

## 2.6 HOW TO WRITE GOOD ITEMS

Over the past 50 years, survey researchers have accumulated a considerable body of knowledge and experience about what makes a questionnaire item good and what the potential pitfalls are. However, most specialists also emphasize that question design is not a 100%

scientific activity because in order to write good items one also needs a certain amount of creativity and lots of common sense. Furthermore, alternative versions of questions must be rigorously piloted because in the absence of hard and fast theoretical rules, “tests of practicability must play a crucial role in questionnaire construction” (Moser & Kalton, 1971, p. 350)

**Well said...**

“The writing of successful attitude statements demands careful pilot work, experience, intuition and a certain amount of flair.”

(Oppenheim, 1992, p. 180)

In the following I will summarize the do’s and don’ts of item writing. Most of the material will concern the most common question types, rating scale items.

**In writing questionnaire items...**

“no amount of textbook admonition can take the place of common sense.”

(Moser & Kalton, 1971, p. 310)

### 2.6.1 Drawing up an ‘item pool’

It is generally recommended by survey specialists that when we get down to writing the actual items, we should start doing so without restricting ourselves to any number limitations. Let our imagination go

free and create as many potential items as we can think of – this collection is referred to as the *item pool*. At this stage, successful item designers rely heavily on their own verbal creativity, but they also draw on two additional sources:

1. *Qualitative, exploratory data* gathered from informants, such as *notes* taken during talks and brainstorming in focus or discussion groups; recorded unstructured/semi-structured *interviews*; and *student essays* written around the subject of the enquiry. The best items are often the ones that sound as if they had been said by someone – so why not include phrases and sentences that have indeed been said by real informants?
2. *Borrowing questions* from established questionnaires. Questions that have been used frequently before must have been through extensive piloting and therefore the chances are that “most of the bugs will have been ironed out of them” (Sudman & Bradburn, 1983, p. 120). Of course, you will need to acknowledge the sources precisely.

### **Provided you acknowledge the sources...**

“The best advice we can offer to those starting out to write attitude questions is to plagiarize. While plagiarism is regarded as a vice in most matters, it is a virtue in questionnaire writing – assuming, of course, that you plagiarize good quality questions.”

(Sudman & Bradburn, 1983, p. 119)

## **2.6.2 Rules about item wording**

### *Aim for short and simple items*

Whenever possible, questionnaire items should be short, rarely exceeding 20 words. They should preferably be written in simple sen-

tences rather than compound or complex sentences, and each should contain only one complete thought.

**Quite so!**

“... short questions are good questions.”

(Brown, 2001, p. 45)

*Use simple and natural language*

As a rule, in questionnaire items we should always choose the simplest way to say something. Items need to be kept clear and direct, without any acronyms, abbreviations, colloquialisms, proverbs, jargon, or technical terms. We should try to speak the ‘common language’ and find synonyms for the “polysyllabic and Latinate constructions that come easily to the tongue of the college educated” (Converse & Presser, 1986, p. 10).

Oppenheim (1992) argues that the most important rule in writing rating scale statements is to make them *meaningful* and *interesting* to the respondents. As he points out, “There are many attitude scales which falter because the items have been composed in the office according to some theoretical plan and fail to arouse much interest in the respondents” (p. 179). The best items are the ones that sound like being taken from actual interviews, and Oppenheim encourages item writers not to refrain from using contentiously worded statements that include phrases relating to feelings, wishes, fears, and happiness.

### *Avoid ambiguous or loaded words and sentences*

It goes without saying that any elements that might make the language of the items unclear or ambiguous need to be avoided. The most notorious of such elements are:

- Nonspecific adjectives or adverbs (e.g., good, easy, many, sometimes, often).
- Items containing universals such as 'all,' 'none,' 'never.'
- Modifying words such as 'only,' 'just,' 'merely' – these should be used with moderation.
- Words having more than one meaning.
- Loaded words (e.g., 'democratic,' 'modern,' 'natural,' 'free,' etc.), because they may elicit an emotional reaction that may bias the answer.

