

This paper renews the line of research into the effects of changes in survey question wording and form which occupied researchers during the 1940s. We suggest two reasons for the cessation of such research: the idiosyncratic nature of many of the items experimented with and the near exclusive focus on single-variable distributions. In the present study, the experiments are based on decisions that face all survey investigators: whether to use agree-disagree statements or forced choice items; whether to ask open or closed questions; whether and how to balance alternatives offered; whether to include a middle alternative; and whether or not to filter for no opinion. Furthermore, we examine the consequences of these decisions not only for univariate distributions but also for an item's relationship to education. The results from SRC national probability samples suggest that for the first two types, as well as for items involving variations in tone of word, the decisions may affect inferences about correlations with education. For the other three types the effects are restricted mainly to changes in marginals, although the no-opinion type shows a more limited kind of interaction with education. Finally, we present evidence that index construction is not an adequate solution to the question-wording problem.

QUESTION WORDING AS AN INDEPENDENT VARIABLE IN SURVEY ANALYSIS

HOWARD SCHUMAN
STANLEY PRESSER

The University of Michigan

during the 1940's a number of experiments on attitude question wording and form 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 experi-

AUTHORS' NOTE: *We are indebted to Jean M. Converse who worked with us in the development of the experiments described here, to Jane Fountain for statistical and computing advice, and to Otis Dudley Duncan and William M. Mason for advice on several important technical points. We alone are responsible for any errors. This research has been supported by grants from the National Institute of Mental Health (MH 24266) and the National Science Foundation (GS-39780 and Soc 76-15040). An earlier version of the paper was presented at the American Statistical Association meetings, Atlanta, August 1975.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 6 No. 2, November 1977
© 1977 Sage Publications, Inc.

ments became most widely known through the Cantril et al. (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?" Approximately 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 early 1950s such question-wording experiments had largely disappeared from major surveys. We think there were several reasons for their demise. 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 almost all these early reports of experiments were restricted to univariate results, while survey *analysis* is typically concerned with relationships between variables. The assumption developed among investigators that marginals cannot be trusted owing to question wording uncertainties, but that associations between variables are not subject to this same sort of instability. Exactly this assumption is made, for example, by Stouffer and DeVinney in *The American Soldier* (1949: 168):

one must be careful to focus attention on *differences* in percentages among different categories of men with favorable attitudes on a given item, not on absolute percentages. The fact that the percentages saying the Army was run pretty well or very well are large does not mean, necessarily, that so many men were actually favorable to the Army—such percentages are artifacts of question wording and of the check-list categories arbitrarily selected as 'favorable' But when we focus on *differences* in percentages responding favorably to the *same* questions, among men in different categories, the differences can be meaningful in a sense in which the absolute values cannot. [emphasis in original]

What Stouffer and DeVinney state explicitly appears, in the absence of warnings to the contrary, to have been assumed by many other survey methodologists and practitioners.¹

The assumption of "form-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* form effects is rarely spelled out.² 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 lay in the ad hoc character of most of the early work. Even in terms of univariate results, larger theoretical issues of question construction and typology were seldom addressed, and there was little 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.³

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 types of survey questions (e.g., open versus closed), 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 form-resistant correlations, an attempt which grew to include development of a typology of question forms, as well as treatment of some related issues.

Two general hypotheses underlie this research. First, where question wording alters marginals appreciably, it seems unlikely from a theoret-

ical standpoint that those persons being affected are simply a random subsample of all respondents. The effects are a kind of self-selection, and self-selection is rarely a randomizing procedure. Second, and more specifically, those affected by form should usually be the less educated, as well as the less interested or involved in the particular issues asked about. For the present paper, we concentrate on education because it is a fundamental variable in almost all survey analysis, and its interaction with question form would mean that one would draw different conclusions about the relation of education to opinion depending on which question form had been asked.

