

2

Constructing the Questionnaire

Section 1.2.2 contained a long list of potential problems with self-completed questionnaires. My goal was not to dissuade people from using such instruments but rather to raise awareness of these possible shortcomings. It is true that respondents are often unmotivated, slapdash, hasty, and insincere, yet it is also an established fact that careful and creative questionnaire construction can result in an instrument that motivates people to give relatively truthful and thoughtful answers, which can then be processed in a scientifically sound manner. The relevant professional literature contains a significant body of accumulated experience and research evidence as to how we can achieve this. Some of the points highlighted by researchers are seemingly trivial in the sense that they concern small details, but I have come to believe that it is to a great extent the systematic handling of such small details and nuances that will eventually turn an *ad hoc* set of questions into an effective research instrument.

I agree...

“Questionnaires can be designed to minimize, but not eliminate, dishonest, and careless reporting.”

(Aiken, 1997, p. 58)

Constructing a good questionnaire involves a series of steps and procedures, including:

- Deciding on the general features of the questionnaire, such as the length, the format, and the main parts.
- Writing effective items/questions and drawing up an item pool.

- Selecting and sequencing the items.
- Writing appropriate instructions and examples.
- Piloting the questionnaire and conducting item analysis.

This chapter will provide an overview of these issues, offering many practical do's and don'ts to facilitate effective questionnaire construction.

Indeed...

“Questionnaires are so easy to do quickly and badly that, in a way, they invite carelessness.”

(Gillham, 2000, p. 11)

2.1 GENERAL FEATURES

Between the initial idea of preparing a questionnaire for the purpose of our research and actually getting down to writing the first draft, a number of important decisions need to be taken regarding the general features of the would-be instrument. First of all, we need to specify the maximum *length of time* that the completion of the questionnaire could take; then we need to consider general *format characteristics*; and finally we need to think about the issue of *anonymity*, particularly if we are going to target sensitive/confidential topics.

2.1.1 Length

When we design a questionnaire, the general temptation is always to cover too much ground by asking everything that might turn out to be

interesting. This must be resisted: in questionnaire design less is often more because long questionnaires can become counterproductive.

How long is the optimal length? It depends on how important the topic of the questionnaire is for the respondent. If we feel very strongly about something, we are usually willing to spend several hours answering questions. However, most questionnaires in the L2 field concern topics that have a low salience from the respondents' perspective, and in such cases the optimal length is rather short. Most researchers agree that anything that is more than 4-6 pages long and requires over half an hour to complete may be considered too much of an imposition. As a principle, I have always tried to stay within a 4-page limit: It is remarkable how many items can be included within 4 well-designed pages and I have also found that a questionnaire of 3-4 pages does not tend to exceed the 30-minute completion limit.

A further factor to consider is that if we are restricted in the time we can have access to the respondents, for example when we administer a questionnaire to learners during their teaching hours, the maximum length should be set with the slowest readers in mind. For example, in a national survey that involved the group-administration of a questionnaire in hundreds of primary school classes in various locations in Hungary (Dörnyei & Clément, 2001; Dörnyei & Csizér, in press), we could only negotiate a maximum of 30 minutes' access to the children. This meant that the questionnaire had to be cut down to three pages and an estimated 20-minute completion time in order to give everybody a chance to finish within the allotted time.

To summarize

In my experience, only in exceptional cases should a questionnaire:

- be more than 4 pages long;
- take more than 30 minutes to complete.

2.1.2 Layout

Sanchez (1992) points out that the design of the *questionnaire layout* is frequently overlooked as an important aspect of the development of the instrument. This is a mistake: Because in surveys employing self-completed questionnaires the main interface between the researcher and the respondent is the hard copy of the questionnaire; the format and graphic layout carry a special significance and have an important impact on the responses. Over the past 15 years I have increasingly come to the belief that producing an attractive and professional design is half the battle in eliciting reliable and valid data (for a discussion of the role of the layout in increasing respondent motivation, see Section 3.3.8).

