4

Descriptive Study I: Understanding Design

Design methodologies emphasise the importance of investigating the needs of the users and understanding the situation a product is supposed to improve, in particular when this situation is complex and failure of the product is expensive or unacceptable. Developing support for design is no different; designing is a complex activity, and failure of support can be expensive in terms of time, people and money and can have a large effect on practice. Descriptive Studies help understand this complex activity and should provide a sound basis on which to develop support.

This chapter focuses on the second stage of DRM: the DS-I stage. It discusses how, starting with the deliverables from the RC stage – the Initial Reference and Impact Models, the preliminary Criteria and the Overall Research Plan – sufficient understanding of the topic of interest and of the factors that determine its success can be obtained, such that areas for which development of support is realistic and effective can be identified with confidence.

All types of design research will require a DS-I stage to obtain sufficient understanding of the current situation, *i.e.*, to complete the Reference Model. Depending on the research goal (descriptive, prescriptive or evaluative) DS-I may be limited to a detailed review of the literature in potentially relevant areas (as illustrated in the ARC diagram, Section 3.6) or may be more comprehensive, involving a literature review as well as one or more empirical studies.

Referring back to Section 2.6.2, the objectives of the DS-I stage are:

- to obtain a better understanding of the existing situation by identifying and clarifying in more detail the factors that influence the preliminary Criteria and the way in which these factors influence these Criteria;
- to complete the Reference Model including the Success Criteria and Measurable Success Criteria;
- to suggest the factors (possible Key Factors) that might be suitable to address in the PS stage, as these are likely to lead to an improvement of the existing situation;
- to provide a basis for the PS stage for the effective development of support that addresses those factors that have the strongest influence on success, and can be assessed against the Criteria;

- to provide detail that can be used to evaluate the effects of the developed support in the DS-II stage.

The deliverables of the DS-I stage are:

- a completed Reference Model, Success Criteria, Measurable Success Criteria and Key Factors, that:
  - describe the existing situation and highlight the problems;
  - show the relevance of the research topic;
  - clarify and illustrate the main line of argumentation; and
  - point at the factors that are most suitable to address in order to improve the situation;
- an updated Initial Impact Model;
- implications of the findings for the development of support and/or for the evaluation of existing support.

In this book the term ‘Descriptive Study’ or ‘DS’ refers to the two stages of DRM that focus on obtaining a better understanding of the current situation. All the different types of *empirical studies* that can be used to investigate (*describe*) the phenomenon of design can be involved. A Descriptive Study thus covers the three types of studies distinguished in the Social Sciences: exploratory, descriptive and explanatory (Yin 1994).

- An *exploratory study* answers ‘what’, ‘who’, ‘where’ questions, and is intended “to develop pertinent hypotheses and propositions for further inquiry”, that is, to help find a research focus when the understanding is still insufficient or lacking.
- A *descriptive study* also answers ‘what’ questions, but of the type ‘how many’ and ‘how much’, because it is aimed at “describing the incidence or prevalence of a phenomenon or to be predictive about certain outcomes”.
- *Explanatory studies* are used to answer ‘how’ and ‘why’ questions, *i.e.*, “questions that deal with operational links needing to be traced over time, rather than mere frequencies or incidence”.

We will continue to use the term *Descriptive Study* (with capitals) in our methodology to represent the stages in DRM and use the term *empirical study* to represent the nature of the actual investigation, which can be exploratory, descriptive or explanatory, as necessary.

4.1 Schools of Thought

When designing products, the design team usually draws upon support from a variety of domains – such as machine elements, mechanics, materials, ergonomics, marketing, mathematics, cognitive sciences, and economics – in varying degrees depending on the particular characteristics of the problem to be solved. Each domain has its own terminology, theories, approaches (methodologies), rules for verification, *etc*., but only a collaborative effort will result in the best solution. In a similar way, to investigate complex phenomena such as design (involving products, people, teams, tools, organisations and their micro- and macro-economic context) one has to draw upon research methods from a variety of disciplines – such as engineering sciences, social sciences, natural sciences, management science, *etc*., – depending on the focus of interest.

Research in these disciplines bases itself on a vast body of knowledge, and the methodologies and methods used are based on specific paradigms. *Paradigms* are worldviews or belief systems that guide researchers by defining the topic of research and the type of research questions, as well as the research process – for example the role of the researcher in the data-collection process – and thus determine the details of the research methodology and methods applied. Paradigms change over time and new ones emerge. Competing paradigms may exist simultaneously; specifically in less mature sciences (Kuhn 1970) (see Appendix A.1 for more details).

When adopting research methodologies and methods, as well as the related terminology, models, theories and other elements from other disciplines, it is important to be aware of the underlying paradigms, as these might constrain, or put requirements upon, their application for investigating design as well as their use in combination with other methods. As a design researcher it is not necessary to join in debates about the best methodology, but it is important to read primary sources about potentially suitable approaches and methods before making a choice (Section 4.6). This will ensure that the data obtained and conclusions drawn are valid for the purpose intended and that pitfalls in applying these are avoided.

In this section, we address two of the issues raised in these disciplines that are particularly relevant for design research. Our main objective here is to raise awareness. Further literature needs to be consulted.

What Comes First: Theory or Observation?

Many definitions of theory and several different kinds of theory exist. Following the definitions of the social science researchers Frankfurt-Nachmias and Nachmias (1996) scientific theories are abstractions representing certain aspects of the empirical world; they are concerned with the how and why of empirical phenomena, they therefore help us explain and predict phenomena of interest. They are not concerned with what *should* be. Note that for our purpose – to understand as well as improve design – we need to determine 'what is' as well as 'what should (and could) be'.

Theories can be classified in various ways. The classification we found useful is from Parsons and Shils, quoted in Frankfort-Nachmias and Nachmias (1996).

- *Ad hoc* classificatory systems: arbitrary categories – categories not based on a more general theory – that organise and summarise empirical data.
- Taxonomies: systems of categories constructed to fit empirical observations. Taxonomies enable researchers to describe relationships among categories.
- Conceptual frameworks: descriptive categories are systematically placed in a structure of explicit, assumed propositions. The propositions included

within the framework summarise and provide explanations and predictions for empirical observations. They are not established deductively, however.

- Theoretical systems: combine taxonomies and conceptual frameworks by relating descriptions, explanations, and predictions systematically. The propositions of a theoretical system are interrelated in a way that permits some to be derived from others. A specific theoretical system is the formal or axiomatic theory, based on direct causal relationships between concepts that are not testable but stated as being true, the so-called *axioms*.

Regarding the role of theories in research, two main schools of thought exist: (1) starting with a theory, developing hypotheses and then doing empirical research to test these hypotheses, and (2) using the data from empirical research to develop hypotheses and theories. Meanwhile, many scientists agree that in reality these two approaches do not occur in their 'pure' forms – a view we fully support.

Most common is the first school of thought: a *theory-driven* approach. Denzin, *e.g.*, emphasises theories as starting point when he defines research (in his case sociological research) as "those endeavours which take the sociologist from the vague realm of theory to substantive issues in the empirical social world" (Denzin 1978). Frankfort-Nachmias *et al.* highlight the importance of theory as "affecting each stage and being affected by each stage" of the research process (Frankfort-Nachmias and Nachmias 1996). "The process starts with a problem about which tentative generalisations, or hypotheses, are formulated that are then tested *logically* and *empirically*". These, they call, *validation* and *verification*[9] respectively. The more mature a discipline, the more one can build upon existing theories and hypotheses. It is, however, questionable whether the 'pure' approach of starting with a theory or hypotheses can exist, because their initial formulation requires at least some research (see also the discussion in Reich (1995)).

A clear representative of the second school of thought is the *data-driven*, Grounded Theory, approach, where theories are grounded in empirical data. In its original form the researcher is advised "to ignore the literature of theory and facts on the area under study, in order to assure that the emergence of categories will not be contaminated" (Glaser and Strauss 1967). Apart from this being inefficient, it is questionable whether even those researchers attempting to follow such an approach would not at least have a belief that what they are studying is worth doing, which involves assumptions about a possible interesting outcome. In design research this would involve beliefs about a link to success or potential for improvement. In general, it is now accepted, also by the founders of Grounded Theory,[10] that "observation of the world and what happens in it, whether or not aided by instruments, is never free of the theories, beliefs, assumptions and expectations

---

[9] Note that the definition of validation and verification is used differently in other disciplines. In computer science, *e.g.*, the terms are used in the opposite sense. Validation is to ensure that you built the right thing; verification is to ensure that you built the thing right.

[10] Strauss, *e.g.*, states that one can use another's ideas to build complex concepts without violating the grounded theory notion of empirical faithfulness, see Strauss (1970) in (Star 1997).

brought to the task by the observer himself" (Bullock *et al.* 1988). This is called *theory-ladeness*, which covers both the process of observation as well as the terms in which what is observed are described (Bullock *et al.* 1988).

Quantitative or Qualitative

Much has been written about the differences between qualitative and quantitative research. Some authors refer to the type of questions addressed, others to the type of data collected, to the analysis methods used, or to the whole research approach.

Authors such as Frankfort-Nachmias and Nachmias (1996) and Kelle (1997) link quantitative and qualitative research directly to the theory-driven and data-driven approach, respectively. "Quantitative research uses deduction by deriving hypotheses from theory and analysing the data they collect to statistically test the hypotheses. […] Qualitative field research moves in the opposite direction, using a process called analytic induction: collect data, formulate hypotheses based on data, test hypotheses using data and attempt to develop theory. This theory is called Grounded Theory." (Frankfort-Nachmias and Nachmias 1996). "Unlike hypothetical-deductive research, such a theory that consists of empirically contentful statements is not the starting point of the qualitative research process, but its result" (Kelle 1997). "Scientists must gain an empathic understanding of societal phenomena, and they must recognise both the historical dimension of human behaviour and the subjective aspects of the human experience" (Frankfort-Nachmias and Nachmias 1996). Even though other authors do not directly link quantitative and qualitative to the research approaches, the methods they propose tend to be either theory-driven or data-driven.

