

## A General Social Survey Experiment in Generic Words

NORA CATE SCHAEFFER

In *Class and Conformity* (1969), Mel Kohn used a series of items to investigate the values parents felt were important for their children. His study identified a particular child within each family. To obtain parental norms, he asked parents to evaluate the general desirability of 13 traits for a child of the same age and sex as the identified child. Kohn reported that, after controlling for social class, fathers evaluated some of the traits differently depending on the age and sex of the child being considered (1969:54-55).

This finding suggests that an adaptation of these items that influenced respondents to consider a child of a specific age or sex might affect the distributional characteristics of the items. This could be true if (a) a significant number of respondents were sensitive to the variation in the item wording, (b) the variation were interpreted consistently so that the age or sex of the child considered when formulating an answer were fixed, and (c) the way in which the traits were

**Abstract** This paper analyzes a question wording experiment from the 1980 General Social Survey. The experiment used two variants of a series of items that asked for evaluations of qualities for children. The experimental variation used the noun "child" consistently; the standard variation used the pronoun "he" and the noun "child." There were no differences between the two wordings in the way the traits were evaluated. The lack of a question wording effect can be partly attributed to the fact that both wordings are biased in favor of males. Question wording interacted with respondent sex in affecting the sex of the child that respondents claimed to be considering when answering the trait items.

Nora Cate Schaeffer is an Assistant Survey Director, National Opinion Research Center. This research was done for the General Social Survey project, which is under the direction of James A. Davis and is supported by the National Science Foundation, SOC77-03729. The author is grateful to the General Social Survey for including this experiment; to Tom Smith, Jan Dunham, Carol B. Stocking, and John Loft, all of the National Opinion Research Center, for useful comments on earlier drafts of this paper; and to Chris Flinn and Sue Shott for statistical advice.

evaluated actually depended on the age or sex of the child the respondent considered when answering. For example, if the item wording suggested to most respondents that they answer about boys rather than girls, but there were no differences in the way the traits were evaluated for boys and girls, there would be no effect on the item distribution. On the other hand, if a trait were evaluated very differently for boys and girls, the item distributions might be affected, even if only a few respondents were systematically influenced by a wording change.

The trait evaluation items were adapted for use in the General Social Survey (GSS) beginning in 1973. In the GSS, the interviewer hands the respondent a card with the following list (the label which will identify each item in the following tables appears in parentheses):

- That he has good manners (MANNERS)
- That he tries hard to succeed (SUCCESS)
- That he is honest (HONESTY)
- That he is neat and clean (CLEANLINESS)
- That he has good sense and sound judgment (JUDGMENT)
- That he has self-control (SELF-CONTROL)
- That he acts like a boy (she acts like a girl) (ROLE BEHAVIOR)
- That he gets along well with other children (AMIABILITY)
- That he obeys his parents well (OBEDIENCE)
- That he is responsible (RESPONSIBILITY)
- That he is considerate of others (CONSIDERATION)
- That he is interested in how and why things happen (INTEREST)
- That he is a good student (STUDIOUS)

The interviewer asks the respondent four questions about the list (the emphasis in these questions appears in the instrument):

- Which *three* qualities listed on this card would you say are the *most desirable* for a *child* to have?
- Which *one* of these *three* is the *most* desirable of all?
- All of the qualities listed on this card may be desirable, but could you tell me which *three* you consider *least important*?
- And which *one* of these three is *least important* of all?

In Kohn's study the first question specified that the traits be evaluated for a child of a specified age and sex: "Which three qualities listed on this card would you say were the most desirable for a (boy, girl) of (child's age) to have?" Other than this, there are only minor wording differences between the questions and list used in the GSS and those described by Kohn (1969:257).

In the GSS version of these items, the age of the child to be considered is vague, defined only by the age connotations of the word "child." The question and list together, however,

unintended suggestion about the sex of the child to be considered. There are two possible sources of bias. First, although the word "child" is ostensibly neutral with respect to sex, the dominant (possibly unconscious) category in the minds of respondents, even in the face of grammatical neutrality, may in fact be male. Second, the list of qualities consistently uses the pronoun "he," with a single (parenthetical) exception.