What does an 'attractive and professional design' involve? The following list summarizes the five most important points:

- *Booklet format*. Not only does the questionnaire have to be short but it also has to *look* short. I have found that the format that feels most compact is that of a *booklet*. It can be achieved by taking a double-sized sheet (A3 size in Europe), photocopying two normal-sized pages on each of the sides, and then folding the sheet into two. This format also makes it easy to read and to turn pages (and what is just as important, it also prevents lost pages...).
- *Appropriate density*. With regard to how much material we put on a page, a compromise needs to be achieved: On the one hand, we want to make the pages full because respondents are much more willing to fill in a two-page rather than a four-page questionnaire even if the two instruments have exactly the same number of items. On the other hand, we must not make the pages look crowded (for example by economizing on the spaces separating different sections of the questionnaire). Effective ways of achieving this trade-off involve reducing the *margins*, using a *space-economical font* (e.g., 11- or 12-point Times New Roman), and utilizing the whole *width* of the page, for example by printing the response options next to the questions and not below (as illustrated in the following example).

	Strongly disagree	Disagree	Slightly disagree	Partly agree	Agree	Strongly agree
1. Language learning is a burden for me.						
2. Foreign languages are an important part of the school curriculum.						
3. I like the sound of English.						

On length and crowdedness

“Perhaps the most common mistake of the beginner in questionnaire construction is to crowd questions together in the hope of making the questionnaire look short. ... While length is important, the respondent’s perception of the difficulty of the task is even more important on self-administered questionnaires. A less crowded questionnaire with substantial white space looks easier and generally results in higher cooperation and fewer errors.”

(Sudman & Bradburn, 1983, p. 244)

- *Orderly layout.* Even if the page is dense, a well-designed, orderly layout that utilizes various typefaces and highlighting options (e.g., bold characters or italics) can create a good impression, whereas an unsystematic layout, even if it is more spacious, can

appear chaotic. It is also essential that the final version be nicely printed – as Newell (1993) points out, in these days of sophisticated word-processors, people are used to receiving good quality hard copy. So try and find a laser printer and a good photocopier!

- *Paper quality.* Even the quality and color of the paper might make a difference. Newell (1993) describes a colleague who has always produced documents on thick, beige paper because she believes that “(1) it stands out from the mass of other paper which might be received, (2) it is pleasant to handle, and (3) people will not have the heart to throw away such an attractive document. She says it works” (p. 109). Other researchers suggest that it may be useful to separate the various parts of the questionnaires with a certain color-code of the paper used as it clarifies the structure (Robson, 1993); for example, the paper of the cover page or the instructions can be of a different color.
- *Sequence marking.* I normally mark each main section of the questionnaire with Roman numbers, each question with consecutive Arab figures, and then letter all the subparts of a question; as a result, I may have Question 1a or 27d within Section I or III (see the example on page 22). This creates a sense of structuredness. It is also beneficial to include a phrase such as “Continued on back” at the bottom of the first side of a page that is printed on both sides. Finally, it is probably obvious but still worth mentioning that a question should not be split between two pages.

2.1.3 Sensitive topics and anonymity

It was mentioned in Section 1.2.2 that respondents are sometimes reluctant to give honest answers to sensitive questions. Questionnaire items differ greatly in terms of how threatening/imposing/sensitive/embarrassing they feel. It requires little justification that we need to approach the issue of constructing and administering the questionnaire in a very different way if it concerns, for example, the evalua-

tion of the L2 teacher or the school rather than one's interest in travelling abroad.

Example of sequence marking

I. ATTITUDES TOWARD LANGUAGE LEARNING

1. Language learning is an exciting activity.
2. Language learning often makes me happy.

II. LANGUAGE CHOICE

3. If you could choose, which foreign languages would you choose to learn next year at school? Please mark three languages in order of importance.

- (a)
- (b)
- (c)

Continued on back...

Sensitive topics

* 7
'Sensitive' topics are not confined to explicitly illegal or embarrassing subjects but also include basic demographic items such as age or marital status. Indeed, various facts of life can carry such a prominent social and emotional loading that questions targeting them often fall prey to the respondents' 'social desirability' bias (cf. Section 1.2.2). Depending on our core values, we are likely to overreport on what we conceive as a positive aspect and underreport on a negative one.