Our initial reasoning was that poorly educated respondents should be more easily swayed by emotionally toned words or by the presence or absence of a response alternative. Better educated respondents, on the other hand, should more easily grasp the general point of a question and not be as easily affected by emotionally colored words or by the degree to which implications of a response are spelled out. Furthermore, it seemed likely that more educated respondents would feel less deferential toward the interviewer and her questionnaire, and therefore more willing to insist on giving a sensible alternative answer (e.g., a midpoint between two extremes) even where it is not provided by the question frame. For these reasons, plus the general importance of years of schooling as both a cognitive and a status indicator, this seemed to us a strategic starting point for an analysis of question-form and wording effects.⁴

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.⁵ 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 34 years. The 21% difference in 1940 is paralleled by 16% in 1974, and thus the change over time is similar for both forms.⁶ While we have no direct evidence as to the cause of the form effect, one possibility is 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.

TABLE 1
Comparison of Forbid and Allow Marginals in 1940 and 1974

<u>Allow Form</u>			<u>Forbid Form</u>		
Do you think the United States should allow public speeches against democracy?			Do you think the United States should forbid public speeches against democracy?		
	<u>1940</u>	<u>1974</u>		<u>1940</u>	<u>1974</u>
1. Yes (Allow)	25%	56%	2. No (Not Forbid)	46%	72%
2. No (Not Allow)	<u>75%</u>	<u>44%</u>	1. Yes (Forbid)	<u>54%</u>	<u>28%</u>
	100%	100%		100%	100%
N	(*)	(494)	N	(*)	(936)

Response by Form, 1974 data only: $\chi^2_1 = 35.75, p < .001$

*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 shows the relation between response and education for the two question forms. In line with our hypothesis, form seems to make least difference (6%) for those who have been to college, but has a substantial impact (26%) on those with zero to eleven years of schooling; high school graduates fall in between. Using Goodman's (1971) method for analyzing multiway-contingency tables, the likelihood-ratio χ^2 for the three-way interaction is 5.75, $p < .06$.⁷ On replication in 1976, this three-way interaction reaches significance at beyond .02—the college-educated again showing the least effect—but the linear trend is not obtained, high-school graduates and those without high-school diplomas being affected about equally. Thus, tone of word makes a difference here 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 larger gap in connotation or tone of word would remove entirely the association with education on one form, thus changing conclusions about type as well as degree of relationship to this basic background variable.

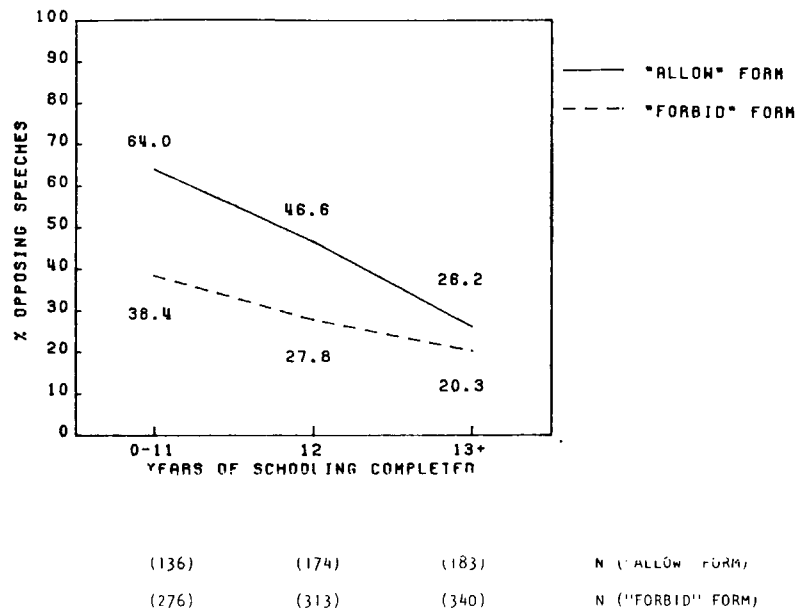


Figure 1: Percent Opposing Free Speech by Education and Form, 1974

These findings point to 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 assume that index construction would remove just this type of idiosyncratic effect, we shall also raise some questions about that strategy in a later section of this paper. 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 form, 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 usually are made on the basis of rule-of-thumb, personal preference, convenience, or simply chance. This led us to formulate five 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 five types. Note that in constructing examples of each question type, we started, wherever possible, from items used in past national surveys, in order to increase the external validity of our results.