We use the terms quantitative and qualitative to express the goal of a particular research question or hypothesis. A *quantitative* approach is applied to investigate or measure *the degree in which phenomena occur*. Methods used are experiments, observations, closed questionnaires, *etc*. The methods are generally well formulated and established, and based mainly on statistics. Quantitative research produces the type of data common to engineering, and engineering design researchers usually learn, how to collect and analyse this type of data using experiments and statistics, how to interpret the findings and how to avoid bias. Examples of such data in design research are design time, number of errors, number of components, percentage of returns, number of warranty claims, *etc*.

A *qualitative* approach is applied to investigate *the nature of phenomena*. Methods used are interviews, observation and written documents, such as open-ended items on questionnaires and diaries (Glaser and Strauss 1967; Patton 1987; Wester 1987). Researchers talk about 'rich descriptions', 'sensitive interpretation', 'growing understanding', all pointing to the different nature of the qualitative research process and its ways of data-handling. As Kelle (1997) writes: "The theoretical knowledge of the qualitative researcher does not represent a fully coherent network of explicit propositions from which precisely formulated and empirically testable statements can be deduced. Rather it forms a loosely connected "heuristic framework" of concepts which helps the researcher to focus his or her attention on certain phenomena in the empirical field". Qualitative data in design research would include sketches, arguments and decisions, gestures, designer opinions, *etc*.

Increasingly, both qualitative and quantitative approaches are combined to obtain a full picture of the object of study, see, *e.g.*, the discussion in Tashakkori and Teddlie (1998). In our opinion it is this combination that provides the richest picture, addressing the various factors involved in the phenomenon of design using the method that is most suitable for each of these. After all, as Einstein said "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted."

4.2 Types of DS-I

In Figure 3.12 two types of DS-I were identified:

- A **Review-based DS**-I, which involves a detailed review of the literature in both the area of research and other potentially relevant areas, as illustrated in the ARC diagram (see Section 3.6). A Review-based DS-I will only cover Steps 1 and 5 in the DS-I process outlined below.
- A **Comprehensive DS-I**, which involves a literature review as well as one or more empirical studies. The empirical studies take place when the literature review shows a lack of understanding about the chosen topic, or when particularly relevant links in the Initial Reference or Impact Models are still poorly understood.

4.3 DS-I Process Steps

For a systematic approach to the planning and execution of a Comprehensive DS-I, the following steps are proposed (see Figure 4.1), which will be described in more detail in Sections 4.4 to 4.8:

1. Reviewing the literature (also for Review-based DS-I). This involves determining the existing level of understanding and, based on this, adapting the Initial Reference Model (and Initial Impact Model where relevant).
2. Determining research focus. This involves identifying and defining factors and links of interest, as well as extending and refining the initial research questions and/or hypotheses.
3. Developing research plan for DS-I.[11] This involves selecting and developing research method(s) and combining these into one or more studies, developing any necessary material and infrastructure to be used, undertaking a pilot study, and adjusting the research plan, method(s) and material;
4. Undertaking empirical study. This involves collecting data, processing data, analysing and interpreting data, verifying the results and drawing

[11] We opted for the term 'research plan' rather than the commonly used term 'research design' in order to avoid confusion with our domain 'design research'.

conclusions. Furthermore, the results are used to update the Reference Model, and to plan for further empirical studies, if not already foreseen.

5. Drawing overall conclusions (also for Review-based DS-I). This involves combining the results of the various studies, modifying and completing the Reference Model and updating Initial Impact Model. Furthermore, suggestions or concepts for support are proposed, and the next stage (continue DS-I, go to PS or revisit RC) and future work determined.

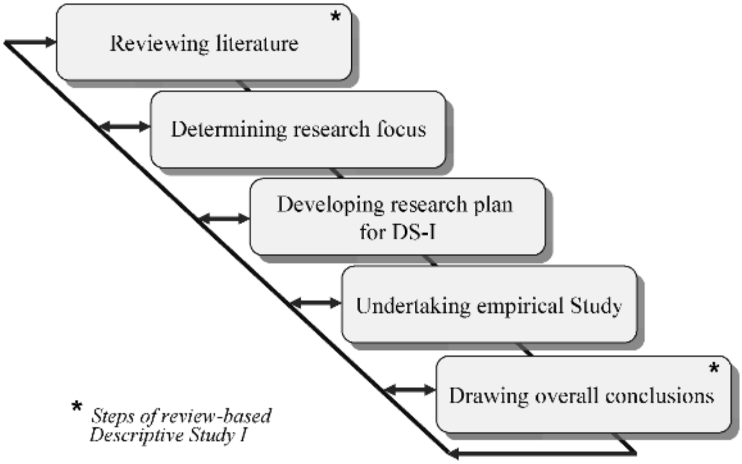

Figure 4.1 Main steps in a Comprehensive DS-I, stars (*) indicating the steps in a Review-based DS-I

This process will involve many iterations; with every study, the understanding increases and may give rise to further empirical studies or the literature reviews. In each cycle one or more methods can be used, which can differ from cycle to cycle depending on the specific research questions and hypotheses to be addressed. For example, the starting point might be a survey amongst a large number of companies to explore the main factors influencing the topic of interest, and then to interview key players to find more details about these factors. One might also choose to start with interviews to obtain a detailed understanding of the topic of interest in the contexts represented by the interviewees, and then undertake a survey to verify whether the findings are true for other contexts. In many instances, research questions rather than hypotheses will be the basis for this stage because the area of design is still relatively unexplored.

4.4 Reviewing Literature

Reviewing the literature is an activity that has to be continued throughout a project in order to keep up-to-date with the latest research findings. This section will focus on the review of the literature relevant for DS-I.

4.4.1 Identifying Literature

The aim of a literature review in DS-I is to extend the level of understanding gained thus far and update the expectations as represented in the Initial Reference and Impact Models, respectively. The resulting level of understanding should help decide whether the aims, identified problems and assumptions are realistic and relevant and hence, to help decide on the next steps in the research process. This involves a detailed study of the literature, with a particular focus on the results of empirical studies.

Studies from different disciplines and with different aims may potentially contain statements, models and theories relevant to the research problem at hand. As discussed in the previous chapter, the literature review should therefore consider other potentially relevant areas. The ARC diagram (Section 3.6) was set up for this purpose. In addition to the literature, exploratory discussions with experts and stakeholders in academia and practice can be very useful. In our reliability example, such discussions shed light on problems and experiences with reliability that were not published.

The Initial Reference and Impact Models indicate relevant factors and can therefore be used to guide the literature review. A possible way to proceed is to:

- check each link in the models against the literature to see the extent to which these have been shown to exist, or can be expected to exist using the evidence available;
- check the literature for additional influencing factors and links not considered earlier;
- verify the relevance and correctness of the preliminary Success and Measurable Success Criteria;
- continue until there is a *complete or at least sufficiently complete link* between the factors that are of interest and the Success Factors.

In our reliability example, the links and chosen Key Factor in the Initial Reference Model shown in Figure 3.10 gives rise to the question: What constitutes reliability? The Initial Impact Model in Figure 3.11 points to the question: How and how well is reliability assessed before details of the product are known, *i.e.*, in the early design stages? This also requires the investigation of currently available Design for Reliability methods, in particular their pre-requisites in terms of which product details have to be known. Investigating how and how well other product properties such as safety, performance, cost, manufacturability, environmental impact, *etc.*, are assessed in the early stages might also provide interesting information.

For a literature review we suggest to:

- first do a quick read of each publication (see Section 3.3);
- if a publication seems relevant or interesting, make a summary (see Section 4.4.2);
- use the DRM framework to place the study in the context of other design research.

4.4.2 Summarising Literature

It is important to write a summary of each publication while reading, not only noting down the statements but also adding page numbers and remarks about their relevance for the research topic at hand. Remarks can refer to the aim, the setup, the analysis and the findings of the study, as well as the conclusions that were drawn. Being critical is important, as long as the criticism is fair and constructive. Critical reviewing includes mentioning the positive elements of a publication. People's contributions to the field have to be acknowledged.

It is important to be careful to distinguish between statements and remarks. One's own opinion should be clearly separated from that of the authors, *e.g.*, by using separate paragraphs for the summary and for remarks. Statements in the summary that are directly taken from the publication should be immediately identifiable as citations, as later on this will no longer be clear. Careful documentation of what was read will benefit the writing up by preventing long searches for particular quotes and reference details (see also Chapter 7 on writing up).

The literature in design should be read carefully to determine whether the statements in a publication are *descriptive* or *prescriptive* in nature, that is, whether they describe how design *takes place* and how a particular support *works*, or how the author believes or suggests that design *should take place* and a particular support *should work*. We found that in many publications this is not made very clear. The first step, therefore, is to try to identify the source of each statement, *i.e.*, whether there is any description of, or reference to, empirical research on which the statement is based.

If such a description or reference can be found, the second step is to find out the strength of the evidence and its relevance for one's own research. On how many cases are the statements based? What research methods were used? What was the context of the study? Do the statements represent actual findings or are they derived from findings through reasoning, *i.e.*, are they interpretations? In the latter case, the assumptions behind the interpretations have to be checked; the statements could be based on speculation or involve unacceptable generalisations going well beyond what the setup and context of the study allow. Whether strong evidence exists or not, does not necessarily reflect upon the quality of the study – the study might have been exploratory.

Reviewing Empirical Studies

In a proper empirical study, the aim, the research questions and/or hypotheses, the type of data collected, the way it is collected, processed and analysed, the interpretations and conclusions should all match. This implies that in order to assess a particular statement for strength, quality and relevance, several details about the study have to be known. This requires a more thorough analysis of the publication, and may require contacting the author(s) if details are missing of a particularly relevant publication. The aim of this analysis is not to criticise existing work, but to develop a true understanding of the topic of interest so that one's own research project can be more effective and efficient.

The importance of such an analysis is illustrated by a publication we came across describing the results of an observational study of design. One of the results was a table showing the percentages of requirements and constraints that were taken from one of three identified sources. As the publication did not provide much detail about the set-up of the study, the author was contacted. The study turned out to be based on the analysis of, what the researcher called "interesting parts" of a video recording of *a single* designer, thinking aloud while he was working on a small design problem provided by the researcher. The observation as such cannot be criticised, those percentages were indeed found; it is the generalised way in which the conclusion was formulated that did not match the study and therefore cannot easily be used as the basis for other studies.