Questionnaire designers need to be aware of this tendency and a good initial rule of thumb is that we should *not* ask any sensitive questions unless absolutely necessary for the project.

In Section 2.6.3, I will discuss several item-writing strategies that might make such questions more palatable, and in Section 3.4 we will look at questionnaire administration techniques that may help to 'sell' these items. Here I would like to highlight the usefulness of an explicit statement or promise of confidentiality in overcoming possible apprehensions. Oppenheim (1992, pp. 104-105) suggests that something along the following line be displayed prominently on the front of the questionnaire:

THE CONTENTS OF THIS FORM ARE *ABSOLUTELY* CONFIDENTIAL.
INFORMATION IDENTIFYING THE RESPONDENT WILL NOT BE DIS-
CLOSED UNDER ANY CIRCUMSTANCES.

In the general instructions of a motivation questionnaire among school learners which included the appraisal of the L2 teacher and course (and was therefore particularly sensitive from the students' point of view), Glikzman, Gardner and Smythe (1982, p. 637) provided the following detailed description of how confidentiality was observed in spite of asking the students to state their names:

Your answers to any or all questions will be treated with the strictest confidence. Although we ask for your name on the cover page, we do so only because we must be able to associate your answers to this questionnaire with those of other questionnaires which you will be asked to answer. It is important for you to know, however, that before the questionnaires are examined, your questionnaire will be numbered, the same number will be put on the section containing your name, and then that section will be removed. By following a similar procedure with the other questionnaires we will be able to match the questionnaires through matching numbers and avoid having to associate your name directly with the questionnaire.

Anonymity

One frequent method used to diffuse sensitive items is to make the questionnaire *anonymous*. For example, in a student questionnaire that asked the learners to evaluate their language teacher and the course (Clément, Dörnyei, & Noels, 1994), using similar items to the ones employed in the Gliksmann et al. (1982) study just mentioned, we felt it unlikely that the 16/17-year-old teenagers in the sample were going to agree to give us honest answers without being assured about the anonymity of the questionnaires. Following the same reasoning – and particularly when legal considerations, such as local research regulations, also necessitate it – researchers often feel ‘forced’ to make the survey anonymous. The main argument to support this practice is that anonymous respondents are likely to give answers that are less self-protective and presumably more accurate than respondents who believe they can be identified (Kearney, Hopkins, Mauss and Weisheit, 1984). Anonymity, however, raises two issues:

- Opinions differ widely as to whether respondent anonymity actually fulfills its purpose in encouraging honesty and willingness to disclose. As Aiken (1997) summarizes, most adults will probably give the same answers to questionnaire items whether or not their responses are anonymous. For example, Sudman and Bradburn (1983) report on a large-scale postal survey of college graduates, in which the researchers placed the mailing label (which naturally contained the respondent’s name) on the back cover of the questionnaires and sent these out in window envelopes. Out of the 40,000 recipients, only five objected to this procedure and scratched out their names. On the other hand, in situations when an honest answer might cause embarrassment or pose actual threat to the respondent, anonymity does obviously matter. Thus, the question to consider is whether our questionnaires really falls into this category.
- Anonymity may not serve the purpose of the investigation. More often than not the researcher would like to link the data from the questionnaires to data coming from other sources; for example,

motivational data obtained by questionnaires is often correlated to achievement scores coming from end-of-term course grades or proficiency tests. Without any identity marking on the questionnaires, we simply cannot link someone's scores in the two datasets. Similarly, if we are conducting a longitudinal investigation we would not be able to follow a person's development if all the answers gathered from the multiple subjects at a time were anonymous.

In sum, sensitive items and anonymity are a serious issue that needs to be considered right from the beginning. In Section 3.4.3, I will present some approaches that have been successfully used in the past to reconcile confidentiality with the need for identification for research purposes.

