

Principles of Survey Research

Part 3: Constructing a Survey Instrument

Barbara A. Kitchenham
 Department of Computer Science
 Keele University, Staffordshire, UK
 barbara@cs.keele.ac.uk

Shari Lawrence Pfleeger
 Systems/Software, Inc.
 Washington, DC, USA
 s.pfleeger@ieee.org

Abstract

In this article, we discuss how to construct a questionnaire. We point out the need to use any previous research results to reduce the overheads of survey construction. We identify a number of issues to consider when selecting questions, constructing questions, deciding on the type of question and finalizing the format of the questionnaire.

Key words: survey construction, question selection, question construction, question types, questionnaire format.

Introduction

In this installment of our series, we turn to how to construct a survey instrument. Survey instruments, which are usually questionnaires, are constructed using the following steps:

- Search the relevant literature.
- Construct an instrument.
- Evaluate the instrument.
- Document the instrument.

We discuss the first two steps in this article and the next two steps in the following article. As usual, we use the three surveys described in Part 1 of this series ([3], [4] and [5], plus the questionnaire in Appendix 1) to illustrate good and bad practice in software engineering survey research.

Searching the literature

As with any good investigative study, we must begin our work by looking through the literature. We need such searches to

- Identify what other studies have been done on the topic.
- Determine how the previous studies' researchers collected their data. In particular, we want to find out what questionnaires or other data collection mechanisms were used.

There are many reasons for knowing what has come before. First, we do not want *unknowingly* to duplicate someone else's research. Second, we want to learn from and improve upon previous studies. For example, if previous studies have developed relevant validated instruments or questions that we can adopt, it makes our own survey easier to administer and validate. Similarly, if other researchers had problems with response rates, we will be aware of the need to adopt measures to address this problem. Finally, other studies may give us ideas about variables and issues we need to consider in designing our own studies.

Getting started: new or reused?

In software engineering, we often start from scratch, building

models of a problem and designing survey instruments specifically for the problem at hand. However, in other disciplines, it is rare to develop a new survey instrument. Researchers usually rely on using existing instruments, perhaps tailored slightly to accommodate variations on a common theme. This reliance on standard instrumentation has two important advantages.

1. The existing instruments have already been assessed for validity and reliability.
2. By using common instruments, it is easy to compare new results with the results of other studies.

When researchers in other disciplines cannot use an existing instrument, they are often able to amend existing instruments. An instrument might be amended if

- It is too long to be used in entirety.
- A different population is being studied from the one for which the original instrument was designed.
- It needs to be translated.
- The data collection method is different in some way from the original instrument's data collection.

However, we must take care when considering amending an instrument. Our changes may introduce complications that make the research more difficult. For example,

- If the original instrument is copyrighted, we may need permission to change it.
- We must repeat pilot testing of the instrument.
- The new instrument must be assessed for validity and reliability.

Unfortunately, because most survey instruments in software engineering research are developed from scratch, we introduce these and many other problems. In particular, software engineering research instruments are seldom properly validated. For these reasons, the remainder of this article focuses on the care needed in designing a new survey instrument.

Question selection

A survey asks the respondents to answer questions, so our starting point in designing the survey instrument should always be the survey's purpose and objectives. These define the context of the survey and tell us what areas to cover and what types of information we need. In general, the purpose of a survey is wider in scope than the survey's objectives. The purpose concerns the hoped-for outcomes of the survey and so constrains the scope of questions. The more clearly the purpose is stated, the better. The specific objectives of the survey refer to the precise information the survey is to collect. Objectives may sometimes be formulated as research questions or research

hypotheses, but they may also be just simple statements. No matter how the objectives are stated, it is essential that the survey questions relate directly to the survey objectives.

When deciding what to ask in the survey instrument, there are several other factors to be considered:

Understanding the respondents.

Questions must be formulated in a way that respondents can answer them easily and accurately. For example, we can make assumptions about reading ability or level when surveying computer science graduates that may be inappropriate when surveying users of software-intensive products. It is also important to make sure that respondents have sufficient knowledge to answer the questions. It can be extremely frustrating to be asked questions you are not in a position to answer. For example, of the three surveys described in Part 1 of this series, two of the surveys (Lethbridge's surveys [3], [4] and the Finnish survey [5]) asked respondents about their personal experiences. In contrast, our survey of technology adoption asked respondents to answer questions such as

"Did your company evaluate this technology?" Yes/No

"Are you now using the technique in some production work or most production work"? Yes/No

In this case, we were asking people to answer questions on behalf of their company. The questions may have caused difficulties for respondents working in large companies or respondents who had worked for the company only for a relatively short period of time.

