

# A Question of Wording

*Philip Gendall and Janet Hoek*

Although researchers have recognised the importance of question wording in survey research for many years, the search for generalisable rules of question wording has proved elusive. This paper reports the results of four studies on question wording: a comparison of an open with a closed question on the same topic, a comparison of negative versus positive statements, a test of different question wording variations, and a test of the effect of one question on another. The conclusion reached is that, while it is possible to write the same question in a number of different ways, simply changing one word may change the whole meaning of a question. However, other questions are more resistant to wording variations and further research in this area may help distinguish between specific variations that affect question meaning and those that do not.

Keywords: question wording attitude questionnaire

## Introduction

One of the paradoxes of questionnaire design is that the more you know, the more you realise how little you know. Questionnaire design is one vocation where ignorance is bliss. Many questionnaire designers take for granted that the questions they ask are the same as those their respondents answer and that all respondents answer the same question, yet Belson (1986) has clearly shown the fallacy of these assumptions.

Most developments in survey research over the last twenty years have been concerned with increasingly powerful analytical techniques and their application to larger and larger data bases. Consequently factor analysis, discriminant analysis, cluster analysis and various types of multidimensional scaling are now routinely applied to survey data. But these developments have often lost sight of the fact that the quality of data analysis depends on the quality of the data analysed. No amount of multivariate analysis can improve the quality of the original survey responses.

Over 40 years ago Gallup (1947) concluded that survey results were influenced more by question design than by sample design. Since then, despite the efforts of Belson, Converse, Kalton, Schuman, Presser and others, the development of knowledge about question wording, and particularly of an underlying theory of question wording, has been very slow. Generalisable rules about question design have been difficult to establish, and, although some authors have offered general rules for practitioners, such rules are often contradictory and are rarely based on exhaustive empirical research.

However, the observation that some question wording variations have little impact on the stability of survey results, suggests that it may eventually be possible to develop a theory of question wording. Labaw (1980), for example, cites the results of a series of six polls concerned with American foreign policy in Panama (see Figure 1). These polls were conducted over a period of two years and each poll involved a differently worded question, but the results were strikingly similar. The obvious implication is that these were essentially the same question.

### Figure 1. Stability of poll results for differently-worded questions

---

"Do you agree or disagree with the statement that our Government should eventually return control of the Panama Canal to the Government of Panama?" (CBS, May 1976)

Agree	24%
Disagree	52
Undecided	24

"Do you think the time has come for us to modify our Panama Canal Treaty or that we should insist on keeping the Treaty as originally signed?" (Roper, Jan 1977)

For modification	24%
Against	53
Undecided	23

"Do you favour or oppose giving the Panama Canal back to the Panamanians even if we maintain our right to defend it?" (Yankelovich, March 1977)

For giving back	29%
For holding on to it	53
Undecided	18

"Do you think the United States should negotiate a treaty with Panama where, over a period of time, Panama will eventually own and run the Canal?" (Caddell, May 1977)

Should negotiate	27%
Should not negotiate	51
Undecided	22

"The Senate now has to debate the Treaties that President Carter signed granting control of the Panama Canal to the Republic of Panama in the year 2000. Do you approve or disapprove of these Treaties?" (CBS, Jan 1978)

Approve	29%
Disapprove	51
No opinion	20

"Do you think the Senate should have approved the Panama Canal Treaties, or should not have approved them?" (Roper, June 1978)

Should have approved	30%
Should not	52
Don't know	18

---

Source: Labaw (1980)

The effect of changing the question is illustrated by the results of another poll conducted during the period in question. Although the question is similarly worded to those in Figure 1, there is an additional emotional element which apparently turns it into a different question (see Figure 2).

## Figure 2. The effect of changing the question

---

"Would you favour or oppose approval of the Panama Canal Treaty if an amendment were added, specifically giving the United States the right to intervene if the Canal is threatened by attack?" (NBC, Jan 1978)

Favour of revised Treaty	65%
Oppose	25
Undecided	10

---

Source: Labaw(1980)

From these results, Labaw concludes that question wording variations as such have little impact on the stability of survey results. Question wording variations only become significant, according to Labaw, when the variations introduce or tap a different concept, reality or emotional level surrounding an issue. This is reassuring for questionnaire designers, because their task would be impossible if every change in question wording, no matter how small, produced a different question. But it still does not establish how researchers can ensure that the questions their respondents answer are the ones they meant to ask.