While there is little literature on the cognitive content of words that is directly relevant, a study by Broverman et al. (1970) is worth reviewing briefly. In this study three groups of subjects (mental health professionals) were asked to characterize a healthy adult male, a healthy adult female, and a healthy adult, respectively. The characteristics chosen by the group rating adult women were significantly different from those chosen by the remaining two groups. The latter were not significantly different from each other. This finding clearly supports the possibility that "child" might not be as egalitarian a noun as its definition would lead one to expect. While no similar experiments exploring the properties of the pronoun "he" were found, the fact that this pronoun is used to refer specifically to males, in addition to its generic use, makes it ambiguous and therefore suspect.

In the 1980 GSS, an experiment using two random subsamples was introduced to test whether the use of the pronoun "he" in the list of traits affected respondent evaluations. One-third of the respondents were given the list of traits which had been used in previous surveys, and one-third were given a list which substituted the noun "child" in the statement of each trait, for example, "that a child is honest." (One-third of the sample was used in an experiment on the response categories and is not used in this analysis.) In addition, a follow-up question was included: "When you rated the importance of various qualities for children were you thinking mostly about boys, mostly about girls, or about both boys and girls equally?"

The experiment and follow-up question provide an opportunity to investigate two separate issues: (1) are there mean differences in the trait evaluations given to the two question wordings,<sup>1</sup> and (2) regardless of whether there are differences in responses to the two question wordings, does the experiment provide some understanding of the

<sup>1</sup> The effects of question wording on item variances will not be analyzed. Such an analysis would be inconclusive because of peculiarities in the shape of the distributions of these items, discussed in a later section. However, if in reality traits were evaluated very differently for boys and girls, and if the question wording influenced some respondents to consider a child of a particular sex, it would not be safe to assume that the effect of question wording on item variances was inconsequential.

way people respond to an ostensibly "generic" use of the words "child" and "he" that might be useful in other contexts?

With respect to the first issue there are two possible outcomes, each of which suggests slightly different questions. If there are differences in responses to the two wordings of the question, can they be attributed to the fact that one wording has a slightly stronger suggestion about the sex of the child the respondent should consider? If there are no differences between the answers to the two wordings, the results are harder to evaluate. There are several possibilities, two involving the effectiveness of the suggestion and one involving the way the qualities are evaluated.

If the noun "child" in the question introduction and trait list (whether or not followed by the pronoun "he") evoked a male image in spite of its definitional and grammatical neutrality, there would be no difference in responses to the two question wordings. In this case both question wordings would contain a systematic bias in favor of males and would be indistinguishable. On the other hand, if "child" and "he" were recognized as including both boys and girls equally, there would be no significant differences in responses to the two forms. In this case, respondents might randomly think about boys and girls (or choose one or the other for reasons that are independent of question wording), or take some sort of "average" across sex of child when formulating their answers, regardless of the way in which the traits were presented. Finally, if there is little difference in the way these traits are evaluated for boys and girls in this population, no amount of suggestion would produce a systematic difference between the forms.

With respect to the second general issue, these data provide only a weak test of what the noun "child" and the generic pronoun "he" connote. The follow-up question is useful in trying to understand how people respond to "generic" words, but it may be a difficult question to answer accurately. The question has the serious disadvantage that respondents may not be able to reconstruct their (probably largely unconscious) thought processes exactly, and choose the category "both boys and girls equally" because it seems most fair or reasonable, grammatically accurate, or requires the least effort. There is evidence that respondents selecting the "both" category are significantly older ( $p < .02$ ) and less educated ( $p < .05$ ) than respondents claiming to have thought of "mostly boys" or "mostly girls" (data not shown). Both age and education may reasonably be expected to affect a respondent's ability to answer this question. In addition, there is the substantive question of what it means to consider both sexes "equally" if previous research indicates that a trait is evaluated

differently for males and females. Responses to this question, and particularly the selection of the "both" category, should be regarded cautiously.