We must also take care to consider the timeframe from which we want the respondents to draw their answers. We may get inaccurate responses if we ask people about events that happened long in the past. Moreover, we must make the timeframe clear. For example, suppose the technology under consideration is object-orientation. If we ask

"Did your company evaluate this technology?" Yes/No

we may find that the company evaluated OO before there were supporting tools and methods. It may not be fair or suit our objectives to compare the answer from this respondent with one who evaluated the technology after the supporting tools and methods were available. Thus, it is essential that we specify the time period from which we would like our answers drawn.

Asking an appropriate number of questions.

Each of us has, at one time or another, received a questionnaire that is daunting simply in terms of the number of questions posed. For example, the long form of the U.S. census has dozens of questions and can take a very long time to fill out. Thus, it is important to keep in mind that the number of questions you can realistically ask in a survey depends on the amount of time respondents are willing to commit to it. If we undertake a cold-call telephone survey on a general topic, we should not expect to require more than 10 minutes of the respondent's time. In a self-administered questionnaire, we can ask more questions, but we need to be sure that we do not overload respondents; if we do, they may provide answers that are expedient but not particularly accurate or thoughtful. Some people may be willing to spend 30 to 60 minutes answering a questionnaire. However, unless the topic of the questionnaire is

of paramount importance to them, they are not likely to spend more than two hours. Remember that the length of time to respond is not always related to the absolute number of questions. For instance, Lethbridge's survey is very long. Although he asked only four main questions, he asked them about 75 different subjects and added a further 12 demographic questions.

If we have too many questions, we may need to remove some. Questions can usually be grouped together into topics, where each topic addresses specific objectives. One way to prune questions is to identify a topic that is addressed by many questions, and then remove some of the less vital ones. Another way is to remove some groups of questions. Keep in mind, though, that such pruning sometimes means reducing the objectives that the questionnaire addresses. In other words, you must maintain a balance between what you want to accomplish and what the respondents are willing to tell you.

Standardizing response formats.

One way to reduce the time taken to complete a survey is to have standardized response formats. For example, in attitude surveys, responses are usually standardized to an ordinal scale of the form:

Strongly Agree, Agree, Disagree, Strongly Disagree.

If all responses are standardized, respondents know their choices for each question and do not have to take time to read the choices carefully, question by question. Thus, respondents can usually answer more standard-format questions in a given time than non-standard ones.

Question construction

Once we have an idea of what we want to ask, we must give some thought to how we want to pose the questions. We have seen that standardized responses can be useful. There are additional guidelines to help us formulate the questions we ask. In particular, we want our questions to be purposeful and concrete.

Purposeful questions

The questions should be worded so that the respondent can see relationship between the intention of question and survey objectives. If the purpose of the question is not clear, the respondent may very well ignore it or provide a less-than-thoughtful response.

Concrete questions.

Each question must be precise and unambiguous. Often, we need to add clarifying detail; for instance, we may include a time period to make a given question more precise. We must pay particular attention to wording to avoid two-edged questions; that is, if a question contains two different ideas, it may confuse the respondent.

Improving question construction

There are several steps we can take to improve the general construction of our survey questions. First, we can use time periods that are related to the importance of the question. In her book about asking survey questions, Arlene Fink [2] recommends that the timeframe of a year or more is acceptable for questions dealing with major life events. However, less

important matters may draw from periods of a month or less.

Second, we must be sure to use conventional language. Not only does this mean relying on standard grammar, punctuation and spelling, but also using complete sentences that express a single idea. Never assume that your respondents know your terminology; if necessary, define a set of terms for them, to be sure that you are working with the same understanding of vocabulary. For example, the words "release" or "version" may have very particular meanings to someone with a background in configuration management, and very different meanings to someone who is a tester or requirements analyst.

To see how wording can affect results, consider the two Lethbridge surveys. Each was on the same topic, but he changed his last question about each subject. In the first survey [3], question 4 was:

"How useful would it be (or have been) to learn more about this (e.g. additional courses)?"

In his second survey [4], question 4 was:

"How much influence has learning the material had on your thinking (i.e. your approach to problems and your general maturity), whether or not you have directly used the details of the material? Please consider influence on both your career and other aspects of your life."

The first version of the question is considerably better than the second version, because the second version is more complex and thus more difficult to interpret and understand. In particular, the second version appears to be two-edged (referring both to approach to problems and to general maturity) and rather imprecise (since it may not be clear what "general maturity" really means).

When wording questions, we should try avoid the use of abbreviations, slang and colloquial expressions. Similarly, we should be careful about using jargon and technical expressions.