2.2 THE MAIN PARTS OF A QUESTIONNAIRE

Bearing in mind the general considerations just discussed, we are now set to start drawing up the first draft of the questionnaire. Before we get down to describing the various item types, let me briefly summarize the main components of a questionnaire.

2.2.1 Title

Just like any other piece of writing, a questionnaire should have a title to identify the domain of the investigation, to provide the respondent with initial orientation, and to activate various content schemata. Because uninformative titles fail to achieve these objectives, Aiken (1997) suggests that we should try and avoid title words like "questionnaire" or "survey." For better identification, the title might be accompanied by the date of the survey administration and the name of the organization conducting or sponsoring the study.

2.2.2 Instructions

The title of the questionnaire is followed by instructions. These cannot be too long and yet need to be informative and well pitched because they play an important role in determining the respondents' feelings toward the questionnaire and in specifying how they should go about answering the items. Instructions are of two types:

- *General instruction* (or 'opening greeting') at the beginning of the questionnaire.
- *Specific instructions* introducing each new task.

General instruction

As a minimum, the general instruction (or 'opening greeting') should cover the following points (see also Section 3.2.1, for special instructions for mail surveys):

- What the study is about and why it is important or socially useful.
- The organization responsible for conducting the study.
- Emphasizing that there are no right or wrong answers; requesting honest answers and trying to elicit integrity in general.
- Promising confidentiality.
- Saying 'thank you.'

For better readability and emphasis, the instructions should be graphically highlighted, such as being printed in boldface type, and the main pieces of information can also be given in a format such as bulleted points. I would expect the following sample instruction would be suitable for most purposes.

Sample 2.1. General instruction

We would like to ask you to help us by answering the following questions concerning foreign language learning. This survey is conducted by the Language Research Group of the University of X to better understand... This is not a test so there are no “right” or “wrong” answers and you don't even have to write your name on it. We are interested in your personal opinion. Please give your answers sincerely as only this will guarantee the success of the investigation. Thank you very much for your help.

Specific instructions

Specific instructions explain and demonstrate how respondents should go about answering the questions. It is obvious that this is a crucial part. Each new task-type requires instructions, and in order to separate these instructions from the rest of the text, they should be graphically highlighted, for example by printing them in bold (just like the general instruction).

A very important role of the instructions is to explain how various rating scales (cf. Section 2.4.1) work and what the various rating criteria are. For example, if we ask the respondents to produce evaluations on a five-point scale (i.e., giving marks ranging from 1 to 5), we need to explain very clearly what each numerical category stands for. Then, to avoid misunderstandings and mistakes, a short summary of this explanation will need to be repeated at least twice on each new page. Samples 2.2 and 2.3 on pages 28-29 provide examples of instructions for two common rating scale types (see also, Sample 2.4 on page 41).

Sample 2.2. Instructions for numerical rating scales

In the following section we would like you to answer some questions by simply giving marks from 1 to 5.

1 = not at all 2 = not really 3 = so-so 4 = quite a lot 5 = very much

For example, consider the following item. If you like hamburgers very much, write '5' in the space in front of the question:

_____ How much do you like hamburgers?

Please write one (and only one) whole number in front of each question and don't leave out any of them. Thanks.

2.2.3 Questionnaire items

After the instructions comes the central part of the questionnaire, the actual items. They will be discussed in detail in Sections 2.3 – 2.7. Two points need to be made here:

- Questionnaire items rarely take the form of actual questions that end with a question mark. The item type found in Sample 2.3, for example, is far more common than that in Sample 2.2 (which is a real question).
- The items need to be very clearly separated from the instructions. This is where different typefaces and font styles come in handy.

Sample 2.3. Instructions for Likert scales

Following are a number of statements with which some people agree and others disagree. We would like you to indicate your opinion after each statement by putting an 'X' in the box that best indicates the extent to which you agree or disagree with the statement. Thank you very much for your help.

For example:

Pickled cucumbers are unhealthy.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Strongly disagree	Disagree	Slightly disagree	Partly agree	Agree	Strongly agree

If you think, for example, that there is something true about this statement but it is somewhat exaggerated, you can put an 'X' in the fourth or the fifth box.