Table 4.1 shows a checklist we developed to support the review of empirical studies in design. Details of each dimension and an example can be found in Appendix A.2. The assumption is that empirical studies can be characterised by the options chosen by the researcher(s) for a set of dimensions shown in the first column of the checklist. The choice is guided by the aim of the research; by the specific research questions, hypotheses, models or theories that were defined or used; and by the specific context and constraints of the research project. The choice determines the potential findings and possible generalisations.

Many of the options are interrelated, *e.g.*, the decision to go into industry will limit the possible data-collection techniques, and a particular data-collection technique is likely to affect the number of cases that can be investigated. Not all dimensions and options are relevant to each research method. When multiple methods are used independently, it is useful to apply the checklist for each method. If methods are used together in one study, the specifics of each method can be separated for each dimension.

Table 4.1 Checklist for determining the characteristics of empirical studies, not all dimensions and options apply to all studies (adapted from Blessing (1994))

Dimensions	**Options**
Aim, research questions, hypotheses	The aim of the research project and of the study, main research questions and hypotheses, Success Criteria and/or Measurable Success Criteria and possible constraints
Nature of the study	Observational or interventional (*i.e.*, whether the study involved intervention in the design process by the researcher), comparative or non-comparative
Theoretical basis	Paradigms, methodologies, theories, views, assumptions, *etc.*, that guided the researcher
Unit(s) of analysis	The element(s) for which findings are reported and about which to draw conclusions that are intended to be generalised
Data-collection method	The method(s) used, such as direct observation, participant observation, document analysis, questionnaire, interview
Role of researcher	Type of involvement of the researcher in the research process

Table 4.1 *(continued)*

Dimensions	**Options**
Time constraint	Time constraint imposed by the researcher, *e.g.*, available design time, available time to answer a questionnaire, time of the observation (in case the phenomena observed lasts longer)
Continuation	Continuous data collection or sampling
Duration	Length of the process studied and length of the whole process (note that these can be different)
Observed process	Starting point and required deliverables of the observed process: *e.g.*, specification as starting point, layout drawing, prototype or product as deliverable
Setting	Location of the study, including whether the setting was contrived or natural
Task	Type and complexity of task. Nature of the observed tasks: real, realistic or artificial
Number of cases	Number of data sets collected, *e.g.*, the number of experiments, interviews, observed groups, products
Case size	Number of persons, product elements, employees, *etc*., within each case
Participants	Level and type of experience, background, size of organisation, *etc*.
Object	Description of the design object, company, project or documents involved
Coding and analysis method(s)	Methods used to process, code and analyse the data, *e.g.*, use of pre-determined coding schemes or not, and statistics applied
Verfication method(s)	Methods used to verify the results
Findings	Main statements, model, theory, conclusions resulting from the study
Notes	Anything remarkable or important in the publication, that is not covered by the other dimensions, missing information, relevance for one's own project, *etc*.

The chosen options for a particular study, together with the main findings, aid in:

- comparing studies, their setup and their findings;
- formulating justified comments, *e.g.*, regarding the amount of evidence;
- determining whether pieces of evidence from different studies can be brought together to form stronger evidence;
- finding possible explanations for contradicting evidence;

- establishing whether findings can be used as the basis for one's own research, *e.g.*, based on the amount of evidence and the context in which the study took place.

In addition, reviewing the literature using the checklist provides an overview of the various methods that have been applied and the ways in which studies have been set up and conducted (as has been done with an earlier version of the checklist in Blessing (1994) and Dwarakanath *et al.* (1995)). The overview can also inspire and help plan one's own empirical studies (see Section 4.6.3).

4.4.3 Updating Reference and Impact Models

The literature review will result in:

- a summary of, comments on and a comparison of relevant theories, models and other findings;
- a summary of, comments on, and comparison of commonly available support (details will be investigated in the PS stage);
- more specific research questions and/or hypotheses;
- above all, more detailed Reference and Impact Models, Success and Measurable Success Criteria, and Key Factors.

This can then be used to determine whether the available understanding is sufficient, or whether empirical studies are necessary (see Sections 4.8.5 and 4.8.6).

Reliability Example

In our reliability example the detailed literature review provides the following understanding:

- In general, early failure detection and analysis do reduce the number of iterations in a design process. A large number of iterations increases lead time. These and related statements from different sources are combined into a new set of influencing factors and links and represented as a partial Initial Reference Model, see Figure 4.2.
- The quality of a concept contributes to the quality of an embodiment. Similarly, the quality of the embodiment contributes to the quality of the detail design. Since reliability is a major component of quality, the researcher argues that hence reliability of a concept should contribute to the reliability of its embodiment, which in turn should contribute to the reliability of its detail design.
- Reliability of a product depends on the reliability of its detail design, the quality of production, the quality of the bought-in components and the quality of use. The latter is determined by the clarity of the instruction and the motivation behind the product's use, that is, whether the user likes to use the product and can freely decide to use it, or whether the user *has* to use the product, whether he likes or not, *e.g.*, in a work situation.
- Existing Design-for-Reliability methods require a level of product detail that is not available until the detail design stage.

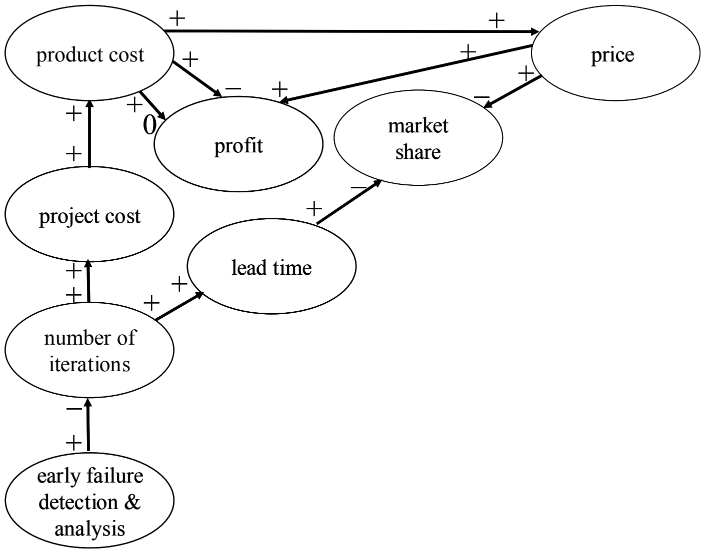

Figure 4.2 Partial Initial Reference Model related to the effects of failure detection and analysis in an early stage

- How designers assess reliability during embodiment design has not been investigated.
- Several general design principles related to a variety of properties exist for supporting embodiment design. The principles are derived from best practice and were found to have a positive effect on the quality of the product. None of the principles, however, focus directly on reliability and the effects of the application of individual principles on reliability or on product quality has not been investigated.
- The principles are based on basic design rules, also derived from best practice and are applicable throughout the embodiment design stage. These rules state that clarity, simplicity and unity, the so-called internal properties of a product, have to be maximised. They relate to components, interfaces and their configuration. They are easy to understand, but not very specific. Designers apply these rules, but are often not aware of this. They are said to have a positive influence on the so-called external product properties (of which reliability is one) and hence the quality of the product, but the effect of the application of the rules, individually as well as together, has not been investigated.
- Reliability involves the quality of the components, but more importantly the quality of the interfaces between the components.

Based on the information obtained from the literature, the researcher draws several partial Reference Models (as shown in Figures 4.2, 4.3 and 4.4) that are then used to update the earlier Initial Reference and Impact models.

From these models the researcher infers that: (1) best practice results in good designs and these are likely to be reliable; (2) overlap exists between what the basic design rules 'clarity', 'simplicity' and 'unity' address and what reliability depends on (quality of components, interfaces and configuration); (3) the basic design rules can be applied during early embodiment; and (4) early assessment of reliability should, like early failure detection, reduce the number of iterations.

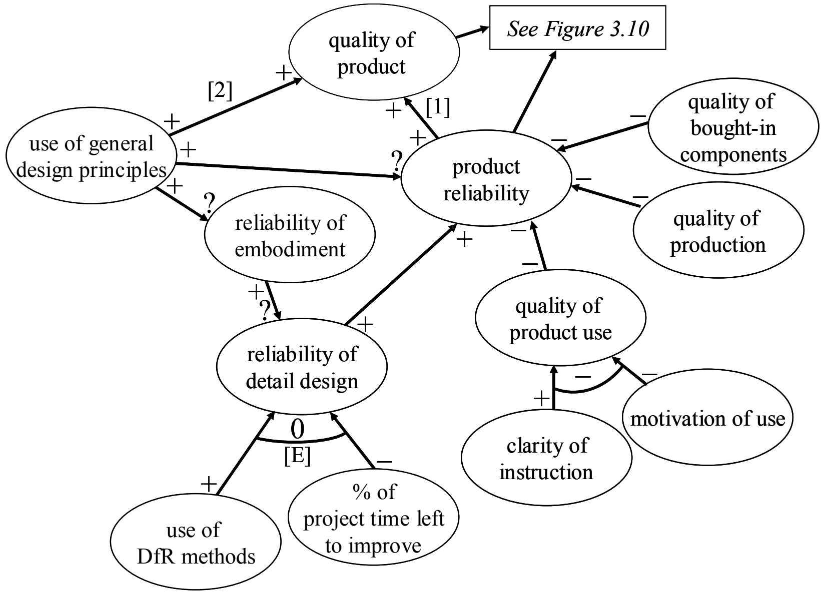

Figure 4.3 Partial Initial Reference Model based on some of the findings in the literature related to reliability

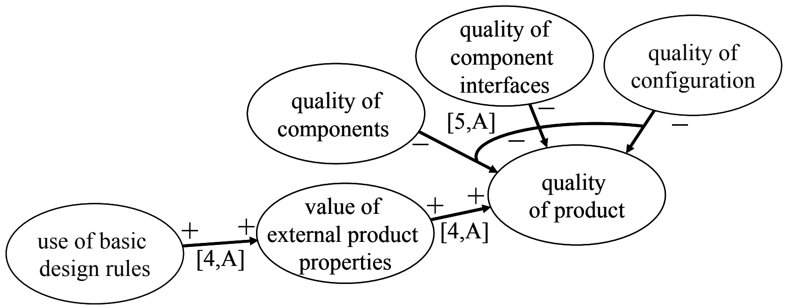

Figure 4.4 Partial Initial Reference Model based on some of the findings related to design rules

From this combination, he concludes that the basic design rules seem potentially useful for achieving the research aim, namely assessing and improving reliability in an early stage of the design process. He also concludes that the level of understanding of the application and effects of these rules is insufficient.