It is also obvious that loaded questions such as "*Isn't it reasonable to suppose that ...?*" or "*Don't you believe that...?*" are likely to bias the respondent toward giving a desired answer and should be rephrased in a neutral way.

### *Avoid negative constructions*

Items that contain a negative construction (i.e., including 'no' or 'not') are deceptive because although they read OK, responding to them can be problematic. For example, what does a negative answer to a negative item mean? In order to avoid any possible difficulties, the best solution is to avoid the use of negatives altogether. In most cases negative items can be restated in a positive way by using verbs or adjectives that express the opposite meaning.



### *Avoid double-barreled questions*

Double-barreled questions are those that ask two (or more) questions in one while expecting a single answer. For example, the question "How are your parents?" asks about one's mother and father, and cannot be answered simply if one of them is well and the other unwell. Indeed, questions dealing with pluralisms (children, students) often yield double-barreled questions, but compound questions also often fall into this category (e.g., "Do you always write your homework and do it thoroughly?"). With double-barreled questions even if respondents do provide an answer, there is no way of knowing which part of the question the answer concerned.

### *Avoid items that are likely to be answered the same way by everybody*

due to answers are widely known  
the lack of variance this way cause

In rating scales we should avoid statements that are likely to be endorsed by almost everyone or almost no one. In most cases these items are not informative and they are certainly difficult if not impossible to process statistically. Here is a recent example from my own research (Dörnyei & Clément, 2001): A questionnaire item asked students to rate the international role/importance of six countries, including the United States. As can be imagined, most respondents gave the U.S. the top score. However, as we found out in the analyses, this did not provide enough variance to compute certain statistical results involving this item, and in some cases – when in a particular subgroup (e.g., a school) every single person gave the top score – the computer treated the responses as missing data because of the total lack of variance.

### *Include both positively and negatively worded items*

to avoid having answers at only one side of the scale

In order to avoid a response set in which the respondents mark only one side of a rating scale, it is worth including in the questionnaire both positively and negatively worded items. In addition, a balanced

mixture might also reduce the harmful effects of the 'acquiescence bias' (cf. Section 1.2.2). The term 'negatively worded item' means that it focuses on negative rather than positive aspects of the target, and we should note that it is all too easy to fall into the trap of trying to express this negative aspect by using some sort of a negative construction (which has been previously warned against): I have found more than once in the past that even carefully designed and seemingly fine 'negatively worded items' had to be excluded from the questionnaire after a *post hoc* item analysis.

**Ellard and Rogers' (1993, p. 17) "Ten Commandments of Question Writing"**

- I. <sup>You shall</sup> Thou shalt not create double-barreled items.
- II. Thou shalt not use 'no' and 'not' or words beginning with 'un.'
- III. Thou shalt match the vocabulary used in items to the vocabulary of those who will respond to them.
- IV. Thou shalt not use complex grammatical forms.
- V. <sup>large adequacy</sup> <sup>?</sup> Thou shalt have 40% to 60% true- or agree-keyed items.
- VI. Thou shalt not use redundant or irrelevant items.
- VII. Thou shalt not permit any loaded questions to appear in your questionnaire.
- VIII. Thou shalt not mix response formats within a set of questions.
- IX. <sup>?</sup> Thou shalt not permit a non-committal response.
- X. Thou shalt pretest questions before collecting data.

### 2.6.3 Writing sensitive items

If the previous section has (hopefully) suggested that item writing requires special attention to details, then this is even more so when writing *sensitive items*, that is, questions addressing issues that are not easy to talk about because they may ask about *confidential personal information*, *undesirable social behavior*, or information that might pose *potential threat* to the respondent.

#### *Confidential personal information*

With regard to questions that ask about *personal information* that is usually considered private, the best advice is that the fewer of them, the better. If they are really necessary for the survey then some sort of a justification and a renewed promise of confidentiality are in order (e.g., “Finally, in order to help us to better interpret and classify your answers, would you mind telling us more about your personal and language learning background?”).

#### **Quite so!**

“Classification questions ... need a special introduction. After all, a respondent who agrees to answer questions about his leisure pursuits or to give his opinion about television may legitimately wonder why he should supply details about his family, his age, his education, his occupation, and even his income.”