2.2.4 Additional information

Depending on circumstances, the questionnaire may contain, usually at the end, a short additional information section in which the author can address the respondent concerning a number of issues:

- Unless the researcher or a representative is present during the completion of the questionnaire, it might be worth including a contact name (e.g., the researcher's or an administrator's) with a

telephone number or address and some explicit encouragement to get in touch if there are any questions.

- In 'distant' situations, it might also be worth summarizing briefly how the questionnaires should be returned, and even when a return envelope is provided, we should print on the questionnaire the name and the address of the person to whom the completed questionnaire is to be sent.
- It is a nice gesture (unfortunately too rarely used) to include a brief note promising to send the respondent a summary of the findings if interested (see Section 3.3.9, for a discussion of this point).
- Sometimes questionnaires can also end with an invitation to volunteer for a follow-up interview.

2.2.5 Final 'thank you'

It is basic courtesy, yet it is all too often overlooked, that the respondents should be thanked for their cooperation at the very end of the questionnaire. After all, they have done us a favor. Although I usually do not include any drawings in my questionnaires, if I did it would be located here: a smiling face or some little figure that can be seen as a nice gesture. Modern word processing packages offer many graphic designs, such as:



2.3 QUESTIONNAIRE CONTENT AND MULTI-ITEM SCALES

The first step in preparing the questionnaire items is to specify their content in explicit terms. Although this may sound obvious, it does not always happen, and vague content specifications can pose a serious threat to the validity and reliability of the instrument, particularly in two areas:

- the appropriate sampling of the content;
- the preparation of multi-item scales.

2.3.1 Appropriate sampling of the content

Ad hoc questionnaire design involves jotting down a number of relevant questions without any rigorous procedure to ensure that the coverage is comprehensive. The problem with this method, as Davidson (1996, p. 10) highlights, is that “You cannot analyze what you do not measure.” That is, not even the most sophisticated data analysis techniques will be able to compensate for leaving out some important questions from the data collection by accident. Certain omissions are bound to occur even in otherwise very thorough studies (as attested by the countless anecdotes one hears at professional conferences) but when the sampling of the content is not theory-driven, the chances for something irrecoverable to happen are obviously much greater.

On the other hand, forewarned by the potential threat of a lack of comprehensiveness, researchers may be tempted to make the questionnaire too long by covering every possible angle. Although this is undesirable, without any explicit content specifications it is almost impossible to decide what limit to put on the range of questions. So, the initial stage of questionnaire design should focus on clarifying the research problem and identifying what critical concepts need to be addressed by the questionnaire. To facilitate this, it is often recommended that the questionnaire design phase be preceded by a small-scale qualitative study (e.g., focus group interviews) to provide information on the relevant points and issues.

Yes!

"The temptation is always to cover too much, to ask everything that might turn out to be interesting. This must be resisted."

(Moser & Kalton, 1971, p. 309)

Once a theoretically sound shortlist of specific content areas has been drawn up, it becomes possible to eliminate all the questions that are only of peripheral interest but not directly related to the variables and hypotheses that the questionnaire has been designed to investigate. Such a shortlist is also necessary to be able to produce 'multi-item scales' (see below), without which no questionnaire can be reliable.

To illustrate this process, let us take a concrete example: the design of a short questionnaire to assess student attitudes toward the language teacher. Which aspects of the teacher shall we concentrate on? Without any theoretical guidelines we could be producing an infinite number of items, all seemingly targeting important teacher characteristics. In a study where we faced this problem (Clément et al., 1994), in order to follow a more systematic approach we first conducted a review of the relevant literature and identified four main dimensions of teacher appraisal: *competence*, *rapprochement* (with the students), *motivation*, and *teaching style/personality*. We then used this list to guide us in generating the item pool.