The following research questions remain:

- Does a relationship between reliability and clarity, simplicity and unity (together and separately) exist?
- What constitutes this relationship and is it causal?

The researcher decides to undertake a Comprehensive DS-I and, if a relationship exists, to focus on how to achieve reliability through using the basic design rules.

4.5 Determining Research Focus

In many cases, the updated Initial Reference and Impact Models will reveal that it is too early to start developing support: too many assumptions, rather than evidence, link the Key Factors to the Success Criteria. A detailed empirical study is required to gain knowledge about the missing and contradicting links in order to complete the Reference Model.

A pre-requisite for an effective and efficient empirical study is a good set of research questions and/or hypotheses. Data collection should focus on data that can be used – preferably directly – to answer the questions or verify the hypotheses: it is very easy to end up with large amounts of data that contribute little to the research aim. Furthermore, data collection should focus on data that can be collected within the constraints of the research project. If the latter is not the case, the research questions and hypotheses need to be adapted. The research questions and hypotheses resulting from the RC stage are a good starting point, but usually incomplete and not sufficiently well defined for a Comprehensive DS-I, since the literature review in DS-I has led to a better understanding of the current situation.

After a short section on identifying and defining factors and links of interest (Section 4.5.1), Section 4.5.2 discusses in more detail the formulation of research questions and hypotheses. In Section 4.5.3 methods are proposed for refining the questions and hypotheses, followed in Section 4.5.4 with suggestions for focusing the resulting set.

4.5.1 Identifying and Defining Factors and Links of Interest

In most projects it is not possible to investigate in detail all factors and links in the models that are inadequately understood, despite their expected influence. Some factors might fall outside the scope of the research project or the expertise of the researcher(s), others might be considered to have a relatively low impact. In our reliability example (see Figure 3.10) the factor ‘quality of product’ is influenced by ‘quality of production’ but this will not be pursued further because the focus in this case is on improving design and not production, and because the company involved in the project does not consider production quality to be an issue.

The most important reason to reduce the number of factors and links to be investigated is time. Detailed empirical studies are generally very time consuming and most projects are limited in time and resources. In our opinion, in a detailed empirical study it is generally better to have a deep understanding about a few factors, than a shallow understanding of a large number. We found that research students tend to grossly underestimate the effort required and to be far too optimistic about the number of factors and links they can address.

The investigation should focus on the weak links between the thus far chosen Key Factors and Measurable Success Factors, as these provide the core argumentation for developing support. If this still involves too many factors and links given the available time and resources, it may be necessary to change the Key Factors and Measurable Success Factors or reduce their number.

4.5.2 Formulating Research Questions and Hypotheses

The initial set of research questions and hypotheses from the RC stage and the Initial Reference Model will trigger new and more detailed questions and hypotheses. Taking our example, the question ‘Does a relationship between reliability and clarity, simplicity and unity (together and separately) exist’ will trigger questions such as: What is reliability? What is clarity, what is simplicity and what is unity? Are these terms familiar to designers? Do designers explicitly determine clarity, simplicity and unity of their designs? At which stages do designers determine reliability? How do they determine reliability? How are clarity, simplicity and unity related? Does increased clarity/simplicity/unity increase reliability? Does their combination increase reliability?

This section describes how to derive and formulate research questions and hypotheses.

Research Question

As defined in Section 3.4.2, a *research question* is a question for which no answer exists yet. The type of question determines the research approach and, in particular, the methods that can be used. The selection of the most suitable methods is discussed in Section 4.6.1. In our example, some of the research questions are: What causes a lack of product reliability? How can reliability of an embodiment be assessed?

Research questions can relate to any of the facets of design shown in Figure 1.1, such as:

- What is creativity? How important is creativity for a company’s success?
- What role do gestures play in communication between designers?
- How are requirements generated, evaluated, used and managed?
- How do physical characteristics of a product relate to the emotions it evokes?
- Why do designers typically generate very few product alternatives?
- What is the effect of available time on planning the design process?
- What kinds of CAD system are used in small and medium sized enterprises (SMEs)? How, when and for what are they used? Why and how were they chosen? What are their effects?

- How does a product evolve from idea to embodiment? How are products represented throughout this process and why are certain representations chosen?
- What and who determine product quality?
- How does the organisational structure influence teamwork? How does distributed design influence the design process and the product?
- How are customer requests and complaints dealt with in the consumer goods industry?
- How do new products influence social behaviour and culture?
- How do macro-economic factors influence innovative behaviour in practice?
- What is sustainable development and what is the role of design?

Research questions can be (Trochim 2006):

- *descriptive*: when the aim is "to describe what is going on or what exists";
- *relational*: when the aim is "to look at the relationships between two or more variables";
- *causal*: when the aim is "to determine whether one or more variables causes or affects one or more outcome variables".

Research questions can be very general, in particular in the early stages, but in order to find answers through an empirical study, the research questions have to be sufficiently detailed so that these are (adapted from the description of hypotheses given by Frankfort-Nachmias and Nachmias (1996):

- *clear*: all of the variables should be conceptually and operationally defined, *i.e.*, they should be defined such that they can be observed and assessed. (experience, *e.g.*, requires an operational definition, such as 'having worked in industry for more than 8 years and having designed the same products before' as discussed later in this section);
- *unspecific*: in order to avoid bias the expected direction of the answers and the conditions under which the answer holds should not be given, *i.e.*, they should be answer-free (see 'bias' in Section 4.7.1);
- *answerable:* it should be possible to find an answer with available methods;
- *value-free*: which is particularly important in a social context.

A few words about the term *variable*, and its relationship to the term 'influencing factor' are useful at this stage. Variables are characteristics of a situation or phenomena that can change in quantity or quality. They can take on at least two values. Note that value does not necessarily refer to a numerical value. (See also the discussion on scales of measurement in Section 4.7.2.) The influencing factors in our Reference and Impact Models are variables.

A distinction is made between dependent variable and independent variables. The *dependent variable* (or criterion variable) is the variable the researcher wishes to explain. The variable that is expected to influence the dependent variable is the *independent variable* or (explanatory or predictor variable). Independent variables are actively changed and their effect on the dependent variable measured. The variables used need to be mutually exclusive, *i.e.*, their definitions should not be overlapping and a particular variable should not be a subset of other variables. The relationship between dependent and independent variables can be spurious, that is, a relation is found but this is actually caused by another variable affecting both the dependent and independent variable. To avoid this, *control variables* are introduced. Control variables are those variables that could have an effect and are involved in alternative explanations of the observed relationship.

In our reliability example, we may wish to study the factors influencing ‘reliability of detail design’ (see Figure 4.3). In that case ‘reliability of detail design’ would be the dependent variable, the independent variables would be ‘reliability of embodiment’, ‘use of DfR methods’ and ‘% of project time left to improve’. Possible control variables to consider are ‘experience of designers’ and ‘type of product’ to check whether an observed correlation between reliability and use of DfR methods is indeed caused by the methods and not by the experience of the designers using these methods or by the type of product (the methods may not apply well to certain types of products). If control variables are found to have an influence these should be added to the Reference Model.

Although research questions do not include an answer, careful analysis often reveals underlying assumptions, some of which are expressed in the aim or criteria of the research project. No one is free of assumptions, even if it is only the assumption that the topic is worth investigating for a particular reason. Identifying these assumptions may lead to further questions or explicit assumptions (hypotheses). The following example may illustrate this. The research question is formulated as ‘How often do designers iterate in order to make corrections to earlier solutions?’. If the focus of the study is on corrections, this question will look at how corrections influence the number of iterations. However, if the focus of the study is on iterations, the question assumes that making corrections is the main, if not the only reason for iterations. This may not be the case. Questioning this assumption will lead to another research question, namely: Why do designers iterate? Or to several questions in the form of ‘how often do designers iterate in order to do X?’. This obviously requires knowledge of the possible reasons (Xs). Most importantly, different research methods may be required to answer these questions.

Hypotheses

In Section 3.4.2 we defined *hypothesis* as a tentative answer to a research question in the form of a relationship between two or more variables, or in our case between two or more influencing factors, including the Success Factors. That is, an hypothesis is a claim or a statement about a characteristic of a situation, or a proposed explanation for a phenomenon. Hypotheses are tested as to whether they can be accepted or have to be rejected given the available evidence. Hypotheses should be formulated such that they are ’refutable’, that is, that they can be disproved or demonstrated to be false or erroneous. For example, an hypothesis which contains the word ‘might’ cannot be refuted, as it will always be true: the hypothesis that ‘product reliability *might* influence product quality’ will hold when the influence is observed and will hold if the influence is not observed (‘might’ implies ‘might or might not’).

Examples of hypotheses, chosen to be in line with some of the research questions listed earlier, are the following:

- A high level of creativity within the design team increases a company's success.
- Communication between designers who are not able to observe each others gestures leads to an increased level of misinterpretation.
- Requirements and solutions co-evolve during the design process.
- The requirements list generated at the beginning of a design project is not managed consistently through the project.
- The colour of a product has a strong influence on its perceived attractiveness.
- The use of discussion forums on particular products has increased the influence of customers on product development.
- Distributed design increases the number of iterations.
- Financial incentives of local governments increase the number of small companies involved in innovation.

As these examples show, the formulation of an hypothesis is far more specific than that of a research question. One research question would require several different hypotheses.

All hypotheses express a co-occurrence and correlation between variables (a descriptive relationship), but not necessarily a causal relationship, that is, that one variable is responsible for the other(s). The aim is to verify or falsify these hypotheses. All but the third and fourth of the example hypotheses above suggest a causal relationship. Note that while the first three hypotheses denote expected links, the fourth denotes the expected value of a factor ('consistency of managing' is low). In design research, we are ultimately interested in causal relationships; by knowing causes and effects we can address the causes by developing support in PS. However, causal relationships are much more difficult to verify than relational hypotheses (see Section 4.7.3 for a discussion about causality).

