

Here are some questions about other countries. Do you agree or disagree with this statement? "The Russian leaders are basically trying to get along with America."

	Total	Education		
		0-11	12	13+
(IF VOLUNTEERED) No Opinion	15.2%	27.2%	12.7%	7.9%
Agree	49.9	39.7	47.1	60.5
Disagree	34.9	33.1	40.1	31.6
	100 (499)	100 (151)	100 (157)	100 (190)

$X^2_4 = 29.81, \quad p < .001$

Response by form. (Collapsing "agree" and "disagree"): $X^2_1 = 66.72, \quad p < .001$

Response by form (excluding DK): $X^2_1 = 1.24$, n.s.

None of the response by education by form interactions is significant.

*Constructed for this experiment.

respondents. One item dealt with the Soviet Union, one with the Middle East, and one with the 1974 revolution in Portugal -- which we took to represent an ascending order of public ignorance. The two forms that comprised this set of experiments are illustrated in Table 5 by the Soviet Union pair of questions.[9]

Several conclusions can be drawn from these three experiments. First, it clearly is possible to increase the percentage of "don't know" responses substantially by making their legitimacy clear -- the increase being 22% in the Soviet example and about the same in the others. Second, while the univariate change in DK's as such is highly reliable in each case, when all DK responses are removed there appears to be no univariate difference between the two forms that even approaches significance. In other words, the ratio of the agree to disagree responses remains the same across form in each experiment, despite the shift of approximately a quarter of the respondents between DK and substantive categories. If this holds up, it seems to us quite a remarkable finding, suggesting that in this type of format manipulation, the DK filter serves to select people in a way that is essentially random relative to the substantive categories.

However, our third conclusion, though more tentative, is that although education does not interact significantly with form and response for the three items, for two of the three there is a trend for the DK difference between forms to be smallest for the college group. If the three items, dichotomized into DK vs. non-DK responses, are summed to form a "DK Index," then the correlation of DK with education is -.39 on the filtered version and -.26 on the unfiltered. (The difference between the two correlations is significant beyond the .05 level; the same result is obtained testing the corresponding regression coefficients.) Given the size of the marginal effect on DK itself it is certainly possible that distinctive three-way interactive results will emerge from further analysis, but for now we must say that the inclusion or exclusion of a DK filter seems not to change either the ratio of the agree and disagree responses to each other, or their relation to other variables. Other than increasing or decreasing one's estimate of the DK percentage itself, it appears to make no difference whether one screens or not, though it is not certain whether these conclusions will hold for subject matter outside the foreign affairs area.

One incidental finding from these three graded items is the percentage of people willing to say DK on difficult items even when not explicitly encouraged. It is sometimes asserted that people are willing to answer any survey question, but at least on these foreign policy issues that is not the case. On the form which does not screen out DK responses, 15% of the sample nevertheless volunteered DK to the Soviet item, 23% to the Middle East item, and fully 63% to the item on Portugal. We did not expect many people to know much about the Portuguese revolution in the fall of 1974, and more than three-fifths of the sample were honest enough to admit this even on the form that did not encourage such admission.

Index Construction

We would like to address one more issue in this paper; that of whether index construction, at least in a typical form in which it occurs, necessarily minimizes the interactive problems that constitute our primary concern here. In addition to the four main types of items described thus far, two more miscellaneous items were taken from Stouffer's well known 1955 study of Communism, Conformity, and Civil Liberties. Although Stouffer was a master craftsman in the construction of survey questions, we noticed that the items shown on the left side of Table 6 were worded in a way that might discourage civil libertarian responses, much as the "not allow" form in the example with which we opened this paper. We therefore wrote slightly different versions of the two items which leaned, we think not unfairly, in the more libertarian direction (See Table 6, right side). The two original Stouffer items were placed in the same form as the item on "allowing" speeches against democracy; the two amended items in the form with the "forbid" speeches item. By scoring each item as 1 or 2, and adding these scores, we created a brief "Civil Liberties Index" for each form.[10] Let us call the first form the "hard form," the second the "easy form," in terms of encouraging a libertarian position. The cor-

TABLE 6

Two Civil Liberties Items Based on Stouffer (1955)

Original Stouffer Items	Our Variations
This next question is about a man who admits he is a Communist. Suppose he wrote a book which is in your public library. Somebody in your community suggests the book should be removed form the library. Would you favor removing the book, or not?	This next question is about a man who admits he is a Communist. Suppose he wrote a book which is in your public library. Somebody in your community suggests the book should be removed from the library. Somebody else in your community says this is a free country and it should be allowed to remain. Would you favor removing the book, or not?
There are always some people whose ideas are considered bad or dangerous by other people. For instance somebody who is against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed to speak, or not?	There are some people who are against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed the freedom to speak, or not?

relation of this index with education is .34 for the easy or facilitative form, .51 for the hard. Although the two correlations are in the same direction, they differ significantly ($p < .01$), and education can be said to account for more than twice the variance in index scores for the hard form than for the easy one. The difference between the two regression coefficients is also significant beyond the .01 level. The results item by item are somewhat more complex than this summary indicates, but the conclusion does seem sound that the general problem we address in this paper is not one that can always be finessed by mechanical resort to index construction. This is by no means to argue against index construction -- on the contrary, our work with individual items makes it clear that these are sometimes very unstable -- but simply to note that many of the simple additive indices used in social research may be subject to cumulative biases of the kind dealt with in this research.