(1) *Agree-Disagree versus Forced Choice*. Speed and convenience of administration frequently seem to recommend the use of agree-disagree items in questionnaires and interviews. But some past research, as well as a certain intuitive regard for fairness in presentation of issues, suggests the desirability of providing forced-choice forms, 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 form differences in marginals significant beyond the .05 level, and two of the five provide significant three-way interactions with education, as illustrated in Table 2.

The item on crime presented in Table 2 involves three forms: two separate agree-disagree statements that are logical contraries (A & B) and a forced choice version (C). Responses to both the first agree-disagree form and the forced choice form show a significant relationship to education, but those on the other agree-disagree form show no such relationship. Thus one would draw different conclusions about the *existence* of a relationship to education in this case, not only about its magnitude. The response by education by form interaction reaches significance in the comparison of the two agree-disagree forms, and is of borderline significance in the B-C comparison. Moreover, in the latter case, the three-way interaction with education occurs despite almost identical marginals! One would not have had even the initial clue of a two-way difference (response by form) to alert one to the possibility of this important interaction.

Although in only one of the other experiments (women in politics) does the three-way interaction with education reach significance, in three of the remaining four the variation in response by form is greatest for the least educated just as it is in the crime example. Thus, overall, the assumption of form-resistant correlations is not well supported with respect to these agree-disagree items.

TABLE 2
Agree-Disagree Example
Percent Saying Individuals Are to Blame for Crime
by Education and Form*

		<u>Total</u>	<u>Education</u>		
			<u>0-11</u>	<u>12</u>	<u>13+</u>
A. Please tell me whether you agree or disagree with this statement: " <u>Individuals</u> are more to blame than <u>social conditions</u> for crime and lawlessness in this country." (Do you agree or disagree?)	% Agreeing	59.6	67.7	62.0	51.7
	N	(473)	(133)	(163)	(176)
		$\chi^2_2 = 8.55 \quad p < .02$			
B. Please tell me whether you agree or disagree with this statement: " <u>Social conditions</u> are more to blame than <u>individuals</u> for crime and lawlessness in this country." (Do you agree or disagree?)	% Disagreeing	43.2	41.9	41.8	45.3
	N	(472)	(136)	(153)	(181)
		$\chi^2_2 = 0.54 \quad n.s.$			
C. Which in your opinion is more to blame for crime and lawlessness in this country-- <u>individuals</u> or <u>social conditions</u> ?	% Individuals	46.4	55.3	44.5	40.5
	N	(448)	(141)	(155)	(148)
		$\chi^2_2 = 6.76 \quad p < .05$			
		<u>A&B</u>	<u>A&C</u>	<u>B&C</u>	
Response by Form: $\chi^2_1 =$		25.55	16.12	0.96	
		p < .001	p < .001	n.s.	
Response by Education by Form: $\chi^2_2 =$		6.64	0.67	4.98	
		p < .05	n.s.	p < .10	

*Form C is a slight modification of a Gallup question. See "The Gallup Opinion Index," Report No. 65, November 1970, page 15.

(2) *Formal versus 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

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 (455)	68.9 (119)	70.8 (168)	72.7 (165)
		$\chi^2_2 = 0.49, n.s.$		
B. Would you favor <u>or</u> oppose a law which would require a person to obtain a police permit before he could buy a gun?	71.7 (455)	69.9 (103)	71.6 (162)	72.6 (179)
		$\chi^2_2 = 0.24, n.s.$		
C. Would you favor a law which would require a person to obtain a police permit before he could buy a gun, <u>or</u> do you think such a law would interfere too much with the right of citizens to own guns?	67.3 (431)	62.5 (96)	63.9 (166)	73.4 (169)
		$\chi^2_2 = 4.80, p < .10$		

Response by Form (A and B): $\chi^2_1 = 0.05, n.s.$

Response by Form (B and C): $\chi^2_1 = 2.00, n.s.$

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.