### The Data

The 1980 GSS used a full probability sample of the noninstitutionalized, English-speaking population of the continental United States, 18 years of age or older, and had a response rate of 76 percent. The National Opinion Research Center national probability sample used by the GSS is a stratified, multistage area probability sample of clusters of households, areas at successive stages being chosen with probability proportional to size (Davis, 1980). The tests presented in this paper are standard tests assuming simple random sampling. The sample is not weighted in the tables presented.

The items have an ipsative response structure and were coded as follows:

- 1 = the most desirable trait
- 2 = one of the three most desirable traits, but not the most desirable
- 3 = not mentioned
- 4 = one of the least desirable traits, but not the least desirable
- 5 = the least desirable trait

Thus, a higher score indicates a lower desirability rating. (As a result of the ipsative structure of these items, the outcome of the statistical test on the thirteenth item is fixed given the outcome of the first 12 tests.)

The distributions of most of these items are highly skewed. In particular, five traits—honesty, judgment, obedience, responsibility, and consideration—are almost never scored as one of the three least desirable traits. Three items—cleanliness, acting like a boy or girl, and being studious—are rarely mentioned as one of the most desirable. Because of these extreme skews in the item distributions, the items were collapsed for analysis, and tests were performed on the proportion considering a trait desirable, although considerable information is lost in this way.

### Differences Between the Two Question Wordings

If question wording had a systematic effect such that respondents given the standard version of the items considered boys more frequently than girls, while respondents given the experimental version considered boys and girls "equally," it might be expected that the

traits that had been evaluated differently for boys and girls in the past would show a difference by question wording here. Kohn reports that among the fathers he interviewed five traits—success, honesty, cleanliness, sex role behavior, and interest in how things happen—were rated significantly differently ( $p < .05$  or better) depending on the sex of child considered (all but cleanliness were rated more desirable for boys). However, Kohn's published results are of only limited usefulness for evaluating the possibility just mentioned in these data. Although the table in which he presents the relevant findings describes them as "parental" values, the text indicates that they are actually "paternal" values. The GSS sample is heterogeneous by comparison, since it includes fathers, mothers, and nonparents.

Table 1 presents the proportion rating a trait desirable and item standard deviations by question wording for each of the traits. None of the differences are significant at the .05 level, even before adjusting for the number of tests performed.<sup>2</sup> In 5 of the 13 cases, the proportion citing a trait as desirable is larger in the standard version, but the differences are quite small. (In four cases, the variance is larger in the standard form, but again the differences are small.) The traits for which Kohn found sex-of-child effects do not stand out in any way. Ignoring the information in the tails of the distributions of the item responses, there is no difference in the way respondents evaluate these traits when they are presented using the combination "child" and "he" and when they are presented using "child" alone. Apparently either both words refer to boys and girls in approximately the same way or there are no systematic differences in the way the traits are evaluated for boys and girls in these two heterogeneous groups.

### QUESTION WORDING AND SEX OF CHILD CONSIDERED

Although there were no mean differences in the evaluations of the traits between the two versions of the question, the follow-up question provides a limited opportunity to examine whether respondents considered boys or girls when making their evaluations. There is a significant departure from independence in Table 2 ( $p < .01$ ). While the data indicate that respondents overwhelmingly claimed to be thinking about both boys and girls equally, this category is selected by 9 percent more of the subsample who were given the experimental wording than the subsample given the standard wording. This finding cannot be taken at face value because of the ambiguity of the "both"

<sup>2</sup> A multivariate test on 12 of these items might have revealed a common significance not evident in the items tested singly. The size of the differences among the items, however, is generally small.

Table 1. Proportion Citing a Trait as Desirable, by Question Wording: 1980 GSS<sup>a</sup>