The only way to develop a comprehensive and reliable set of rules for question design is to systematically test alternative question forms, sequences and wording variations in a variety of situations, until those results which are generalisable can be differentiated from those which are situation-specific. Inevitably this will be a long, slow process, but the benefits of the knowledge gained would be far more than peace of mind for questionnaire designers. It would guarantee that the information collected in surveys was as reliable as the techniques used to analyse and interpret it imply.

This paper presents the results of four small pieces of research in the area of question wording: a comparison of an open and a closed question on the same topic, a comparison of responses to positive and negative versions of the same attitude statements, a test of different question wording variations, and a test of the effect of one question on another.

## Method

The vehicle for this research was the 1989 Palmerston North Household Omnibus, which is conducted annually by students from the Marketing Department of Massey University. The survey area covers households within the Palmerston North city boundary, and the sample is based on clusters of four interviews (two with males, two with females, 15 years of age or older) around randomly-selected starting points. Substitutions are made for households where an interview is refused or where no contact can be made after three attempts.

Two versions of the Omnibus questionnaire were used to allow the four pieces of research, each of which employed a split sample technique, to be conducted. The responses to each version of the questionnaire were weighted so that the age-sex distribution of the two subsamples - those who answered Version 1 and those who answered Version 2 - were the same. The subsample sizes were 347 for Version 1 and 311 for Version 2.

## Results and Discussion

### Open Versus Closed Questions

Version 1 of the questionnaire included the following open question:

"What do you think is the single most important problem facing New Zealand right now?"

In Version 2, this question was replaced with a closed alternative:

"Would you please look at this card and tell me which of these you think is the single most important problem facing New Zealand right now . . . ?"

AIDS

INFLATION

HIGH EXCHANGE RATE

LAW AND ORDER

INTEREST RATES

UNEMPLOYMENT

THE ECONOMY IN GENERAL

RACIAL PROBLEMS

The open version of the question and the response categories for the closed version were taken from the regular Heylen omnibus. An additional response category, "AIDS", was added at the top of the list to test the effect of item order and the presence of an item not generated by open questioning. Responses to the open question were coded into the "closed" categories of the alternative question wherever possible, and the resulting distribution compared with that for the closed question. Table 1 shows the results of these comparisons.

As expected, the pattern of responses to the open and closed question differed, although the general picture which emerged from both was similar (for example, unemployment was clearly the most important issue among all respondents at the time). Some response categories generated by the open question were not included in the prespecified list of responses to the closed question, and AIDS was never given as a response to the open question, despite the fact that 6% of those asked the closed question gave this as their response.

Analysis of the open responses was, of course, subjective and a better list of closed alternatives may have been generated by some qualitative research preceding the design of this question. However, this exercise supports earlier findings (Schuman & Presser 1981; Kalton & Schuman 1982; Belson 1986; Converse & Presser 1986) that:

1. the patterns of responses to open and closed versions of the same question are likely to differ, and
2. the pattern of responses to closed questions is heavily influenced by the choices presented to respondents.

**Table 1. Comparison of open and closed responses**

<b>Most important problem facing New Zealand</b>	<b>Closed Question %</b>	<b>Open Question %</b>
Unemployment	56	50
Economy in General	22	11
Law and Order	9	5
Racial Problems	7	10
AIDS	6	-
Don't Know	1	-
Unstable Government	-	5
Declining Moral Standards	-	3
Other*	-	16
<b>Total</b>	<b>100</b>	<b>100</b>

\* Including: schooling, drugs, lack of Christian faith, laziness, David Lange, breakdown of family.

### **Positive Versus Negative Attitude Statements**

Both versions of the questionnaire included the following question:

"I'm going to read you some statements about rabbit meat. Please tell me whether you agree or disagree with each statement. If you have no opinion, just say so. Would you agree or disagree that . . .?"

The statements which followed were similar for each version, except that in Version1 all the statements were "positive" while in Version 2, all the statements were "negative".