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," and our hypothesis was that it would have little or no effect, since the original form is a question rather than an assertion and negative answers are 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 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 one case, there is a significant difference in 1974 but upon replicating that experiment in 1977 we obtained no difference at all, hence the original finding is probably best regarded as a chance event.) 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 nonsignificant exception being the gun control item. (In a replication of the gun control experiment the substantive effect did reach significance. See Schuman and Presser, 1977, for a detailed examination of those results.) 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 versus forced-choice problem, we cannot reject the null hypothesis for three-way form effects, at least in relation to education.

(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 four items that we varied in this way, all show significant differences in the middle-category percentage, but in no case does the division of opinion between the polar positions differ significantly by form. This is particularly striking because in one instance the middle alternative rises by almost 40% and yet the ratio between the other positions is unaffected.

Turning to the results by education, the form-resistant correlation assumption fares quite well for this type of item. In no case does the response by education by form interaction approach significance. This is so whether one collapses the polar positions and compares them with the middle alternative or simply compares the polar positions, excluding the middle alternative.

(4) *Opinion Screening Filters*. It is well known that on many issues a large fraction of the public has no opinion, for reasons of lack of information or interest. In asking survey questions, one can attempt first to screen out those who admit having no opinion, as the Institute for

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	<u>11.1</u>	<u>12.8</u>	<u>9.2</u>	<u>11.3</u>
	100 (882)	100 (250)	100 (294)	100 (335)

$$\chi^2_{14} = 9.95, p < .05$$

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 government, 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	<u>9.2</u>	<u>7.9</u>	<u>10.1</u>	<u>9.8</u>
	100 (434)	100 (139)	100 (148)	100 (143)

$$\chi^2_{14} = 13.65, p < .01$$

Response by form (collapsing "too much" and "not enough"): $\chi^2_1 = 22.62, p < .001$

Response by form (excluding "right amount"): $\chi^2_1 = 0.04, n.s.$

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

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

Social Research election studies have done for many years. This step has the seeming merit of reducing the number of what Converse (1970) has called nonattitudes. 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.

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?

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

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

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."

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

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

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

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

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

*Constructed for this experiment.

To study this problem we constructed three items dealing with foreign affairs, intentionally varying the presumed familiarity of the issue for respondents. One item dealt with the Soviet Union, one with the Middle East, and one with the 1974 revolution in Portugal which had occurred just prior to the survey. We took this to represent an ascend-

ing order of public ignorance. The two forms that comprised this set of experiments are illustrated in Table 5 by the pair of questions about the Soviet leaders.

Several conclusions can be drawn from these three experiments. First, it clearly is possible to increase substantially the percentage of "don't know" (DK) responses 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 is no significant univariate difference between the two forms. In other words, the ratio of the agree to disagree responses is very similar across form in each experiment, despite the shift of approximately a quarter of the respondents between DK and substantive categories. Third, omitting DK responses, substantive distributions do not differ in their relation to education between the two forms. That is, the response (agree versus disagree) by education by form interaction does not approach significance for any of the three foreign policy items.

It is also of some interest to collapse the substantive responses into a single category ("opinion") as against the DK or no opinion category. When this is done, the three-way interaction (education by form by opinion versus DK) is not significant for any of the three items, but for two of the items (the Middle East and Portugal) the linear component of the interaction reaches borderline levels ($p = .10$).⁹ Moreover, if each item is dichotomized into DK versus all nonDK responses, and the three recoded items 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, and the same significance level is obtained for the corresponding regression coefficients.) Thus, education is more highly correlated (negatively) with saying DK when the latter is explicitly offered than when it is merely accepted if volunteered. This suggests that in cases where education is very strongly related to opinion position (which is not so in the present examples), filtering, by changing the educational distribution of those who give an opinion, would alter the marginal split between the substantive positions.