As you can see in Appendix 1, our survey of technologies posed questions about evaluation procedures in terms of how the respondent's company performed its evaluation studies. In particular, we asked questions about soft and hard evaluation techniques by defining them at the top of two of the columns:

Soft evaluation techniques: Read case studies, articles, talking with peers, lessons learned or other more anecdotal evidence? Yes/No

Hard evaluation techniques: feature comparison, performance benchmark, or other more quantitative evidence? Yes/No

As you can see, these questions include jargon terms related to evaluation that may not be well-understood by the potential respondents. Similarly, if you examine the list of technologies listed in the left-hand column of our questionnaire, you will see that we use jargon there as well: CASE tools, Rapid Application Development, 4GLs, and more. Were we to design questions for such a survey now, we would spell out each technology and include a glossary to describe each one. Such information gives us a level playing field, making sure that the respondents have a common understanding of the terminology we use.

One final caution: We should avoid using negative questions or

statements, such as "Did you not use inspections because you had inadequate evidence of their effectiveness?" Negative questions are usually more difficult to understand than straightforward ones, since they require an exercise in logical reasoning.

Question Types

When formulating questions for a survey instrument, you can express them in one of two ways: open or closed. A question is *open* when the respondents are asked to frame their own reply. Conversely, a question is *closed* when the respondents are asked to select an answer from a list of predefined choices.

There are advantages and disadvantages to each type of question. Open questions avoid imposing any restrictions on the respondent. However, there are many different ways respondents may choose to answer a question. Moreover, no matter how carefully we word the question, open questions may leave room for misinterpretation and provision of an irrelevant or confusing answer. Thus, open questions can be difficult to code and analyze.

By restricting respondents' choice of reply, closed questions are easier to analyze. In fact, closed questions are the preferred format for mail surveys. However, if we want to use closed questions in our survey, we may need to include a "None of the above" or "Don't know" answer when asking categorical questions. If we are using an ordinal scale, we may need to include a neutral response, such as "No preference," as the mid-point of the scale. There is some disagreement in the social science community about whether to include neutral or "don't know" responses. Some researchers feel that such choices allow respondents to avoid answering a question. In our view, it may be counter-productive to force people to answer questions they don't want to, or to force them to make a choice about which they feel ambivalent. However, your decision about using a neutral category must depend on the particular requirements of the survey.

With closed categorical questions, the respondent chooses one or more responses to a question. If we want the respondent to provide only one answer, it is important to ensure that the categories are mutually exclusive. That is, it should be impossible for an answer to fit into two or more categories at the same time. For example, suppose we are asking questions about defects, and we provide a set of defect type categories that includes "design defect" and "interface defect" but allows only one choice. This situation causes problems for a respondent who wants to report an interface defect that is also a design defect.

We cannot emphasize enough the need for all response categories to be clearly understood. For attitude surveys, we can usually assume that terms like "Strongly agree," "Agree" and so on are well-understood and do not need further definition. However, understanding may not be obvious if we construct our own scales. For example, Lethbridge gives some indication of the detail needed to define an ordinal scale in his survey. Each of his four main questions has its own associated ordinal scale with responses defined in the context of the question. For instance, the question "How much did you learn about this at university or college" had the following scale:

Score	Definition
-------	------------

- 0 Learned nothing at all
- 1 Became vaguely familiar
- 2 Learned the basics
- 3 Became functional (moderate working knowledge)
- 4 Learned a lot
- 5 Learned in depth, became expert (learned almost everything)

Although the intermediate points on the scale are a little vague, the end points are clear and unambiguous. The respondents were asked to use the numerical value to respond to the questions. The numerical value is referred to as the "score" for the response category. Lethbridge's scale conforms to the normal standard of using between 5 and 7 choices along an ordinal scale. Ordinal scale data can cause analysis difficulties, which will be discussed in later article.

Lethbridge's scale is a reasonably balanced one. A scale is *balanced* when the two endpoints mean the opposite of one another *and* the intervals between the scale points appear to be about equal. Creating equal distances between the scale points is called anchoring the instrument. It is difficult to create an anchored scale and even more difficult to validate that a scale is properly anchored.

Most questionnaires include demographic questions (that is, questions that describe the respondent). Many survey instruments put them at the front of the questionnaire. However, Bourke and Fielder [1] suggest putting them at the end instead. They point out that demographic details may be off-putting at the start of the questionnaire and so may discourage respondents.

Questionnaire Format

For self-administered questionnaires, it is important to consider both the format of the questionnaire and the questionnaire instructions. For formatting printed questionnaires, use the following checklist (much of which applies to Web-based questionnaires, too):

- Leave a space for the respondents to comment on the questionnaire.
- Use space between questions.
- Use vertical format, spaces, boxes, arrows, etc. to maximize the clarity of questions. However, do not overwhelm the respondent with "clever" formatting techniques (particularly for Web Questionnaires).
- Consider the use of simple grids.
- Consider the use of a booklet format.