The rationale for this design was that a "positive" or "negative" set of statements might influence respondents' answers to the statements. In other words, a set of positive statements might produce higher agreement than the level of disagreement for a set of negative statements.

The results of this exercise are shown in Table 2, where the proportions of respondents who had an opinion are reported. The percentage of "don't knows" ranged from 8% to 78%, and was generally of the order of 30% to 40%.

Although some of the differences between the equivalent responses to the different versions were quite large (up to 15%), there was no discernible pattern to these differences and no evidence to support the hypothesis that a positive or negative set of statements has a predictable effect on respondents' answers. Despite the fact that only the largest of the observed differences is statistically significant at the 5% level, the results suggest that some words or statements which appear to be "opposites" are not perceived by respondents as such. This issue is discussed in more detail in the following section.

**Table 2. Respondents' reactions to "positive" and "negative" statements**

<b>Rabbit Meat . . .</b>	<b>Version 1 Positive Set % Agree</b>	<b>Version 2 Negative Set % Disagree</b>	<b>Difference V2 - V1 %</b>
Is an everyday/not an everyday meat	10	10	0
Is readily available/not readily available in supermarkets	10	12	+2
Is a healthy/unhealthy meat	87	89	+2
Is a low/high fat meat	98	96	-2
Is easy/difficult to prepare	87	81	-6
Is good/poor value for money	54	61	+7
Is a suitable/not a suitable meat for guests	51	59	+8
Has a pleasant/unpleasant smell when cooking	56	64	+8
Is a cheap/expensive meat	61	53	-8
Has little/a lot of waste	90	75	-15*

\*Difference significant at 5% level.

### Question Wording Variations

Respondents were presented with a set of attitude statements covering a wide range of social issues and asked to agree or disagree with each statement. About a quarter of the statements were identical in both questionnaires; the rest contained various question wording variations.

**Table 3. Responses to identical attitude statements**

<b>Attitude statements common to both questionnaires</b>	<b>Agree %</b>	<b>Version 1 Disagree %</b>	<b>DK %</b>	<b>Agree %</b>	<b>Version 2 Disagree %</b>	<b>DK %</b>
Compulsory military training would be good for young unemployed people	65	32	3	66	31	4
Most unemployed people are basically lazy	20	75	5	19	76	4
The death penalty should be reintroduced	29	62	10	34	58	9
There should be stricter censorship of videotapes	62	31	8	60	35	5
Women should have more opportunity to work outside the home	78	17	5	76	19	5
More childcare should be provided for working mothers	68	24	8	65	28	7
Women are still treated as second class citizens	40	55	5	47	47	6
The country cannot afford to support solo mothers	47	41	12	43	44	14
Women should stay at home while their children are young	59	34	7	60	34	5

Table 3 shows the responses to the nine statements which were identical in Version 1 and Version 2. For all but two of these statements the pattern of responses for each questionnaire is virtually identical. This suggests that the weighting procedure described previously was successful in producing two balanced subsamples, and, furthermore, that any differences in the responses to other statements were the result of differences in question wording rather than the result of differences in the composition of the subsamples presented with each version of the questionnaire.

### Introducing a New Concept

The pair of statements shown in Table 4 was designed to test the effect of introducing a new concept into a question.

**Table 4. The effect of introducing a new concept into a question**

Attitude Statements	Agree %	Disagree %	DK %
Women should get equal pay for equal work	93	6	1
Women should get equal pay for equal work, even if this results in higher unemployment	78	14	7

In this case, the introduction of the qualifying phrase, "even if this results in higher unemployment", influenced the pattern of responses in a predictable way (less agreement, more disagreement, more uncertainty). Thus, while it is clear that most respondents supported the concept of "equal pay" for women, it is equally clear that the level of support depends on the trade-off presented.

### "Reversible" Statements

The two statements shown in Table 5 are identical in all respects except that the order of two words, "men" and "women" is reversed in the second statement. Although these statements are linguistically reversible, it is clear that this change affected the way respondents interpreted them. When "women" became the object of the question there was a dramatic increase in the proportion of "don't knows", and, among those with an opinion, a higher level of agreement that women are better suited emotionally for politics than men.