Similar to research questions, hypotheses can themselves be based on assumptions. These underlying assumptions can be about the domain in which the hypothesis is expected to be valid, the distribution of the population (which is relevant for statistical tests), the type of products to which the hypothesis refers, *etc*. Hypotheses that can be accepted in one situation might have to be rejected in another situation. Making the underlying assumptions explicit is thus relevant for setting up empirical studies, as these will point to factors that have to be considered as these might provide alternative explanations.

When hypotheses can be formulated, an empirical study can be focused more easily because it is clear: (1) what needs to be known, namely, whether the relationship expressed in the hypothesis can be accepted or has to be rejected given the available evidence; (2) what has to be measured,[12] namely the variables in the

---

[12] As emphasised earlier, 'measuring' is used in the meaning of assessing the value of a factor, whether absolute or relative, whether in qualitative or quantitative terms. Classifying would thus be a way of measuring.

hypotheses and the type of relationship between these; and (3) how (at least partially) the setup would have to be.

Hypotheses should be (Frankfort-Nachmias and Nachmias 1996):

- *clear*: all of the factors and links should be conceptually and operationally defined, that is, they should be defined such that they can be observed and assessed;
- *specific*: the expected direction of the relationships between the variables (in the case of causal relationships) and the conditions under which the relationship holds should be given;
- *testable*: it should be possible to find an answer with available methods;
- *value-free*: which is particularly important in a social context.

Hypotheses can be derived in a variety of ways: deductively from theories (in our case using the literature and the Reference and Impact Models), inductively on the basis of direct observations, intuitively, or by using a combination of these approaches (Frankfort-Nachmias and Nachmias 1996). In an approach based on hypothesis formulation rather than on research questions, intuition is required where neither theories nor observations provide explicit hypotheses. In design, very few established theories exist, and only in the last decade have results from direct observations in design become available. However, potentially relevant theories may have been developed by other disciplines, such as theories on problem solving, decision making, or technical systems. As discussed in Section 3.6, these should be used wherever applicable. If no seemingly relevant theories are available, we would not recommend the formulation of hypotheses based on intuition because of our, as yet, limited understanding of design – in particular when the researcher has little experience in design. In this case it is better to base the study on research questions instead and use a more *data-driven* approach.

Coverage

While formulating and refining research questions and hypotheses, the researcher should take into account:

- the research goal (to remain focused);
- the level of understanding that could be obtained from the literature (to remain efficient);
- possible effects on the findings by other factors (to remain open-minded);
- project constraints that are beyond the researcher's control (to remain realistic).

The latter two will be discussed in this section.

The set of research questions and hypotheses needs to be expanded to include design-related factors that might influence the findings but are outside the immediate focus of interest, and research-related factors caused by the setup of the study. An understanding of these factors and their influences allows the researcher to determine possible alternative explanations for the findings. Such factors therefore have to be included in the research plan, which – in our experience – is often forgotten. It is important to imagine what could influence the phenomenon studied and to ensure that related questions and hypotheses are added to the research plan such as to take these influences into account. This requires going through the whole study, from collecting the data, imagining the context, trying to take the position of those involved, processing the data, analysing the data, *etc*. We found it useful, where possible, to use available data sets, *e.g*., video recordings or interview notes, to try to get a grasp on the type of data that can be collected with a particular method. In many instances, researchers are too optimistic about the type of data and the precision of the data that can be collected.

To identify the design-related factors, we again refer to the use of the facets of design (Figure 1.1) as a checklist. Examples are the type of product, the experience of the participants, the type of practice, or the country. To identify the research-related factors, the literature on the chosen methods has to be consulted. Examples are the effects of being observed or interviewed, of the interest and motivation of the participants, of their role within their organisation, the environment in which the study takes place, the material provided or available to the participants, and the role of the researcher. In Appendix A, some of the effects of methods are described. Additional research methods might be needed to be able to study these factors. For example, if the main data-collection method is observation, a questionnaire might be required to obtain data about experience, motivation, *etc*. Analysis of documents and/or interviews might be required to understand the historical development of a product.

Although the goal is to make the set of research questions and hypotheses as complete as possible as early as possible, new questions and hypotheses will arise as research progresses that may be useful or even necessary to address. Increased understanding will give rise to alternative, more in-depth or precise questions and hypotheses.

Conceptual and Operational Definitions

Concepts constitute the professional language of the researcher. A concept "is an abstraction – a symbol – a representation of an object or one of its properties, or of behavioural phenomenon". "Concepts do not actually exist as empirical phenomena – they are symbols of phenomena, not the phenomena themselves." They have four functions (Frankfort-Nachmias and Nachmias 1996):

- they provide a common language, which enables scientists to communicate;
- they give scientists a perspective – a way of looking at phenomena;
- they allow scientists to classify their experiences and to generalise from them;
- they are components of theories – they define a theory's content and attributes.

Examples of concepts in design research are: requirement, function, product quality, experience, evaluation, team working.

If concepts are to be used to communicate in science, they need clear definitions. Concepts can be defined by using other concepts, that is, given a *conceptual* definition. For example: 'evaluation is the activity of a designer in which he or she assesses the object on which he or she is working'. Primitive terms are concepts that cannot be defined in other concepts, but most importantly, their meaning is generally agreed upon. Other terms are defined using primitive terms.

Conceptual definitions (Frankfort-Nachmias and Nachmias 1996):

- must point out the unique attributes or qualities of whatever it defines. It must include all cases it covers and exclude all cases it does not;
- should not be circular;
- should be stated positively, that is, point to attributes that are unique only to the concept they define;
- should use clear, generally agreed terms.

Certain concepts cannot be directly measured or observed, or for which no generally agreed measurements exist, *e.g.*, experience. This type of concepts is called *construct*. Constructs can only be measured by measuring certain other characteristics of behaviour and background. For experience, one could measure the number of years since the last education, the number of years involved in the particular task, the number of projects carried out, or a combination of these.

Concepts have to be given an *operational* definition in order to empirically establish the existence of a phenomenon described by these concepts. Operational definitions define 'what to do' and 'what to measure'. For example, weight can be conceptually defined as 'a measurement of gravitational force acting on an object'. A possible operational definition is 'a result of measurement of an object on a Newton Spring scale'. The operational definition of a construct might be difficult to formulate and require a combination of measurements that enable its indirect measurement. In the area of design, many conceptual definitions exist for product quality. For the purpose of comparing design processes, an operational definition of product quality was defined in Blessing (1994) as "the degree to which the product fulfils (on a scale of 1–4) the set of requirements given in the task description and a set of general requirements defined by the researcher". This was followed by a description of what to do, involving averaging the quality measure determined by two individual experienced designers, not involved in the experiment. The degree of fulfilment can be said to be a proxy, or a predictor of product quality, as the product does not exist yet and the assessment is based on an embodiment drawing of the product.

As stated earlier, variables are characteristics of a situation or phenomena that can change in quantity or quality and that are measured in order to answer research questions or verify hypotheses. The degree of fulfilment of the set of requirements mentioned above is an operational definition of the concept of product quality, and is the variable used in the study to represent product quality. Note that if it were possible to measure product quality directly, this in itself would have been the variable. Table 4.4 shows an example of the use of concepts (in this case constructs) and variables.

4.5.3 Techniques for Refining Research Questions and Hypotheses

We propose three complimentary techniques that we developed to help refine the initial set of research questions and hypotheses, so that an effective and efficient empirical study can be undertaken:

- Question and Hypothesis Analysis;
- Answer Analysis;
- Question-Method Matrix Analysis.

For using these techniques, the recommendation given earlier still applies: focus should be kept, yet questions and hypothesis that could help identify potential alternative explanations should be included.

Question and Hypothesis Analysis

This first technique involves a direct analysis of each research question and hypothesis as they are formulated, by asking:

- What needs to be measured to be able to answer the question or verify the hypothesis? Do all terms have operational definitions?
- Who or what can provide the data needed to answer the question or verify the hypothesis, and would this data count as strong evidence?[13]
- Is a particular type of research method or are particular options for the dimensions listed in Table 4.1 required, and is this possible with the available resources?

Answering these questions will lead to further questions and hypotheses, to reformulation of the questions and hypotheses and to a clearer focus of the research. Furthermore, the answers will provide an indication of the required research methods. As an example, let us take the following research question (Q):

Q *How do designers set up a requirements list?*

This question is far too open. There is no indication about what to measure, that is, what data to collect and in which context. Using the various facets of design shown in Figure 1.1 as a checklist of what may be involved in setting up requirements lists, can lead to the following questions and further considerations:

Q *Who is involved in setting up a requirement list, directly and indirectly? (Only involving designers might be too limited.)*

- All terms need to be defined.
- Academic terms might not be established in the domain of study. Is, for example, the term requirements list used in the context in which the study takes place?
- The possibility to identify these persons and their involvement needs to be considered. Can a single researcher observe this, or are other methods needed?

Q *In which ways does the type of product influence the process of setting up a requirement list?*

[13] For example, if the participants involved in a study are not selected carefully, the data obtained may not provide strong evidence. Such as when novice designers are asked about how design practitioners deal with certain problems, the answers may not be as representative as when experienced designers are asked.

- To answer this question, several processes need to be investigated. Are there sufficient resources?
- A definition of product types is required, and certain types need to be selected for investigation.
- Thinking or reading is required about possible influences, as what is not considered will not be measured.

Q *What activities are undertaken to set up a requirements list?*

- What counts as a requirements list?
- What are activities? Whose activities are investigated?
- Can the activities be observed: at all; within the timeframe of the project; with the number of researchers available (activities can take place in parallel); or should those involved be asked?

Q *How often do designers set up a requirements list?*

- What is meant by setting up?
- Can this be assessed within the timeframe of the research project, or should this question be answered through an interview?

Q *What is a good requirements list?*

- Can this be recognised by the researcher? By those involved?

During this process of Question and Hypothesis Analysis, possible data-collection methods begin to emerge, such as experiments, observation or document analysis. Some of the concepts in the questions can be measured directly by particular methods, but many will need further refinement and definition, such as the terms ‘project stage’ and ‘good requirements list’. Taking into account possible assumptions behind the research questions may lead to the emergence of hypotheses.

The process of refining *hypotheses* using Question and Hypothesis Analysis may result in additional hypotheses (including contradicting hypotheses) and in research questions that have to be answered in order to be able to verify the hypotheses. Take, for example, the following hypothesis:

H *An extensive requirements list developed early in the design process reduces the number of changes in the later stages of the design process.*

(Note that this hypothesis is a relative statement (‘reduces’) and therefore requires a comparative study.)