(Moser & Kalton, 1971, p. 316)



### *Undesirable social behavior*

With regard to responses that might be felt will meet with *disapproval*, several strategies have been suggested in the literature. Wilson and McClean (1994) recommend that they can be diffused by the use of categories, or brands, for respondents to tick. In their seminal book on questionnaire design, Sudman and Bradburn (1983) devote a great deal of space to discussing sensitive items. Their practical suggestions to mitigate the undesirable nature of certain behaviors include:

- Wording the question in a way that it suggests that the behavior is rather common (e.g., "*Even the most conscientious teachers sometimes...* ").
- Assuming the occurrence of the behavior and asking about frequencies or other details rather than whether the behavior has occurred.
- Using authority to justify behavior (e.g., "*Many researchers now think...* ").
- Adopting a 'casual approach' (e.g., "*Did you happen to...?* ").
- Including reasons that explain the behavior (e.g., "*Does your busy schedule sometimes prevent you from...?* " or "*Have you had time to ... recently?* ").

Aiken (1997) further suggests that by phrasing the question in a way that it refers to "other people" can encourage truthful responses, and the perceived importance of sensitive questions can also be reduced if they are embedded among other questions dealing with both sensitive and nonsensitive topics.

### *Potential threat*

With regard to items in which an honest answer can pose some *real threat* to the respondent (e.g., questions about illegal activities, or asking students to evaluate their language teacher), the main task is to

convince the respondents that their answers will remain confidential. Obviously, offering complete anonymity in such cases might be helpful, but this may not be feasible in certain complex research projects where we need to match the data with information obtained from other sources (cf. Section 2.1.3). In any case, additional gestures emphasizing confidentiality are always welcome. In a classroom study already mentioned (Clément et al., 1994) where a questionnaire was administered to class groups, we asked students to evaluate both the L2 teacher and the course, and applied three confidence-building strategies:

- The questionnaire administrator was a representative of the university and thus external to the school – a fact that was sufficiently emphasized.
- We handed out envelopes in which students put their completed questionnaires and which they then sealed.
- The questionnaire administrator went around the classroom and stamped the envelopes with a university stamp on the seals.

Some questions can pose a threat not only to the respondent but also to the people or institutions the questionnaire is about. For example, few teachers are likely to be happy to allow the administration of a questionnaire in their classes that explicitly asks the students to evaluate the quality of their teaching. Interestingly, Gardner and Smythe (1981) report that educational institutions found semantic differential scales (cf. Section 2.4) less objectionable than complete evaluative statements when talking about such sensitive issues. It seems that the fact that these items do not spell out the issues in detail but only provide pairs of bipolar adjectives make them less offensive.

## 2.7 GROUPING AND ORDERING ITEMS

Once all the items to be included in the questionnaire have been written or collected, we need to decide on their order. Item sequence is a significant factor because the context of a question can have an im-

pact on its interpretation and the response given to it. Indeed, the meaning of almost any question can be altered by the adjacent questions. However, it is usually acknowledged that research has not as yet generated any specific theoretical rules to order questions, beyond some broad suggestions (Robson, 1993). Let us have a look at the four main ordering principles.

### *Clear and orderly structure*

The most important aspect of sequencing questions is to ensure that the respondents' overall impression is that the structure is well-organized and orderly. If the ordering of questions is unpredictable or seemingly haphazard, it will frustrate respondents and make the study appear ill-considered and amateurish (Newell, 1993). Neither the content nor the style of the questionnaire should "jump around" (Aiken, 1997) – the items should seem as a series of logically organized sequences. To achieve this, we need to follow certain organizing principles.

One organizing principle should be the *item format*. If the questionnaire contains items of different types, these need to be clustered together into well marked sub-sections, separated from each other by a clear set of instructions to highlight the format change for the respondent. Similarly, questions that deal with the same *topic* should be grouped together. In order to make the progression from topic to topic smoother, we may include short linking sentences such as, "*In this section we'll move on to look at more specific aspects of...*". Content-based organization, however, does not mean that the items in a multi-item scale (cf. Section 2.3.2) should be next to each other – the repetitive content may frustrate the respondents. What I usually do is take 4-5 content areas that are related to each other and then mix up the constituent items randomly.



