This section aims at providing some basic knowledge on designing cost-effective survey studies. It presents a simple operative framework stating the basic steps required for *reliable* and valid *results*. The concept of *reliability* refers to the extent to which a measurement can be replicated. It concerns the problem of generalizing results of a survey to other users, conditions, times, and places. Reliability means consistency and is a logical precondition to *validity*. This last concept refers to the appropriateness of the survey to measure exactly what the researcher intends. It concerns whether the concepts being investigated are actually the ones being measured.

Generally speaking, survey researches sample many respondents who answer the same set of questions. The method follows a deductive approach. It begins with a research problem and ends with empirical measurements and data analysis. A schematic representation of the process is presented below. The framework is based upon basic methodological knowledge coming from the social sciences (Breakwell et al. 1995; Neuman 1997; Oppenheim 1996) revised to fit the need of system designers. It concentrates specifically on survey studies employing questionnaires as the principal means of data collection.

- *Establishing goals*. The first step in any survey study is clearly stating what the researcher wants to measure and why. Well-defined research goals are instrumental to good questionnaire design and effective data analysis. Goals have to be expressed in a few clear and concise sentences, along with working hypotheses and expected results. When the survey is focused on the collection of user requirements, this step has to be accomplished by an interdisciplinary team: computer scientists and designers indicate the knowledge they need to gather; social scientists indicate the best way of collecting them.

- *Sampling procedure*. Once goals and hypotheses have been defined, the researcher has to choose the type of respondents and the population. Sampling procedures can be more complex that one might expect (for an introduction, see Fife-Schaw 1995), in HCI research sampling is normally achieved by similarity with the target users. Poor sampling introduces unknown sources of error in the survey so that statistical results may be difficult to interpret and to generalize. At this stage, the researcher should also define the procedure of data collection and the statistical techniques she will use to analyze the data.

- *Developing the survey instrument*. Building a questionnaire is an iterative process, beginning with the definition of salient thematic areas and ending only when the researcher feels confident that the questionnaire can do the job for which it has been designed. Questionnaires have to be created or adapted passing through a number of refinement steps, in which they are tried out, improved, and then tried out again. The definition of the thematic areas strongly benefits from an interdisciplinary approach. The translation of these areas into questions should be cared by social scientists.

- *Collecting the data*. In HCI, questionnaires are normally self-administered (the respondents personally fill it). Interviews schedule are administered by an interviewer who asks the questions and register the answers. Beside this basic distinction, there are a number of ways for collecting data (e.g., telephones web-based and e-mail surveying). They strongly affect the content, the format and the layout of the questionnaire.

- *Analyzing the data.* Data must be carefully analyzed bearing in mind the class of measurement they belong to and their particular distribution. As a general rule, survey studies belong to the class of correlational research. This is intended to describe the extent to which variations in factor A are related systematically to variations in factor B. It is impossible to imply causality from correlation. An example may help illustrate this point. There is a strong correlation between attitudes towards and experience with technology. People who hold positive attitudes are normally more experienced than people who hold negative attitudes. Does this mean that the experience caused the attitude, or vice versa that the attitude caused the experience? On the basis of the correlation alone we cannot say. Both hypotheses might be true. We can only conclude that there is an association between attitudes and experiences, so that the two variables tend to increase together. To establish causality requires controlled experiments.

- *Interpreting the data*. Once the data have been analysed, the role of the researcher is in interpreting the meaning and determining the generality of the findings. If the survey is focused on the collection of user requirements, this step has to be accomplished by an interdisciplinary team. The phase ends with the generation of specific system requirements.

4.1 Questionnaire design

Questionnaires are a relatively well understood method of data gathering, which allows collecting large amount of information at a low cost. They are probably the most widely used data-gathering techniques in sociology, social psychology and marketing. In the last decade, they have gained increasing importance also in HCI as a means of assessing user satisfaction. The most known examples are the QUIS questionnaire developed by Norman Kent at the University of Maryland (Shneiderman 1992) and the WAPI questionnaire recently developed for assessing WEB usability. There are numerous guides to questionnaire design, explaining the many technicalities underlining their proper use. Yet, devising effective questionnaires requires experience, sensibility and pilot testing. It is fundamental to understand that no questionnaire can ever provide perfectly accurate measurement because questionnaires collect self-report data. Therefore, they gather estimation rather than direct measurement, as would be the case of direct observation.

Due to the lack of correspondence between questionnaire estimation and direct measurement, the discipline of psychometrics has developed a detailed set of procedures and models for statistical examination of the questionnaire reliability. The discussion of these theories is far beyond the scope of this paper (interested readers can refer to Frederiksen et al. 1993). Nevertheless, questionnaire score might be affected by errors deriving by both systematic and unsystematic sources. The systematic error is built into the questionnaire. It can be due to mistakes in the design of the instruments, such as to ambiguous questions or inappropriate rating scales. In addition, it can be due to uncontrolled variables affecting the questionnaire score. Given that it is impossible to control for every variable that influences human behavior in any given situation, part of the task of the questionnaire designer is to control for variables that represent as much of the variance as possible. The systematic error affects the questionnaire scores consistently in one direction. Hence, it can induce overestimation or underestimation.

The unsystematic error refers to variability sources, which are external to the questionnaire itself (e.g., the mood of the respondents, his/her level of attention, the environment where the questionnaire is administered). Hence, unsystematic error is assumed to be random, such that it might equally results in overestimation or underestimation. The classical test theory is built upon the assumption that the questionnaire has been constructed with sufficient care that systematic error is negligible and only unsystematic error exists. In the following we will briefly summarize a number of guidelines for minimizing systematic errors.

The systematic error generates response biases. This term describes the situation in which respondents fail to answer the question accurately. There may be many reasons for response biases, varying on the degree of faking intentionality. Biases can be induced by motivational factors, such as social desirability. Another common bias is known as *response set*. Here, the idea is that respondents possess an in-built tendency to choose a specific answer (e.g. a preference for "yes" over "not" or for the middle point of a scale). There is no simple solution to the problem of biases, although there are some ways of reducing it. A good questionnaire is an integrated whole composed by introductory remarks, filling instructions, questions, and answering format. All together, these elements contribute to reduce biases.

Adequate question wording is instrumental (Neuman 1997). Questions should be written in simple and plain language, avoiding slang and abbreviations, double negatives and hypothetical construction. In this context, it is worth noticing that words have implicit connotative as well as explicit denotative meanings. Both the aspects have to be controlled in order to avoid inducing inappropriate answers. Emotional language has to be banned by questionnaire. Finally, each question should be accompanied by a precise format of answer, with exclusive and balanced response category. More truthful answers can be obtained also by guaranteeing participants that their responses are anonymous and will not have any direct influence on their personal life. For the same reason, is worth stressing that there are not right or wrong answers. Moreover, more items measuring the same concept at different point of the questionnaire allows testing for consistency in reporting. The idea is that the probability of logically inconsistent reports is much higher when respondents are misrepresenting their behavior, than when they are correctly reporting it. A strategy for reducing the possibility of a response set, is introducing questions that are reversed in meanings from the overall direction of the scale (reversal items).

localuser 3/9/00 22:55
Deleted: es

localuser 3/9/00 22:58
Deleted: It should come as no surprise that sensitive and socially undesirable behaviors are often misreported, if reported at all.