The above hypothesis gives rise to additional questions:

Q *Why does such a requirements list reduce the number of changes? (Improving understanding)*

Q *Is it possible to distinguish stages in a design process? (Questioning the research methods)*

Q *What are the characteristics of an extensive requirements list? (Clarifying terminology)*

It may be relevant to include opposing hypotheses as the rejection of an hypothesis is not necessarily the same as accepting its opposite, as we pointed out while discussing the Reference Model in Section 2.4. If it was found, for example, that an extensive requirements list early on in the process reduces the number of changes later on, this does not imply that a less extensive requirements list automatically increases the number of changes. Other factors such as the maintenance of the requirements list may play an important role.

It is important to question generally accepted statements when the validity of these is at the core of the research argument and accept that this may lead to new research questions or the introduction of new hypotheses.

Answer Analysis

A second useful technique for refining research questions and hypotheses involves the analysis of answers. This technique works backwards from the documentation of the research to the data that needs to be collected. The intention is not to bias the result by preparing the answers, but to anticipate the types of answer and to think of possible representations of the data that help answering the questions or verify the hypotheses. This will lead to a refinement and extension of the set of questions and hypotheses and provide indications for the most suitable setup of the study. Answers can be:

- descriptive (*e.g.*, there is motivation behind this activity), interpretational (*e.g.*, ‘the motivation is private’) or explanatory (*e.g.*, the guiding principle, pattern, theme and/or causal links) (Miles and Huberman 1984);
- comparative (showing differences), relative (showing different ranks) or absolute (having a particular value);
- related to time, frequency or content;
- qualitative or quantitative.

Different types of answers require different representations. A simple example using the hypothesis mentioned earlier – an extensive requirements list developed early in the design process reduces the number of changes in the later stages of the design process – may illustrate this point. To verify this hypothesis the researcher imagines the graph sketched in Figure 4.5a left, showing the number of changes against time for each designer, thereby ranking the designers by the quality of their requirements list. This leads to questions such as:

- How many changes does a particular designer make?
  - This requires a different focus from ‘How often is the product changed?’ as in a team the changes might be made by various team members.
- At what point in time are the changes made?
  - This requires continuous observation to be able to measure time or time intervals.

- What is the quality of the individual requirement lists?
  - The way in which quality is defined – ranking (1st, 2nd, *etc.*), classification (high, medium, poor), percentage of overall number of requirements captured – affects the representation, the formulation and type of research questions, and the choice of research methods.

Had the researcher imagined a bar chart with the number of changes in each stage for each designer, again linked to the quality of the requirements list (see Figure 4.5b), this would have resulted in different research questions and hypotheses as well as data-collection and -analysis methods. It would, *e.g.*, not have been necessary to measure time to identify when designers make changes. Counting occurrences during a particular stage of the design process would have been sufficient. This representation obviously requires the identification of design stages, which in turn is not required for the first representation.

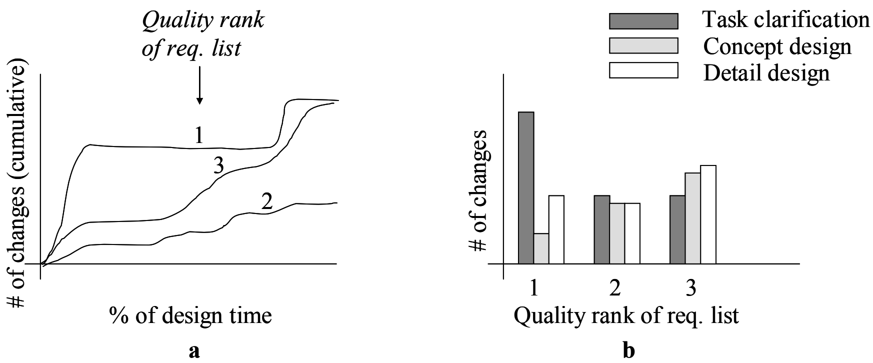

Figure 4.5 Different possible ways of presenting the results, affecting the research questions to be asked and research methods to be chosen

The requirements and constraints for the selection of the most suitable data collection and analysis methods resulting from the Answer Analysis are one of the strengths of this technique. In the above example, capturing time or not will result in a different data-collection method. If time is not captured, this cannot be added during the analysis stage. We have observed several instances, where the collected data was unsuitable for answering the research questions or verify the hypotheses, because the data could not be processed or analysed as required. We observed that the intention to reduce the effort required of the participant was one of the main reasons why – unintentionally – the wrong type of data or insufficient data was collected. Reducing the effort required is important, but only as long as this does not affect the data to such an extent that conclusions cannot be drawn any longer. This might render *all* effort in vain. In one study, *e.g.*, a questionnaire was used to ask managers for the turnover of their company and the number of designers. To reduce the effort required for answering, possible answers were grouped into five categories (turnover < 50 000, turnover between 50 000 and 100 000, *etc.*, and similar for the number of designers). The respondents only had to tick the appropriate category, rather than give actual numbers, which they might have had to look up. When analysing the data, however, it was found relevant to determine the turnover per designer, as one of the basic assumptions was that a large effort on designing would increase the added value and thus turnover. This calculation was not possible with the data collected. Absolute numbers, rather than ranges, should have been collected. A rank correlation was possible, but this turned out to be inconclusive.

Question-Method Matrix Analysis

The third technique for refining questions and hypotheses, the Question-Method Matrix Analysis, makes use of the relationship between research questions and hypotheses on the one hand, and data-collection methods on the other. It is intended to support the selection of research methods but at the same time it also refines the research questions and hypotheses. Because of its emphasis on method selection, this technique will be discussed in Section 4.6.2.

4.5.4 Focusing the Set of Research Questions and Hypotheses

It is relatively easy to come up with a large number of questions and hypotheses that all seem interesting and potentially useful: usually too many to deal with given the project's resources. Focusing the research is essential: as stated earlier, clear statements about a few facts have to be preferred above fuzzy statements about many facts. Focusing, however, does not mean to exclude alternative explanations.

To prioritise the questions and hypotheses and to focus the study, the following questions are useful:

- What is the reason for including this question or hypothesis? How important or essential is this question or hypothesis?
- Do the questions and hypotheses relate to one another? Would the answers provide a coherent picture?
- What use will be made or can be made of the answer? This refers not only to practical issues, but also to ethical issues surrounding studies that involve human beings.

The above-mentioned Question-Method Matrix Analysis aids in answering these questions.

The basis rules for product development – clarity, simplicity and unity – can also be applied to the final set of research questions and hypotheses. The formulations should be clear and simple such as to easily find sound answers. The set as a whole should form a unity.

The questions and hypotheses that have to be left out can still represent interesting avenues for investigation. To avoid losing these, they should be written down as directions for future research, which is usually the final section in a publication.

4.6 Developing Research Plan for DS-I

The time needed to plan an empirical study should not be underestimated. It is important to pay ample attention to detail in the design and preparation of an empirical study and to do a pilot study to try out the chosen methods. Empirical studies can be very time consuming, also for the participants, and redoing a study with the same participants – in particular in a design environment – is usually not possible because of time constraints and the learning effect. The aim should be 'right first time'. This requires careful planning.

The research plan of an empirical study should describe in detail:

- research goal and objectives for this study;
- research questions and hypotheses to be addressed;
- data-collection method(s) and setup;
- data-processing method(s);
- data analysis and interpretation method(s);
- method(s) to verify the results.

Strong relationships exist between these research activities. A useful starting point is the selection of suitable data-collection methods based on the set of research questions and hypotheses and on project constraints. Once the data-collection methods are determined, the other elements in the research plan can be detailed. Setting up the research plan is an iterative process and the plan may be subdivided into multiple studies, each covering a particular set of questions and hypotheses.

The freedom of selection is restricted by the inherent limitations of each method and by the various constraints that are outside the researcher's control, such as available time and resources, and restrictions in recording imposed by the context in which the study takes place.

We have found that many young researchers go for the method that is most commonly used or seems to require least time, and mainly worry about how to apply the method, rather than considering the suitability of the method for their research questions, hypotheses and constraints. This is like deciding to use a drill without knowing yet what to make: perfect holes might be the result, but that may not necessarily what is wanted. It is important to investigate the suitability of a variety of methods.

4.6.1 Selection of Methods

The literature review on studies into design provides a useful overview of methods that have already been applied. Consulting the literature on the research methods or the authors is nevertheless essential, as usually the details of the methods are not published. In Table 4.2, Table 4.3 and Appendix A.4 we have provided a short description of the most common data-collection methods, some suggestions based on our experience, and references to the literature. Because of the variety of factors involved in design, the study of design often requires the selection and combination of research methods from various disciplines. It may be important, *e.g.*, to look into the methods used by sociologists investigating group dynamics, by psychologists investigating decision making and the workplace, by computer scientists and ergonomists investigating human-computer interfaces, by management scientists investigating change management, by engineering scientists investigating optimisation and manufacturing, and by physicists investigating physical principles. The ARC diagram was introduced to identify the relevant areas and can therefore guide the search for potentially suitable methods from other areas.

It is important to consider using multiple methods, because "studies of multi-dimensional problems such as design activity require multi-level, multi-method approaches" (Bessant and McMahon 1979). Usually the formulated research questions and hypotheses already indicate that different methods are needed to answer or verify them. Furthermore, different methods may be required for each cycle of the research process to accommodate our increased understanding. Moreover, findings can be verified by locating similar data using different methods. This is called triangulation (Denzin 1978).

Excellent books have been written about the various methods and need to be consulted. When the research methods of a different discipline seem relevant, it is wise to also consult someone from this discipline. Ideally, this person should be involved in the relevant parts of the research project, as co-researcher, advisor or co-supervisor, because of their training and experience in applying the methods of their discipline.

Each method should in principle be used as intended and for the purpose for which it has been developed, although the use of methods in a different domain and in a different – but well-argued – way may give interesting results and shed a new light on a particular phenomenon. An example is the work of Suchman, who used Conversation Analysis (used in sociology for the analysis of conversational interaction between human beings) for the analysis of human–machine interaction, in this case users of copiers, in order to inform design (Suchman 1987).