In this case, shifting the focus of the question from men to women changed the emotional content of the second statement. Even though both statements contained the same words, the different results generated by changing their order suggests that respondents neither perceived nor treated them as logical opposites. Question designers face the problem that both are equally plausible statements, yet the pattern of responses produced by each is quite different.

**Table 5. Comparison of responses to "reversible" statements**

"Reversible" Statements	Agree* %	Disagree* %	DK %
Most men are better suited emotionally for politics than most women	17 (19)	73 (81)	11
Most women are better suited emotionally for politics than most men	20 (30)	46 (70)	34

\*Figures in parentheses are proportions of those with an opinion.

### "Mirror Image" Statements

Eight of the attitude statements used in the 1989 omnibus were constructed in such a way that a "mirror image" of each statement appeared on each version of the questionnaire. If these statements were true "mirror images", the proportion of respondents agreeing with one version would be the same as the proportion disagreeing with the other version, and vice versa. The actual results are shown in Table 6.

**Table 6. Responses to "mirror image" attitude statements**

"Mirror image" statements	Version 1 Positive Set % Agree	Version 2 Negative Set % Disagree	Difference V2 - V1 %
The law should allow/forbid public speeches which promote racism	13	31	18*
I am in favour of/opposed to rugby tours to South Africa	45	57	12*
Nuclear armed ships should be allowed to visit/prevented from visiting New Zealand	32	28	-4
The police should be/should not be armed	31	42	11*
Homosexuality should be/should not be regarded as a crime	20	29	9*
Abortion should be legal under some circumstances/ should not be legal under any circumstances	81	75	-6
Most Maori get a fair go/don't get a fair go in New Zealand	77	72	-5
People should be able to choose/should not have a choice whether they belong to a trade union	89	78	-11*

\*Difference significant at the 5% level.

Differences between the equivalent responses to the eight "mirror image" statements varied from 4% to 18%, suggesting that most, if not all, of the statements were not in fact interpreted as logical opposites. Furthermore, there was no discernible pattern to the differences between the two sets of responses, nor was there an obvious explanation for them.



Previous research has shown that the substitution of the words "forbid" and "allow", which are logical opposites, has a predictable effect on the pattern of responses (Kalton, Collins & Brook 1978; Schuman & Presser 1981; Kalton & Schuman 1982; Converse & Presser 1986; Hippler & Schwarz 1986). More people are willing to "not allow" something than are willing to "forbid" it. This phenomenon was confirmed by our research (83% of respondents disagreed that the law should allow public speeches which promote racism, but only 66% agreed that the law should forbid such speeches). However, it is difficult to generalise beyond this particular pair of antonyms. The terms "in favour of" and "opposed to" produced a similar pattern of responses to those for the allow/forbid statements, but "allowed to" and "prevented from" did not. Similarly, the responses to the statements concerning abortion and trade unions would have produced positive differences in Table 6, if the proposition that respondents are more willing to "not allow" something than they are to forbid or prevent it was generalisable.

The most that can be said on the basis of the results in Table 6 is that changes in question wording can produce differences in respondents' responses, but not necessarily, and that the effect is often unpredictable.

### **Influence of One Question on Another**

The final piece of research involved a small experiment designed to test whether the response to one item influences respondents' answers to a subsequent item.

Each version of the questionnaire contained a "mirror image" attitude statement, one suggesting that people with AIDS have themselves to blame, the other suggesting that they are not to blame. For both versions the next statement asked respondents to judge whether people with AIDS received too little sympathy from society.

The hypothesis was that respondents presented with the "favourable" initial statement might, as a result, respond more charitably to the next statement. Such behaviour would be supported by the phenomenon of acquiescence, the tendency of respondents to agree with any statement regardless of its content, (Wright 1976; Kalton & Schuman 1982). If a respondent "acquiesced" with the favourable version of the first statement, logically he or she should disagree with the second statement, and vice versa. The outcome of this test is shown in Table 7.

Despite the different responses to the questions which differed in each version of the questionnaire, the distributions of responses to the question common to versions 1 and 2 were virtually identical. In other words, there was no support for the hypothesis that response to this question was influenced by the response to the preceding question.