2.3.2 Using multi-item scales

Multi-item scales are the key components to scientific questionnaire design, yet this concept is surprisingly little known in the L2 profession. The core of the issue is that when it comes to assessing attitudes, beliefs, opinions, interests, values, aspirations, expectations, and other personal variables, the actual wording of the questions assumes an

unexpected importance: minor differences in how the question is formulated and framed can produce radically different levels of agreement or disagreement, or a completely different selection of answers (Gillham, 2000). We do not have such problems with factual questions: if you are interested in the gender of the respondent, you can safely ask about this using a single item, and the chances are that you will get a reliable answer (although the item: "*Your sex:*" might elicit very creative responses in a teenage sample...). However, with non-factual answers it is not unusual to find that responses given by the same people to two virtually identical items differ by as much as 20% or more (Oppenheim, 1992). Here is an illustration:

Converse & Presser (1986, p. 41) report on a case when simply changing "forbid" to "not allow" in the wording produced significantly different responses in the item "*Do you think the United States should [forbid/not allow] public speeches against democracy?*" Significantly more people were willing to "not allow" speeches against democracy than were willing to "forbid" them. Although it may be true that on an impressionistic level "not allow" somehow does not sound as harsh as "forbid," the fact is that 'allow' and 'forbid' are exact logical opposites and therefore it was not unreasonable to assume that the actual content of the two versions of the question was identical. Yet, as the differing response pattern indicated, this was not the case. Given that in this example only one word was changed and that the alternative version had an almost identical meaning, this is a good illustration that item wording in general has a substantial impact on the responses. However, there does not seem to be a reliable way of knowing exactly what kind of an effect to expect.

So what is the solution? Do we have to conclude that questionnaires simply cannot achieve the kind of accuracy that is needed for scientific measurement purposes? We would have to if measurement theoreticians – and particularly Rensis Likert in the 1930s – had not discovered an ingenious way of getting around the problem: by using *multi-item scales*. These scales refer to a cluster of several differently worded items that focus on the same target (e.g., five items targeting attitudes toward language labs). The item scores for the similar questions are summed, resulting in a total scale score (which is why these scales are sometimes referred to as *summative scales*), and the underlying assumption is that any idiosyncratic interpretation of an item

will be averaged out during the summation of the item scores. In other words, if we use multi-item scales, "no individual item carries an excessive load, and an inconsistent response to one item would cause limited damage" (Skehan, 1989, p. 11). For example, the question "Do you learn vocabulary items easily?" is bound to be interpreted differently by different people, depending on how easy they consider 'easily,' but if we include several more items asking about how good the respondents' memorization skills are, the overall score is likely to reflect the actual level of the development of this skill. Thus, multi-item scales maximize the stable component that the items share and reduce the extraneous influences unique to the individual items.

A problem indeed...

"When we sometimes despair about the use of language as a tool for measuring or at least uncovering awareness, attitude, percepts and belief systems, it is mainly because we do not yet know *why* questions that look so similar actually produce such very different sets of results, or how we can predict contextual effects on a question, or in what ways we can ensure that respondents will all use the same frame of reference in answering an attitude question."

(Oppenheim, 1992, p. 149)

Because of the fallibility of single items, there is a general consensus among survey specialists that more than one item is needed to address each identified content area, all aimed at the same target but drawing upon slightly different aspects of it. How many is 'more than one'? The most well-known standardized questionnaire in the L2 field, Robert Gardner's (1985) Attitude/Motivation Test Battery (AMTB), contains 4-10 items to measure each scale. It is rather risky to go below 4 items per subarea because if the *post hoc* item analysis

(cf. Section 2.9.3) reveals that certain items did not work in the particular sample, their exclusion will result in too short (or single-item) scales. The technicalities of how to produce reliable and valid multi-item scales will be discussed in the section on "rating scales" (Section 2.4.1).

Of course, nothing is perfect. While multi-item scales do a good job in terms of psychometric reliability, they may not necessarily appeal to the respondents. Ellard and Rogers (1993) report that respondents sometimes react negatively to items that appear to be asking the same question because this gives them the impression that we are trying to "trick them or check their honesty" (p. 19). This problem, however, can be greatly reduced by using effective item-writing strategies (see Section 2.6, for a summary).