No matter how a method is used, there will be results, but only if methods have been carefully selected and correctly applied, and the study carefully designed, is it possible to realise the rigour needed to obtain valid and useful statements. As Patton (1987) states: "the validity and reliability of qualitative data depend to a great extent on the methodological skill, sensitivity, and training of the researcher." This is equally true for the validity and reliability of quantitative data. We have seen many instances of poorly designed and executed investigations, violating basic rules of the application of a particular method, resulting in invalid, useless data and a waste of time and effort, which could have been prevented by consulting the relevant literature. "Systematic and rigorous observation involves far more than just being present and looking around. Skilful interviewing involves much more than just asking questions. Content analysis requires considerably more than just reading to see what's there." (Patton 1987). More on validity and reliability can be found in Section 4.7.4.

4.6.2 Selection of Data-collection Methods

We first focus on the selection of data-collection methods, as these to a large extent determine the selection of the other research methods.

Classes of Data-collection Methods

Data-collection methods can be divided in various ways. We have opted for a division into real-time methods and retrospective methods, indicating whether the methods are applied during or after the events of interest take place. Our focus is on methods that have been used in design research, but many more methods do exist.

Real-time methods (Table 4.2) can produce unadulterated, direct and potentially very rich descriptions of events and their context, because data is captured while the phenomena occur. This is enhanced by the availability of easy-to-use, high-quality instrumentation, *e.g.*, to record and process video data and measure a variety of variables, voice-recognition software as well as powerful databases and analysis software for quantitative and qualitative data. This has made it easier to collect detailed data in large amounts that can be analysed repeatedly and shared amongst researchers. Generally, the use of real-time methods reduces the number of cases that can be studied. Also, the availability of people and organisations may be limited due to the effort involved.

Table 4.2 Real-time data-collection methods used in design research

observation (no involvement of the researcher); • taking field notes; • recording activities against time; • counting occurrences and contents of particular events; • measuring values and occurrences;
participant observation (the researcher as participant); • several of the other techniques have been used to collect the data;
simultaneous verbalisation (audio or video taped); • thinking aloud; • introspection (commenting on one's own mental activity); • interviewing during the actual process; • talking aloud/recording team discussions;
diary keeping (designer as observer, or observing participant); • keeping a diary of the type instructed by a researcher; • keeping a diary as designer/researcher;
recording the evolution of documents through snapshots; • photographing sketches, drawings at regular intervals; • videoing the evolution process of a document; • keeping computer logs;
computer simulation; • spatial visualisation tasks • computer games to obtain information about specific behaviour.

Retrospective methods (Table 4.3) usually summarise events and rely upon memory or documentation, which may be very selective. There is also a danger that subjects will impose hypothetical constructs on the observed situation, so-called post-rationalisation, which may not give an accurate portrayal of the process. An advantage of these methods, however, is their suitability for involving large numbers of cases. Furthermore, they can be used where reflection is required.

Table 4.3 Retrospective data-collection methods used in design research

documents (case history compilation, archival analysis); • collecting formal project and product documentation; • collecting notebooks (informal documentation);
product data (product family data); • functional data; • service and maintenance data;
questionnaires; • open-ended questions; • multiple choice;
interviewing; • structured, semi-structured, unstructured; • focus groups; • reports by subjects.

In design research, we have found a prevalence of real-time methods in laboratory settings involving few cases. Retrospective methods were more common in industrial settings involving many cases, or to supplement real-time methods. A small, but increasing number of researchers uses real-time methods, often in the form of single-case studies, in practical settings.

Any of the methods can be used to study a variety of factors, but not all factors can be studied using a particular method. The same factor addressed by a different method will provide different data. For example, an interview about the interaction between members in a project team will reveal personal opinions about the interaction. Observation of the interaction would reveal very little about these opinions, except for interpreting gestures, postures and remarks. Observation, however, would allow statements about the frequency of interaction, the people (number and function) involved, the frequency of interaction over time, *etc*. These details cannot be obtained using interviews. In our reliability example, archival analysis of available maintenance and service data will allow statements about reliability of products, but will reveal little about how this data is used to improve the products.

A particular problem in design research is the availability of the intended participants, whether organisations or individuals. First and foremost because of the size of the population: suitable participants may be few in number. Other reasons we know from industrial settings are the expected interruption of ongoing work, worries about confidentiality on a personal as well as company level (there is no anonymity even though data can be treated anonymously) and the fact that interesting projects may be commercially sensitive. Increasingly we also hear companies complain about being inundated with requests from researchers. As a consequence they do not wish to participate at all, or select only those topics in which they are really interested. In a contrived setting the reasons for a lack of available participants can also be the time involved (what is the benefit?) and shyness about being the object of a study.

Selecting Methods Using Question-Method Matrix Analysis

The first criterion for selecting a method is whether it is in principle suitable to address a particular research question or hypothesis. Furthermore, the effort required from the researcher and from the participants differs greatly for each method and is likely to be an important factor in selecting suitable methods. In this section we present a technique called Question-Method Matrix Analysis to select the most effective and efficient combination of methods to address the formulated research questions and hypotheses, while at the same time helping refine these. Once selected, the specifics of the methods and the setup of the study have to be determined such that the questions and hypotheses can be addressed in the most effective and efficient way.

Figure 4.6 shows a Question-Method Matrix. The row headings contain the formulated research questions and hypotheses. The column headings contain the research methods considered. Each cell is divided into an upper part, indicating the suitability of the method for addressing the research question or hypothesis, and a lower part indicating the effort required from the researcher and the participants(s). Two ticks (√√) indicate that the method is expected to fully answer the question or verify the hypothesis. When the answer can only be obtained partially or indirectly, a single tick is given (√). The effort for the researcher and the participants to address the research question or hypothesis is indicated with R (small effort) or RR (large effort), and P or PP, respectively. In some cases the effort may be negligible, in which case this part of the cell remains empty. The effort includes everything from preparation, application and processing to analysis. Some methods, such as observation, will require a large effort from the researcher and virtually nil from the participant, if the participant is observed in his or her own context. On the other hand, keeping a diary will put the onus on the participant for collecting the data, although its analysis will still require considerable effort from the researcher depending on how standardised the diary sheets are. A differentiation in effort could be made for each method, rather than for each research question and hypothesis, but this does not provide much support in selecting methods. It is the combination and type of questions and hypotheses that eventually determine the details of the method and thus the effort required. For instance, a questionnaire about the designer's educational background will take up far less time than a questionnaire about the main lessons learnt in a particular project.

It is important to add specific features of the methods or the setup of the study needed for addressing a particular research question or hypothesis into the matrix underneath the method, *e.g.*, using the options for empirical studies shown in Table 4.1. The reason is that the same type of method, *e.g.*, interviewing, may require two different studies. In the example, a difference was made between interviewing users and interviewing designers, each of which requires different features and a different setup, and hence results in a separate column in the matrix to indicate that each requires a separate study.

<table>
<tr><th></th><th colspan="2">Interview users</th><th colspan="2">Interview designers</th><th colspan="2">Reliability assessment excercise</th><th colspan="2">Analyse reliability documentation</th></tr>
<tr><td></td><td colspan="2">Daily users, not buyers</td><td colspan="2">Different companies
Only experienced</td><td colspan="2">Ask experts
Use definitions
Existing designs</td><td colspan="2">Historical development and reliability</td></tr>
<tr><td rowspan="2">Research Question 1</td><td colspan="2">√√</td><td colspan="2">√</td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>R</td><td>P</td><td>R</td><td>P</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Research Question 2</td><td colspan="2">√</td><td colspan="2"></td><td colspan="2">√√</td><td colspan="2">√</td></tr>
<tr><td>R</td><td>P</td><td></td><td></td><td>R</td><td>PP</td><td>RR</td><td></td></tr>
<tr><td rowspan="2">Hypothesis 1</td><td colspan="2">√√</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>R</td><td>PP</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Hypothesis 2</td><td colspan="2"></td><td colspan="2">√</td><td colspan="2"></td><td colspan="2">√√</td></tr>
<tr><td></td><td></td><td>R</td><td>P</td><td></td><td></td><td>RR</td><td>P</td></tr>
</table>

Figure 4.6 Question-Method Matrix

To fill the matrix, we suggest the following procedure:

1. Draw a matrix.
2. Enter the first research question or hypothesis as a row heading.
3. Determine which methods are suitable and enter these as column headings.
4. Enter comments about any specific features of the method or setup relevant to address the research question or hypothesis. Many specific features will not become clear until the same method is chosen to address other research questions and hypotheses, which require other specific features.
5. Enter the expected level of suitability and the expected effort for researcher and participant(s) in the cells.
6. Repeat steps 2 to 6 for each research question and hypothesis, considering the suitability of the entered methods and their specific features, and add new methods as required or add variants of a method if the specific features required to address a particular research question or hypothesis are in conflict with those of an earlier research question or hypothesis using the same type of method.

When filling the matrix, it will become clear that selecting methods is an iterative process:

- research questions and hypotheses have to be refined or divided into sub-questions and sub-hypotheses, if it is not clear which method(s) can be found used;
- terms used have to be clarified to obtain the operational definitions needed to select a suitable method;

- specific features of the methods have to be determined, when a method can only address the research question or hypothesis if it is applied in a particular way. This may result in the need to generate a new variant of a method, resulting in an additional column;
- consequences of certain research questions and hypotheses for the features of the method, the setup and the required effort will become clear;
- the possibility of *triangulation* will be clarified, *i.e.*, the use of multiple sources and methods to gather evidence about a particular phenomenon, so as to strengthen the evidence.

Once all research questions and hypotheses have been addressed with at least one method, a selection of the most suitable set of methods has to be made. The matrix will show:

- methods that are very effective because they address many questions and hypotheses, such as the interviews with users in Figure 4.6.
  - These methods should be included.
- methods that answer some questions and hypotheses, not answered or only partially answered by other methods, such as the reliability assessment exercise and the analysis of reliability documentation in Figure 4.6.
  - Whether these methods are included depends on the effort required, the importance of the research questions and hypotheses they address, and their usefulness for triangulation. If a method is not selected, for example because it requires too much effort, this might have consequences for the set of research questions and hypotheses that can be addressed. Some might have to be left out, others might have to be reformulated to allow them to be addressed using one of the other methods in the matrix. An example of the latter is a question about the reliability of a product. If determining the reliability requires a specific method involving a large amount of effort, this might be avoided by reformulating the question so as to ask for the opinion of the designers about the expected reliability of the product. This will not require much effort and at least would give an indication of the reliability.
- methods that only answer a subset of the questions that are already covered by other methods, such as interviewing designers in Figure 4.6.
  - These methods can be included for triangulation purposes.