One incidental finding from these three graded items is the percentage of people willing to say DK to difficult questions even when not explicitly encouraged. It is sometimes asserted that people are willing to answer *any* survey question, but this is not the case at least on these foreign policy issues. 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 have an opinion about the Portuguese revolution in the fall of 1974, and more than three-fifths of the sample were willing to admit this even on the form that did not encourage such admission.

(5) *Open versus Closed Questions.* The oldest controversy over question form involves the issue of open versus closed questions. On the one hand, the open form does not limit respondents to alternatives within the investigator's frame of reference, and it also avoids suggesting or imposing answers the respondent may not have considered. On the other hand, the closed form restricts responses to those germane to the researcher's aims, and provides data in a form that is a great deal easier to code and analyze. The issue at hand, however, is whether and when the two forms yield basically the same results. Given the antiquity of the controversy (see Lazarsfeld, 1944, for an early but still widely read statement), the general absence of split-ballot comparisons of the two forms is remarkable.

We designed three open-closed experiments, but have analyzed only one thus far for presentation here. The experiment actually involved two stages. First, the standard closed form of a question on work values (see Lenski, 1963, and the NORC General Social Surveys) was asked of a random half of the 1976 Detroit Area Study (DAS), and a comparable open version was asked of the other half. This produced a number of sizeable differences between the forms, but these were difficult to interpret because it is possible they were due simply to defects in the specific closed alternatives employed. The latter may never have been properly developed to correspond to open responses (as Lazarsfeld, 1944, recommended), or they may simply have become out of date since the closed question was first devised in the late 1950s. In order to remedy this, we reconstructed the closed question to provide categories and wording more in keeping with the actual responses obtained in the 1976 DAS open form. This new version of the closed question, shown in Table 6, was then used with a parallel open form in a split-ballot experiment in a 1977 SRC national telephone survey.¹⁰

Although there are a number of significant differences in the results obtained on the two forms in 1977, for reasons of space we focus here only on one of the more interesting ones. As may be seen in Table 6, almost three times as many people (21%) choose the answer "work that

is steady with little chance of being laid off" on the closed version as give a comparable answer on the open form (8%). Apparently a fair number of people think of this as an important job attribute when reminded of it, yet do not think of it unaided. Moreover, the form difference is related to education. The poorly educated are most likely to choose the job security category on the closed version, but least likely to choose it on the open. We are unable to tell which form (if either) provides the more valid picture of the relation between education and concern with security, but clearly one would draw different conclusions depending on the form used. Since for other alternatives in this experiment the two forms lead to similar conclusions (e.g., the higher educated are more likely to emphasize "feeling of accomplishment" as most important on both forms), open-closed differences are by no means inevitable and an analyst should probably feel most comfortable when the same result can be obtained regardless of form.

Further analysis and presentation of this and other open-closed experiments is deferred until a later report. For now, it is apparent that for at least some important comparisons, the assumption of form-resistant correlations must be rejected when open and closed versions of the same basic item are under consideration.

INDEX CONSTRUCTION

We would like to address one more issue in this paper: 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 five main types of items described thus far, two more miscellaneous items were taken from Stouffer's well-known study, *Communism, Conformity, and Civil Liberties* (1955). Although Stouffer was a master craftsman in the construction of survey questions, we noticed that the items shown on the left side of Table 7 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 7, 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

TABLE 6
Open-Closed Example
Percent Saying Security is Most Important by
Education and Form*

	Total	Education		
		0-11	12	13+
This next question is on the subject of work. People look for different things in a job. Which one of the following five things would you <u>most</u> prefer in a job? Work that pays well; Or work that gives a feeling of accomplishment; Or work where there is not too much supervision and you make most decisions yourself; Or work that is pleasant and where the other people are nice to work with; Or work that is steady with little chance of being laid off.	21.0 (590)	32.0 (128)	21.0 (210)	14.8 (244)
		$\chi^2 = 14.67 \quad p < .001$		
This next question is on the subject of work. People look for different things in a job. What would you <u>most</u> prefer in a job?	7.7 (469)	3.6 (84)	13.2 (152)	5.6 (233)
		$\chi^2 = 9.44 \quad p < .01$		
		$\chi^2_1 = 29.57 \quad p < .001$		
		$\chi^2_2 = 10.98 \quad p < .01$		

*Percents represent those who chose the category "work that is steady with little chance of being laid off" on the closed version and those who gave a comparable answer on the open version.