This result conflicts with evidence from a number of studies (Schuman & Presser 1981; Kalton & Schuman 1982; Crespi & Morris 1984; Converse & Presser 1986) which clearly showed that the order of questions in a survey does affect the answers to them. However, these studies also showed that this does not always happen.

**Table 7. The influence of one question on another**

<b>Attitude Statements</b>	<b>Agree* %</b>	<b>Disagree* %</b>	<b>DK %</b>
<b>Version 1</b>			
"Most people with AIDS only have themselves to blame for having the disease"	38 (41)	54 (59)	8
<b>Version 2</b>			
"Most people with AIDS are not to blame for having the disease"	33 (43)	44 (57)	23
<b>Versions 1 &amp; 2</b>			
"People who have AIDS get much less sympathy from society than they ought to get"	62 61	20 21	18 18

\*Figures in parentheses are proportions of those with an opinion.

## Conclusions

From this research there was no evidence to support the proposition that a "positive" or "negative" set of attitude statements could influence respondents' answers in a predictable way. Nor did our results support the notion that the content of a particular question could influence the answer to another (though this effect is well-established by previous research). However, the results did confirm that different question forms (in this case an open and a closed question) produce different responses, and that question wording effects are potentially important, but often unpredictable.

On the first issue, open versus closed questions, the best advice is to use closed questions where possible. Then at least the context of the question is the same for all respondents. However, it is important to remember that the distribution of responses for a closed question is critically dependent on the answer set presented to respondents. If researchers omit an important answer, the inclusion of "other" as a final category will not compensate for this deficiency. Similarly, if they include an unimportant answer, its importance is likely to be overestimated.

The knowledge that any one question may be beset with a host of problems suggests that researchers should use a number of questions to study a particular issue. In this way they might avoid the worst implications of individual question wording effects. In other words, the use of multiple questions on a single issue gives some protection against the unpredictable effects of question wording.

"Agree-disagree" statements appear to be particularly prone to question wording effects, and in most cases forced-choice questions are better if only for the fact that they are more likely to encourage a more considered response. By itself "The Government should see to it that everyone receives adequate medical care" seems plausible, but so does "Everyone should be responsible for their own medical care", yet the two are contradictory. Thus for most

purposes the better question is: "Should the Government see to it that everyone receives adequate medical care, or should everyone be responsible for their own medical care?"

What seems clear from this study and other research on question wording is that it is possible to write the same question in a number of different ways with no effect on respondents' understanding of it. On the other hand, simply changing one word may change the whole meaning of a question. The problem for questionnaire designers is knowing when wording variations have changed a question and when they have not. Only by systematically researching the effect of question wording variations will this distinction become apparent.

## References

- Belson WA (1986). *Validity in Survey Research*. London: Gower Publishing Co.
- Bishop GF; Oldendick RW & Tuchfarber A (1984). What must my interest in politics be if I just told you "I don't know"?. *Public Opinion Quarterly*, 48, 510-519.
- Converse JM & Presser S (1986). *Survey Questions: Handcrafting the Standardised Questionnaire*. New Delhi, India: Sage Publications.
- Crespi C & Morris D (1984). Question order effect and the measurement of candidate preference in the 1982 Connecticut elections. *Public Opinion Quarterly*, 48, 578-591.
- Gallup G (1947). The quintamensional plan of question design. *Public Opinion Quarterly*, 3, 385-393.
- Hippler HJ & Schwarz N (1986). Not forbidding isn't allowing. The cognitive basis of the forbid-allow asymmetry. *Public Opinion Quarterly*, 50, 87-96.
- Kalton G; Collins M & Brook L (1978). Experiments in wording opinion questions. *Applied Statistics*, 27 (2), 149-161.
- Kalton G & Schuman H (1982). The effect of the question on survey responses: a review. *Journal of the Royal Statistical Society*. 145, (1), 42-57.
- Labaw PJ (1980). *Advanced Questionnaire Design*. Cambridge, Massachusetts: Art Books.
- Schuman H & Presser S (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Wright JD (1976). Does acquiescence bias the index of political efficacy?. *Public Opinion Quarterly*, 39, 219-226.

**Philip Gendall is Professor and Head of Department, and Janet Hoek is a Lecturer, in the Department of Marketing, Massey University.**