### *Opening questions*

Just like with any other piece of writing, the initial section of a questionnaire is particularly important in that it sets the tone. This is partly the reason that instructions (cf. Sections 2.2.2 and 3.3.7) play a significant role, and this is also why the first few 'opening' questions should be carefully selected. In order to create a pleasant first impression, the starter questions need to be interesting, relatively simple yet at the same time focused on some important and salient aspect, and certainly non-threatening/sensitive.

### *Factual (or 'personal' or 'classification') questions at the end*

As Oppenheim (1992) concludes, novice researchers typically start to design a questionnaire by putting a rather forbidding set of questions at the top of a blank sheet of paper, asking for name, address, marital status, number of children, religion, and so on. These personal/classification questions tend to be very off-putting: Having been through the various introductory phases, respondents are now ready to look at some interesting questions dealing with the topic of the study. Instead, they are faced with a set of 'personal' questions not unlike those contained in the many bureaucratic forms we have to fill in when, for example, applying for a passport or registering in a hotel. This can result in a kind of anticlimax in the respondents and it may be difficult to rekindle their enthusiasm again. Thus, such personal questions are best left at the end of the questionnaire.

There is also a second reason why factual questions should not be introduced too early, and this concerns their sensitive nature. As discussed in Section 2.1.3, in many cultures issues like age, level of education, or marital status are personal and private matters, and if we ask them near the beginning of the questionnaire they might create some resistance in the respondents ("*What business of yours is this...?*"), or, in cases where respondents are asked to provide their name, this might remind them of the non-anonymous nature of the survey, which in turn may inhibit some of their answers.

### *Open-ended questions at the end*

As discussed in Section 2.5, if we include real open-ended questions that require substantial and creative writing, it is preferable to place them near the end rather than at the beginning of the questionnaire. In this way, their potential negative consequences (e.g., the required work can put some people off; others might get bogged down and spend most of the available time and mental energy agonizing over what they should write) will not affect the previous items. In addition, some people find it psychologically more acceptable to put in the necessary work if they have already invested in the questionnaire and if they know that this is the final task.

## **2.8 COMPUTER PROGRAMS FOR CONSTRUCTING QUESTIONNAIRES**

Because market research – a booming business area – utilizes questionnaires for various types of surveys, several software companies have developed commercial computer programs to cater to these needs: Currently there are over 30 available desktop packages that combine questionnaire design, data collection, and data analysis. However, as Macer (1999) summarizes, few packages rise to the challenge of each stage in the process with the same degree of accomplishment, and development effort often tends to gravitate to some areas at the expense of others. For comprehensive listings and descriptions of the programs on the market, see for example the Research Software Central database (<http://www.macer.co.uk/rscentral/rscentral.html>) or the database of the Association for Survey Computing (U.K.), which contains a classified listing of 123 software packages related to survey research, with attributes and suppliers (<http://www.asc.org.uk/Register/index.htm>).

Here I will introduce one computer program that I am familiar with: *SphinxSurvey*, distributed by Scolari/Sage (the publishers of such well-known qualitative data analysis software as NUD\*IST and NVivo), is an integrated, PC-based Windows package for conducting



questionnaire-based surveys (for a review, see Macer, 1999). It has built-in functions to help the user to design and print professional questionnaires with ease. The program can handle a variety of question types, including open and closed questions. Similar questions can be grouped and conditional jumps can be defined to permit complex question routings (e.g., if people answer 'yes' to Question X, they should move to Question Y). In addition, extensive question libraries can be developed and used to aid the preparation of an item pool.

*SphinxSurvey* is certainly a useful tool in providing a computerized framework for quick and professional questionnaire construction (the data processing functions of the program will be analyzed in Section 4.5). The novice researcher will find various ready-made options to choose from by simply clicking on items in the menu. The display format is quite flexible and the final result is fairly attractive. Because of the paramount importance of the appropriate layout (cf. Section 2.1.2), I would still design the final version of a questionnaire on a more powerful word processor, but in many situations the available formats are sufficient.