"civil liberties index" for each form.¹¹ Let us call the first form the "hard form," the second the "easy form," in terms of encouraging a libertarian position. The correlation 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).¹² Thus it appears that the general problem we address in this paper is not one that can always be finessed by mechanical resort to index construction. This

TABLE 7
Two Civil Liberties Items Based on Stouffer (1955)

<u>Original Stouffer Items</u>	<u>Our Variations</u>
<p>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. Would you favor removing the book, or not?</p>	<p>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?</p>
<p>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?</p>	<p>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?</p>

is not necessarily to argue against index construction—on the contrary, our work with individual items makes it clear that they 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 study.

SUMMARY

Overall, it appears fairly easy to change item marginals to a reliable extent, although we should add that most substantive changes in marginals were under 15% and the average was closer to 5%. For agree-type items, open as against closed items, and for some items involving changes in tone of word, three-way interactions with education occur such that one would draw different conclusions from different forms of what can reasonably be thought of as the same basic question. Thus for these types of question, the assumption of form-resistant correlations appears to be a poor one. The assumption may be more justified, however, with two other types of items discussed: formal versus substantive

balance and middle alternatives. We are unable to show that inferences about the relationship of education to opinion differ reliably by form in these cases. Finally, the no-opinion filter type occupies a middle ground. Education interacts with form to affect the don't know responses, but for the items used in our experiments, once DKs are excluded from analysis, form does not affect the relationship between education and substantive opinion.

Even where education is unrelated to form effects, this does not mean that such effects are random in nature. For example, in the no-opinion area we have discovered that the relations among attitudes on our experimental items differs by form. In the middle alternative area, there is evidence that intensity is related to form effect, people who say they feel strongly about their opinion showing much less difference between forms than those who register weaker intensity. Thus, while additional experiments are needed to understand the extent, causes, and differential validity of results obtained when the same question is asked in two or more ways, our work suggests that form effects occur with enough frequency so that researchers need to be wary of correlations based entirely on a single question form.

NOTES

1. More recent statements of the assumption may be found in Davis (1971: 20; and 1976: 37). Most textbooks we have looked at simply have nothing to say on the issue one way or the other.

2. An important exception is the work done during the 1940s by Guttman and Suchman (in Stouffer et al., 1950). However, their solution requires a very strong scaling model which is frequently not satisfied. For a brief critique of their work see Presser (1977: ch. 1).

3. 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.

4. There is also some research showing education to be a significant interactive variable with respect to one question-form effect, agreeing-response set (see, for example, Lenski and Leggett, 1960; and Jackman, 1973). (This area is the one exception to our earlier statement that research into form effects ceased in the early 1950s.) At the same time, it should be recognized that years of schooling might promote interaction of an opposite sort, where subtle logical features of a question are noted only by the best educated (see Schuman and Harding, 1964).

5. Randomization was done systematically at the housing unit level. (Irene Hess, head of the SRC Sampling Section, constructed the subsampling design.) Since in some experiments we wished to compare 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.

6. 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.

7. Significance levels reported in this paper must be regarded as approximations since we have used SRS tests, even though the SRC national samples involve some clustering. We computed more exact sampling errors (taking into account the clustering) for all the bivariate results reported and in no case did this change an inference.

8. This evolution in wording is particularly apparent in Gallup questions, as shown in another of our project papers: Converse, "A Content Analysis of National Survey Questions Across Organizations and Over Time" (unpublished, SRC, 1975).