- Have a good contrast between print and paper.
- Stick to a font size of 10-12.
- Use a font that is easy to read.
- Avoid italics.
- Use bolding, underlining or capitals judiciously and consistently for emphasis and instructions.
- Do not split instructions, questions and associated responses between pages.

For information and instructions, use the following checklist:

- Explain the purpose of the study.
- Describe who is sponsoring the study (and perhaps why).
- Include a cover letter using letterhead paper, dated to be consistent with the mail shot, providing a contact name and phone number. Personalize the salutation if possible.
- Explain how the respondents were chosen and why.
- Explain how to return the questionnaire.
- Provide a realistic estimate of the time required to complete the questionnaire. Note that an unrealistic estimate will be counter-productive.

Many of these issues are discussed in more detail in [1].

There are still many issues to be addressed before questionnaire construction is completed. We need to be sure that we have done our best to motivate potential respondents and we have avoided personal bias. Most importantly, we need to evaluate our instrument. These issues are discussed in the next article in this series.

References

- [1] Linda Bourque and Eve Fielder, *How to conduct self-administered and mail surveys*, Sage Publications Inc., 1995.
- [2] Arlene Fink, *How to Ask Survey Questions*. Sage Publications Inc., 1995.
- [3] Timothy Lethbridge, 'A survey of the relevance of computer science and software engineering education, *Proceedings of the 11th International Conference on Software Engineering*, IEEE Computer Society Press, 1998.
- [4] Timothy Lethbridge, 'What knowledge is important to a software professional, *IEEE Computer*, May 2000.
- [5] J. Ropponen and K. Lyytinen, Components of software development risk: How to address them. A project manager survey, *IEEE Transactions on Software Engineering* 26(2), February 2000.

Appendix 1: Our technology evaluation questionnaire

Dear Executive,

We are sponsoring a study for (names of researchers). It is only through our cooperative efforts with the academic community that we bring our commercial experiences to the classroom. Thank you for your help.

Please send a copy of the results to the attention of (name) at the address on the reverse.

This is a study of how organizations evaluate system development technologies before adopting them. If you are not sure or don't

know an answer, just leave the line blank. Otherwise, it is important to answer YES or NO to the first section of every Technology/Technique section.

Technology/Technique	Did your company evaluate this technology?		Soft Evaluation techniques: read case studies, articles, talking with peers, lessons learned, or other more anecdotal evidence?		Hard Evaluation techniques: feature comparisons, performance benchmark, or other more quantitative evidence?		Are you now using the technique in some production work or most production work?		
	Yes	No	Yes	No	Yes	No	Some	Most	None
4GLs	Yes	No	Yes	No	Yes	No	Some	Most	None
Automated project management tools	Yes	No	Yes	No	Yes	No	Some	Most	None
Back end lower CASE tools	Yes	No	Yes	No	Yes	No	Some	Most	None
Inspection/walkthroughs	Yes	No	Yes	No	Yes	No	Some	Most	None
Database config. mgmt., repository tools	Yes	No	Yes	No	Yes	No	Some	Most	None
Data warehousing	Yes	No	Yes	No	Yes	No	Some	Most	None
Preliminary feasibility study	Yes	No	Yes	No	Yes	No	Some	Most	None
Formal life cycle methodology	Yes	No	Yes	No	Yes	No	Some	Most	None
Front end upper CASE tools	Yes	No	Yes	No	Yes	No	Some	Most	None
Graphical user interface (GUI) builders	Yes	No	Yes	No	Yes	No	Some	Most	None
Joint Application Development	Yes	No	Yes	No	Yes	No	Some	Most	None
Knowledge-based tools/Expert systems/Fuzzy logic	Yes	No	Yes	No	Yes	No	Some	Most	None
Mega-packages (SAP, PeopleSoft)	Yes	No	Yes	No	Yes	No	Some	Most	None
Metrics and data collection	Yes	No	Yes	No	Yes	No	Some	Most	None
Middleware	Yes	No	Yes	No	Yes	No	Some	Most	None
Object orientation design and code	Yes	No	Yes	No	Yes	No	Some	Most	None
Outsourcing of software development	Yes	No	Yes	No	Yes	No	Some	Most	None
Post-deployment or post-mortem reviews	Yes	No	Yes	No	Yes	No	Some	Most	None
Prototyping	Yes	No	Yes	No	Yes	No	Some	Most	None
Rapid Application Development	Yes	No	Yes	No	Yes	No	Some	Most	None
Reuse of existing components	Yes	No	Yes	No	Yes	No	Some	Most	None
Software process improvement activities	Yes	No	Yes	No	Yes	No	Some	Most	None
Use of COTS components	Yes	No	Yes	No	Yes	No	Some	Most	None