Trait	Question Wording	Mean	Standard Deviation	t	Probability
Manners	Experimental	.290	.454	1.78	.08
	Standard	.239	.427		
Success	Experimental	.187	.390	.96	.34
	Standard	.164	.370		
Honesty	Experimental	.634	.482	-1.19	.24
	Standard	.671	.470		
Cleanliness	Experimental	.060	.237	-.12	.91
	Standard	.061	.240		
Judgment	Experimental	.366	.482	-1.59	.11
	Standard	.415	.493		
Self-Control	Experimental	.144	.351	.03	.98
	Standard	.143	.351		
Role Behavior	Experimental	.037	.189	.92	.35
	Standard	.027	.161		
Amiability	Experimental	.121	.327	-1.01	.31
	Standard	.143	.351		
Obedience	Experimental	.320	.467	1.23	.22
	Standard	.284	.452		
Responsibility	Experimental	.310	.463	-1.05	.29
	Standard	.342	.475		
Consideration	Experimental	.279	.449	.11	.91
	Standard	.276	.448		
Interest	Experimental	.187	.390	.70	.48
	Standard	.170	.376		
Studious	Experimental	.062	.241	.43	.67
	Standard	.055	.229		

<sup>a</sup> N for Standard Version = 489; for Experimental Version = 487.

response category and the questionable accuracy of responses to this question in general. But it suggests that "child" and "he" succeed in referring to both boys and girls, "child" somewhat more successfully.

However, if the words under consideration were as neutral psychologically as they are grammatically, one would expect the remainder of the subsamples to think of boys and girls in approximately equal proportions. In fact, boys are thought of approximately five times

Table 2. Sex of Child Considered, by Question Wording: 1980 GSS

Sex of Child Considered	Question Wording		Total
	Standard	Experimental	
Mostly boys	22%	14%	18%
Both boys and girls	74	83	79
Mostly girls	4	3	3
(N)	(488)	(483)	(971)

$\chi^2 = 11.6$ ,  $df = 2$ ,  $p < .003$ .  
Gamma = .22.

more often than girls, regardless of question wording. The proportion thinking about girls is quite small, and almost identical in the two versions, 4 percent of those given the standard version and 3 percent of those given the experimental version. Boys are thought of by 14 percent of the experimental group and 22 percent of the comparison group. Both words appear to be biased in favor of males, though "he" more than "child." A bias which affects this large a proportion of a sample in response to an almost unnoticeable suggestion (and one which would not be recognized as a suggestion at all if one considered only common language usage) must be taken seriously. Both question wordings appear to be biased in favor of males. The bias is similar enough that even if the traits were evaluated differently by respondents for whom the sex of the child had been specified, there might not be a significant wording effect on the means of the trait evaluations in these data.

#### RESPONDENT CHARACTERISTICS, QUESTION WORDING, AND SEX OF CHILD CONSIDERED

It is conceivable that respondent characteristics interacted with question wording in influencing the sex of child considered when respondents answered the trait questions. Two obvious possibilities are whether or not the respondent is a parent, and the sex of the respondent. There is no difference ( $p < .15$ ) between the responses to the follow-up question given by parents and nonparents. Two percent (2 percent) more parents than nonparents select the "both" response; 4 percent more nonparents than parents select the "boys" category (table not shown). There are sex differences, however.

Table 3 presents the choices of male and female respondents within question wording. The interactions in this table are complex. The table was tested with a log-linear hierarchical modeling procedure, and no model less inclusive than the saturated model fit. (All less inclusive models resulted in chi-squares with probabilities less than .03.) In the full model only the experimentally controlled effects—the question wording marginals and respondent sex by question wording—were clearly nonsignificant ( $p < .47$ ). The respondent sex by sex of child interaction showed borderline significance ( $p < .08$ ), but all other main effects and interactions, including the three-way interaction, were highly significant ( $p < .02$  or better). Statistics for the subtables are presented in Table 3 as an interpretive aid and are used in the following discussion.

There are no significant differences between men and women in the sex of child considered in the experimental version ( $p < .67$ ), but there are in the standard version ( $p < .003$ ). With the standard

Table 3. Sex of Child Considered, by Question Wording and Sex of Respondent: 1980 GSS

Sex of Child Considered	Question Wording			
	Standard Sex of Respondent		Experimental Sex of Respondent	
	Male	Female	Male	Female
Mostly boys	27%	18%	14%	14%
Both boys and girls	72	76	82	84
Mostly girls	1	6	4	2
(N)	(220)	(268)	(204)	(279)

Question Wording by Sex of Child:  $\chi^2 = 11.6$ ,  $df = 2$ ,  $p < .003$ , Gamma = .22.  
Standard Wording—Sex of Child by Sex of Respondent:  $\chi^2 = 11.8$ ,  $df = 2$ ,  $p < .003$ ,

Gamma = .30.