9. For all these analyses we assume a logistic response model, as employed in the Goodman procedures. It may be noted that the Russia item shown in Table 5 reveals large variations in the DK percentage differences by form among the three educational groups: 31.3%, 24.4%, and 14.0% for the 0-11, 12, and 13 + groups, respectively. If a linear probability model is assumed, the response (opinion versus DK) by education by form interaction turns out to be significant ($\chi^2 = 7.92$, $p < .02$ as computed by H. M. Kritzer's program, NONMET). Thus computations based on the two different models yield quite different conclusions, presumably because of floor effects on DK responses for the standard form. In accord with what seems to be dominant statistical thinking, we have based our main conclusions on results with the Goodman procedure and thus the logistic model, though it might be argued that floor effects have meaningful substantive implications in these no opinion experiments.

10. There were actually five versions of the closed item varying randomly the order of the response alternatives. Since the distributions are not significantly affected by this factor, we have collapsed over the five.

11. The average interitem correlation on each form was about .45.

12. The same results were obtained when we replicated the three experiments in 1976: the difference between the two correlations (.38 and .50) is significant at the .02 level.

REFERENCES

- CANTRIL, H. et al. (1944) *Gauging Public Opinion*. Princeton: Princeton Univ. Press.
- CONVERSE, P. E. (1970) "Attitudes and non-attitudes; continuation of a dialogue," pp. 168-189 in E. R. Tufté (ed.) *The Quantitative Analysis of Social Problems*. Reading MA: Addison-Wesley.
- DAVIS, J. A. (1971) *Elementary Survey Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- (1976) "Are surveys any good, and if so, for what?" pp. 32-38 in H. W. Sinaiko and L.A. Broedling (eds.), *Perspectives On Attitude Assessment: Surveys and Their Alternatives*. Champaign, IL: Pendleton.

[170] SOCIOLOGICAL METHODS & RESEARCH

- GOODMAN, L. (1971) "The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classification," *Technometrics* 12: 33-61.
- JACKMAN, M. R. (1973) "Education and prejudice or education and response-set?" *American Soc. Rev.* 38: 327-339.
- LAZARSFELD, P. F. (1944) "The controversy over detailed interviews: an offer for negotiations." *Public Opinion Q.* 8: 38-60.
- LENSKI, G. (1963) *The Religious Factor*. Garden City, NY: Anchor.
- and J. C. LEGGETT (1960) "Caste, class and deference in the research interview." *Amer. J. of Sociology* 65: 463-467.
- McNEMAR, Q. (1946) "Opinion-attitude methodology." *Psych. Bul.* 43: 289-374.
- PAYNE, S. L. (1951) *The Art of Asking Questions*. Princeton: Princeton Univ. Press.
- PRESSER, S. (1977) "Survey Question Wording and Attitudes in the General Public." Ph.D. dissertation, University of Michigan (unpublished).
- RUGG, D. (1941) "Experiments in wording questions: II." *Public Opinion Q.* 5: 91-92.
- SCHUMAN, H. and O. D. DUNCAN (1974) "Questions about attitude survey questions," pp. 232-251 in H. L. Costner (ed.) *Sociological Methodology 1973-1974*. San Francisco: Jossey-Bass.
- SCHUMAN, H. and J. HARDING (1964) "Prejudice and the norm of rationality." *Sociometry* 27: 353-371.
- SCHUMAN, H. and S. PRESSER (1977) "Public opinion and legislative inaction on gun registration." *Public Opinion Q.* (forthcoming).
- STOUFFER, S. A. (1955) *Communism, Conformity, and Civil Liberties*. Garden City, NY: Doubleday.
- STOUFFER, S. A., et al. (1950) *Measurement and Prediction*. Princeton: Princeton Univ. Press.
- STOUFFER, S. A., et al. (1949) *The American Soldier: Adjustment During Army Life*. Princeton: Princeton Univ. Press.

Howard Schuman is Professor of Sociology and Program Director in the Survey Research Center, University of Michigan. He is presently Editor of Sociometry and has previously served as Director of the Detroit Area Study at Michigan.

Stanley Presser is a Research Investigator in the Survey Research Center, University of Michigan.