Summary

We have touched in an illustrative and summary way on many, though not all, of our results to date. Omitted entirely are results from reinterviews with our main 1974 sample which allow us to look at "who" shifts in response to format variation. There are also some selected data on interviewer effects and on sequence effects, and we have several other independent measures of respondent interest and information which can be brought to bear on all the data. We also plan at least one more set of experiments, mainly to replicate effects of borderline reliability reported here, so that we can be more certain that none is due to sampling error.

For now we have tried to indicate the general types of problems under investigation. Clearly it is possible in almost every instance we tested to change item marginals to a reliable extent, though we should also add that most substantive changes in marginals were under 15% and the average was closer to 5%. Many items seem more robust in terms of resisting three-way interactive effects with education, but for agree-type items and probably for some involving changes in tone of word, interactions occur such that one would draw different conclusions from different forms of what can reasonably be thought of as the same basic item. For other format variations we are still uncertain whether important interactions occur, but there is some evidence that formats which attempt to screen out respondents without strong opinions may tend to accentuate answering in terms of response sets. More generally, our evidence points to low educated respondents as a group particularly vulnerable to question effects of all sorts.

Footnotes

1. Some fairly explicit statements of the assumption are also found in Stouffer, et al. (1949) and Kendall & Lazarsfeld (1950). Most text books we have looked at simply have nothing to say on the issue one way or the other.
2. One further reason for the cessation of work in this area may have had to do with the interests of the commercial pollers who provided the funds for many of the early investigations. It is hard to see how the continued display of question effects could be to the advantage of firms that encourage reliance on single-item distributions.
3. Two experiments were also carried out within the 1974 NORC General Social Survey, with the generous assistance of its Director, James A. Davis.
4. Randomization was done systematically at the DU level. Since in some experiments we compared three forms of a question, we created three equal sized subsamples of about 500 each. Where only two forms were used, as in the forbid-allow experiment, one was allocated to two of the subsamples, the third to the remaining one. Close attention was paid to the grouping and sequencing of items administered to each subsample. Irene Hess, Head of the Sampling Section, constructed the subsampling design.
5. Table 1 may be seen as presenting a three-way interaction involving question wording, response, and time. Lacking the 1940 frequencies, we cannot test the interaction for significance but because the differences by form are so similar in the two years, the main finding appears to be constancy of form effect over time.
6. Significance levels reported in this paper must be regarded as approximations. On the one hand, we have used SRS tests, even though the national samples involve some clustering. On the other hand, we generally employ chi square tests insensitive to the ordering that often appears in these data. Quite likely neither source of imprecision is as important as the fact that we have necessarily run many tests on these data and have a fair number of results of border line significance. For these reasons we plan, and have already employed to some extent, replication as an essential part of this project. In most analysis involving education, the education variable was used both in the three-category form shown here and in a five-category (0-8, 9-11, 12, 13-15, 16+) version. In the experiments generally, the trends for both are similar, though significance levels for the two sometimes differ. The five-category version shows a few statistically significant results difficult to interpret, and for the present we have thought it useful to keep base N's large by restricting our main focus to three categories.
7. See Lenski and Leggett (1960), Carr (1971), and Jackman (1973). This area is the one exception to our earlier statement that research into format effects ceased in the early 1950's.
8. This evolution in wording is particularly apparent in Gallup questions, as shown in another of our project papers: Jean M. Converse, "A Content Analysis of National Survey Questions Across Organizations and Over Time," unpublished paper, SRC, 1975.
9. For the experiments on agree-disagree items, the 2nd form shown here served as the contrast with a third form that employed forced-choice

format.

10. The average inter-item correlation on both forms was about .45.

REFERENCES

Cantril, Hadley
1944 Gauging Public Opinion. Princeton: Princeton University Press.

Carr, Leslie G.
1971 "The Srole items and acquiesence." American Sociological Review, 36 (April): 287-293.

Davis, James A.
1971 Elementary Survey Analysis. Englewood Cliffs, New Jersey: Prentice Hall.

Jackman, Mary R.
1973 "Education and prejudice or education and response-set?" American Sociological Review, 38 (June): 327-339.

Kendall, Patricia L. and Lazarsfeld, Paul F.
1950 "Problems of survey analysis." pp. 176-79 in R.K. Merton and P.F. Lazarsfeld (eds.) Continuities in Social Research: Studies in the Scope and Method of "The American Soldier." Chicago: Free Press.

Lenski, Gerhard E. and Leggett, John C.
1960 "Caste, class and deference in the research interview." American Journal of Sociology, 65 (March): 463-467.

McNemar, Quinn
1946 "Opinion-attitude methodology." Psychological Bulletin, 43 (July): 289-374.

Payne, Stanley L.
1951 The Art of Asking Questions. Princeton: Princeton University Press.

Rugg, Donald
1941 "Experiments in wording questions: II." Public Opinion Quarterly, 5 (March): 91-92.

Schuman, Howard and Duncan, Otis Dudley
1974 "Questions about attitude survey questions." pp. 232-251 in Herbert L. Costner (ed.), Sociological Methodology 1973-1974, San Francisco: Jossey-Bass.

Stouffer, Samuel A.
1955 Communism, Conformity, and Civil Liberties. Garden City, New York: Doubleday.

Stouffer, Samuel A. et al.
1949 The American Soldier: Adjustment During Army Life. Princeton: Princeton University Press.

ACKNOWLEDGEMENT

We are indebted to Jean M. Converse who worked with us in the development of the experiments described here, and to Jane Fountain for statistical and computing advice. This research has been supported by grants from the National Institute of Mental Health (MH 24266) and the National Science Foundation (GS-39780) to Howard Schuman. Our names are listed alphabetically.