

## 2.9 PILOTING THE QUESTIONNAIRE AND CONDUCTING ITEM ANALYSIS

Because in questionnaires so much depends on the actual wording of the items (even minor differences can change the response pattern) an integral part of questionnaire construction is 'field testing,' that is, *piloting* the questionnaire at various stages of its development on a sample of people who are similar to the target sample the instrument has been designed for. These trial runs allow the researcher to collect feedback about how the instrument works and whether it performs the job it has been designed for. Based on this information, we can make alterations and fine-tune the final version of the questionnaire.

**Well...**

“if you do not have the resources to pilot-test your questionnaire, don’t do the study.”

(Sudman & Bradburn, 1983, p. 283)

The pilot test can highlight questions:

- whose wording may be ambiguous;
- which are too difficult for the respondent to reply to;
- which may, or should be, eliminated because, contrary to the initial expectations, they do not provide any unique information or because they turn out to measure something irrelevant;
- which – in the case of open-ended questions – are problematic to code into a small set of meaningful categories.

Piloting can also indicate problems or potential pitfalls concerning:

- the administration of the questionnaire;
- the scoring and processing of the answers.

Valuable feedback can also be gained about:

- the overall appearance of the questionnaire;
- the clarity of the instructions;
- the appropriateness of the cover letter (if there is one);
- the length of time necessary to complete the instrument.

Finally, this is also the phase when omissions in the coverage of content can be identified.

The importance of the piloting is in sharp contrast with the reality that so many researchers completely omit the pilot stage from their research design. Although this is understandable from a personal point of view because researchers at this stage are eager to get down to the survey and see the results, from a measurement perspective this practice is untenable. Regardless of how experienced the questionnaire designer is, any attempt to shortcut the piloting stage will seriously jeopardize the psychometric quality of the questionnaire (Moser & Kalton, 1971). Furthermore, my experience is that by patiently going through the careful editing procedures we can avoid a great deal of frustration and possible extra work later on.

Sometimes the omission of the pilot stage is not due to the lack of will/interest but rather to insufficient time. To do it well, piloting takes up a substantial period, which has often not been included in the timing of the research design. As we will see below, piloting is a stepwise process that, when properly done, can take several weeks to complete. This is usually much more than was originally intended for this phase of the research.

**Absolutely!**

“Questionnaires do not emerge fully-fledged; they have to be created or adapted, fashioned and developed to maturity after many abortive test flights. In fact, every aspect of a survey has to be tried out beforehand to make sure that it works as intended.”

(Oppenheim, 1992, p. 47)

So when and what shall we pilot? While it is useful to have ‘on-going piloting’ by continuously discussing every aspect of questionnaire design with a colleague or a friend, there are two points where a more formal trial run is needed: (1) Once the item pool has been com-

pleted, and (2) when a complete, almost final version of the questionnaire has been prepared.

### 2.9.1 Initial piloting of the item pool

The first time in the questionnaire construction process that some external feedback is indispensable is when we have prepared an initial item pool (cf. Section 2.6.1), that is, a large list of possible items, and we are ready to reduce the number of questions to the intended final number. The initial piloting of the item pool usually consists of the following steps:

- Select three or four people who are motivated to spend some time to help you and whose opinion you value. Some of them should not be specialists in the field – they are very useful in locating unnecessary jargon; others may be people who are accustomed to survey research or who know the target population well. In any case, as Converse and Presser (1986) so realistically state, at this stage we are likely to end up with “that familiar source of forced labor – colleagues, friends, and family” (p. 53).
- Ask them to go through the items and answer them, and then to provide feedback about their reactions and the answers they have given. The best method to conduct this phase is for you to be present while they are working: this way you can observe their reactions (e.g., hesitations or uncertainties) and can note and respond to any spontaneous questions or comments.
- Once they have gone through all the items, you may ask for any general comments and can initiate a brainstorming session.

It may be useful to provide your pilot group with some basic guidelines to focus on. These can include the following:

- They should mark any items whose wording they don’t like; if they can suggest an improvement, so much the better!

- They should mark any items whose meaning is not 100 percent clear; again, suggestions are welcome.
- They should mark any items that they consider unnecessary.
- They should try and think of anything else that might be worth asking about.

**Very important!**

“you may find that you have put so much personal time and effort into developing the questionnaire that it becomes ‘your baby.’ If someone is subsequently critical of it, you may find yourself reacting as if you have been personally attacked. Perhaps, rule number one in the critiquing/revision process is that the creator should never take the criticism personally.”

(Brown, 2001, p. 62)

**2.9.2 Final piloting (‘dress rehearsal’)**

Based on the feedback received from the initial pilot group we can normally put together a near-final version of the questionnaire that ‘feels’ OK and that does not have any obvious glitches. However, we still do not know how the items will work in actual practice, that is, whether the selected respondents will reply to the items in the manner intended by the questionnaire designers. There is only one way to find out: by administering the questionnaire to a group of respondents who are in every way similar to the target population the instrument was designed for. This is usually an ‘undeclared’ pretest whereby the respondents are not told that this is a questionnaire under construction. (Converse & Presser, 1986)



How big should this final pilot group be? It need not be very large; the typical sample size at this stage is around 50 (+/- 20). This number will already allow the researcher to conduct some meaningful item analysis, which is the next, and final, step in the questionnaire construction process. In addition, if the final piloting phase did not result in major changes, it may be possible to use at least some of the obtained data for the purpose of the 'real' investigation.

### 2.9.3 Item analysis

Item analysis can be conducted at two different points in the survey process:

- After the final piloting stage – in this case the results are used to fine-tune and finalize the questionnaire.
- After the administration of the final questionnaire – after such a '*post hoc analysis*' the results are used to screen out any items that have not worked properly.

The procedures in both cases are similar. They usually involve checking three aspects of the response pattern:

- (1) *Missing responses* and possible signs that the instructions were not understood correctly. If some items are left out by several respondents, that should serve as an indication that something is not right: Perhaps the item is too difficult, too ambiguous, or too sensitive; or perhaps its location in the questionnaire is such that it is easy to be overlooked. Also, a careful visual examination of the completed questionnaire might reveal some further response irregularities, for example in the way respondents marked their answers.
- (2) *The range of the responses* elicited by each item. It was argued in Section 2.6.2 that we should avoid including items that are endorsed by almost everyone or by almost no one because they are

difficult if not impossible to process statistically (since statistical procedures require a certain amount of variation in the scores). Although, as Brown (2001) remarks, the lack of variation may well be the true state of affairs in the group, it may be useful in many cases to increase item variation by adding additional response categories or rewording the question.

- (3) The *internal consistency* of multi-item scales. The gist of Section 2.3.2 was that – for the sake of reducing the unpredictable impact of item wording – questionnaires should contain multi-item scales, rather than single items, to focus on any particular content domain. It is obvious, however, that multi-item scales are only effective if the items within a scale work together in a homogeneous manner, that is, if they measure the same target area. In psychometric terms this means that each item on a scale should correlate with the other items and with the total scale score, which has been referred to as Likert's criterion of 'Internal Consistency' (Anderson, 1985). Following this principle, a simple way of selecting items for a scale is to compute correlation coefficients for each potential item with the total scale score and to retain the items with the highest correlations. There are also other, more sophisticated statistical methods to check and improve internal consistency – these will be summarized in Section 4.3.5.