Experimental Wording—Sex of Child by Sex of Respondent:  $\chi^2 = .84$ ,  $df = 2$ ,  $p < .67$ ,

Gamma = -.02.

Male Respondents—Sex of Child by Question Wording:  $\chi^2 = 13.44$ ,  $df = 2$ ,  $p < .001$ ,

Gamma = .40.

Female Respondents—Sex of Child by Question Wording:  $\chi^2 = 6.14$ ,  $df = 2$ ,  $p < .05$ ,

Gamma = .06.

wording 9 percent more men than women thought about boys, while 5 percent more women than men thought about girls. Both men and women chose the "both" category more often in the experimental than the standard version, and both thought about boys more often in the standard than the experimental version. The latter difference is larger for men (13 percent) than for women (4 percent). Thus, question wording affects the sex of child considered for both men and women, but the effect is weak and of borderline significance for women ( $p < .05$ ), and moderate and significant for men ( $p < .001$ ). While the experimental question wording did not result in girls being thought of as often as boys, it did eliminate a sex-of-respondent effect on the sex of child considered, an effect that was stronger for men than for women.

An analysis of the way trait evaluation might have been affected by the sex of child considered should include not only respondent sex, but education and age as well. However, not only are the item distributions in this case too skewed for analysis of variance, but the number of respondents thinking about girls is too small to permit introduction of controls in contingency table analysis without a discouraging number of empty cells appearing.

### Discussion

The experimental variation in question wording did not produce important differences in the evaluations of the traits compared to the standard version. The reasons for the lack of effect cannot be fully

analyzed with the available data. However, responses to the follow-up question indicate that the absence of a question wording effect cannot be attributed to the fact that respondents thought about boys and girls in a random fashion. On the contrary, there were systematic differences by question wording and respondent characteristics in the sex of child considered. The suggestion of a male child was very weak in the standard version of the question, but over 20 percent of the sample responded by thinking about boys, 18 percent more than thought about girls. While the experimental form lessened the bias toward male children in the minds of the respondents, the bias was not eliminated.

The procedure that took place in the minds of respondents who said they considered boys and girls equally can only be speculated on. What it means to consider boys and girls equally in this case is not at all obvious, given that the sex of the child considered was found by Kohn to be an important predictor of evaluations—that is, sex specifies distributions of evaluations with different means (and possibly different variances as well). The best rewording of these items would specify both sex and age of child.

The dangers of assuming that generic words like "he" or "child" include males and females equally is clear. Obviously, if responses to an item are known to be unaffected by which sex is referred to, or if sex differences are too small to be substantively interesting, the choice of male, female, or "generic" words is open to the investigator's taste. However, if a variable can reasonably be expected to take a different value depending on whether males or females are referred to, they should be referred to in separate items. Failure to do this may result in a misestimation of population values—means and variances—or an inability to say clearly to what population estimated values refer.

### References

- Broverman, I., S. Vogel, D. Broverman, F. Clarkson, and P. Rosenkrantz  
1970 "Sex role stereotypes and clinical judgments of mental health." *Journal of Consulting and Clinical Psychology* 34:1-7.
- Davis, James A.  
1980 *General Social Surveys, 1972-1980; Cumulative Codebook/Principal Investigator*, James A. Davis; Associate Study Director, Tom W. Smith; Research Assistant, C. Bruce Stephenson. Chicago: National Opinion Research Center.
- 1980 *General Social Surveys, 1972-1980. [Machine readable data file]*. Principal Investigator, James A. Davis; Associate Study Director, Tom W. Smith; Research Assistant, C. Bruce Stephenson. NORC, ed. Chicago: National Opinion Research Center producer, 1980; Storrs, CT: Roper Public Opinion Research Center, University of Connecticut distributor.
- Kohn, Melvin L.  
1969 *Class and Conformity*. Homewood, Ill.: Dorsey Press.