The analysis of the Question-Method Matrix in Figure 4.6 suggests to definitively include the first method, interviewing users. This method addresses both research questions and one of the hypotheses, and the effort is not too large.

The second method, interviewing experienced designers, overlaps with the other methods and does not provide full answers. The inclusion of this method may still be useful for triangulation purposes, but the usually limited availability of experienced designers might be a reason not to include this method. Obviously, some methods might be included later in the DS-I stage, for example to verify the data.

To decide about the third and fourth methods, it is necessary to look at the importance of the question and hypothesis they address compared to the effort they require. The reliability assessment exercise only contributes to one research question and requires considerable effort of the participants. However, it is the only method to provide a full answer to the second research question. If the question is important, the method must be included. If the question is interesting and the effort required is available, the method might be included. Before deciding not to include the method, the consequences of only having partial or indirect answers to research question 2 must be weighed. The decision about the fourth method, analysis of reliability, follows a similar line of reasoning. Here, however, the effort is mainly with the researcher, which is usually favoured compared to a method requiring considerable effort from the participants.

General Issues

In general, several issues have to be addressed before rejecting a suitable method because of the effort required:

- Is the required effort really a problem?
- Can one of the other methods be adapted to address this question or hypothesis?
- Can the method be adapted so that it can address more questions and hypotheses?
- Can the method be adapted so that it can address the question or hypothesis more effectively?
- Can the method be adapted so that less time is required from the participants?
- Can another method be found to address the question or hypothesis that requires less effort?
- If all of the above fails: Is the question so important that it justifies a separate method that requires considerable effort of the participant?

4.6.3 Detailing the Research Plan

The process of selecting the data-collection methods will have revealed some of the specific features of these methods and the setup of the study. Further detailing of the methods and the setup, as well as the development of the required data collection instrumentation, such as the recording equipment, task descriptions, questionnaires, introduction material, setting, *etc.*, is still necessary before the methods can be applied. The checklist of options in Table 4.1 can be used to determine the various dimensions of the setup that need to be determined.

A data-collection method, however, cannot be detailed without at the same time choosing the data processing and analysis methods. The research questions and hypotheses determine the type of data to be collected as well as – given this type of data – the analysis methods and possible data-processing methods. The analysis methods in turn determine how the data should be processed and the amount of data to be collected. This in turn determines the details of the data-collection method. We have seen on many occasions that the details of data processing and analysis were only determined after data was actually collected with the consequence that the research questions could not be answered and the hypotheses not verified.

Some recommendations for detailing the research plan are the following.

- It is necessary to be creative in adapting a data-collection method to the given situation, while at the same time taking care not to violate the assumptions on which the method is based. Ideas about possible variants of methods can be found in the literature on descriptive studies.
- The details of the data-collection method should match the behaviour of the phenomena investigated to avoid the method influencing this behaviour. For example, using paper in an essentially verbal environment is a change that may affect the findings and consequently the conclusions that can be drawn. This also relates to the setting in which the study will take place, such as a laboratory or a practical setting. Whether the factor of interest occurs in practice does not imply that the empirical study needs to take place in a practical setting. This depends on the actual research question or hypothesis and the operational definitions used.
- The details and scheduling of the methods should be chosen such that the whole research plan (including processing and analysing) can be conducted properly within the constraints of the project. In particular, the ways of recording the data should be chosen such that later processing and analysis is accommodated, without biasing data collection.
- The possibility of using methods in parallel and the consequences for the resources involved should be checked. If only one researcher is involved, certain methods might not be possible to use in parallel, *e.g.*, because they require different roles of the researcher or a different focus, or because one method requires continuous involvement, thus preventing the use of other methods at the same time unless additional researchers are involved.
- Data-collection methods should not be chosen just because they are easy for the participants, if not at the same time it has been assured that the data can be analysed as intended. An example was given at the end of Section 4.5.3, where participants were given multiple-choice questions containing ranges of turnover and ranges of the number of designers, in order to make it easier for participants to fill out the questionnaire. This caused serious problems during data analysis, as it was impossible to calculate turnover per designer.
- Data should be collected as directly as possible from the original source, requiring as little interpretation or translation as possible before processing. An example we observed where this was not the case is the following. In a questionnaire a 5-point sliding scale was used rather than tick boxes in order to ease the process of answering. The participants did not have to select a value or be very precise, but only had to put a vertical line somewhere on the scale. This decision caused problems during data analysis, as the researcher could not directly calculate the average values required to answer the research questions. The researcher resorted to measuring the distance from the origin of the scale to the lines put by the

participants, transferred these into values with one decimal between 1.0 and 5.0, interpreted these as the values intended by the users, and then used these to calculate the average value with two decimals. This example not only illustrates the problem of not considering data processing and analysis when developing the data-collection method, but also the invalid use of precision for the purpose of analysis, a precision not at all contained in the original data.

- Coding is used to abstract or index the collected data to facilitate retrieval, organisation and analysis. Depending on the approach taken, determining the data-coding schemes should be part of detailing the data-collection method (see Section 4.7.2).
- It has to be checked that interpretation and verification of the results is possible (see Sections 4.7.3 and 4.7.4).
- Participants involved in the study have to be contacted well before the details of the research plan are fixed, to ensure that the plan can be executed. In design research it is notoriously difficult to find participants who have the time or are allowed to participate in research. This can seriously affect the number of cases or even the quality of the data and thus the analysis method and outcome.
- Care should be taken to try to anticipate behaviour of participants that may potentially bias the findings and to address this by addressing the related influencing factors in the chosen methods, *e.g.*, through additional questions or factors to be observed, or by adjusting the setup. "The subject is not a mere passive responder to stimuli but an active participant whose perception of the total situation may profoundly affect his behaviour" (Orne 1962). Participants may, *e.g.*, try to guess what is expected of them in order to behave as a good subject (this is called "demand characteristic" (Denzin 1978; Orne 1962)), or participants may feel special because of the attention they receive and therefore work with more motivation that can strongly affect the outcome (the so-called Hawthorne effect, discussed in many sources).
- A careful analysis of the details of the data-collection method is necessary to guarantee that the collected data is indeed the data one needs to answer the question or verify the hypothesis. We have frequently observed that when research students explained their choice of setup, questions asked or tasks used, they do so by using terms that do not appear in their questions, tasks or factors they are going to observe. Often, what they want to know cannot even be derived from the chosen setup, questions or task. For example, when participants are asked to tick on a list the factors that influence the duration of the design process, this cannot be used to determine what the most important influencing factors are. The importance cannot be derived from the frequencies; some factors may be an influence in all cases, but only a small one. The question should have included the term 'most important'.
- When research questions or hypotheses require data to be related or compared, the details of the data-collection method(s) should guarantee that this data relates to the same situation, project or product. For example,

asking designers about the methods they use to assess reliability and about how often they are successful or unsuccessful in assessing reliability of a product does not allow conclusions about which method is more successful than another. Having used two separate questions, the link between a specific method and a successful and unsuccessful assessment is not made explicit. When this link is of interest, this has to be explicitly asked, or the two questions have to be linked, for example by asking how often the designers were successful for each of the methods they mentioned.

- A pilot study is *always* required to verify that the whole research plan, and not only the data collection, works as intended (see Section 4.6.4).
- Despite a pilot study, many things can happen during data collection that could threaten the investigation. Although not everything can be foreseen, putting some thoughts into contingency plans is important to avoid situations that prevent data collection as intended, or render data useless or invalid. Some situations we have encountered are the following:
  - fewer participants then expected can be recruited;
  - results are considerably different from the expected results;
  - a participant being observed in a specially prepared room decides to take a walk in the nearby park "to ponder it (the design task) over", another designer wishes to use the telephone to ask his colleague;
  - the participants ask topical questions to the researcher;
  - the researcher is used in company politics;
  - participants object against the recording of a particular event;
  - the company objects to video recordings.

Table 4.4 provides an example of some of the detail necessary to be well prepared for an empirical study.

Table 4.4 Example of some concepts and their operationalisation that were used for evaluating C-Quark, a design method for novice designers (after Weinert (2001))

Constructs	**Variables**	**Type/ amount of data**	**Method**
Task complexity	Subjective grade of complexity. Definitions of complexity: variety of tasks	Quantitative/ Ratings [once for every task = 12]	Interview with supervisor, ask him to rate the tasks.
Team satisfaction	Subjective grade of satisfaction: are you satisfied with the results?	Quantitative (n = 49)	Feedback form Z (for all three groups)
Experience in design	1. Name and year of degree 2. Previous work history	Qualitative (can be rated too (n = 49)	Background information part of feedback form I

The various terms used in the hypotheses are defined such that they can be observed: the variables to be measured are listed, as well as the type of data these represent and how much data can be expected. In the last column the method is specified. This level of detail is necessary to determine data processing and data analysis.

Reliability Example

In the reliability example it is decided:

- To focus on the relationship between reliability (mechanical reliability) and clarity, simplicity and unity of a product or assembly;
  - This resulted in a more detailed but focused set of research questions and some hypotheses.
- To undertake case studies, examine the documentation of products and assemblies with known levels of reliability and determine the levels of clarity, simplicity and unity of their embodiments.
  - This required, amongst others, operational definitions of clarity, simplicity, unity and reliability. Unity was defined using the mechanical strength of the components; simplicity was defined using the number of components and the number of interfaces. Both can be measured directly. As the literature did not provide a clear definition of clarity, it was to decided to define clarity as the average of the values between 0 and 5 (with 0 being no clarity) given by two independent experienced designers judging the product or assembly. Reliability was defined as (1 – failure probability). The failure probability was calculated from the available warranty data.
  - The focus of the case studies was on the product characteristics, rather than on the process of using the rules, as the latter seemed impossible to trace (the literature suggested that the rules are often used implicitly).
  - Data collection consisted of two parts. First the product and warranty data of different configurations of three different subsystems were analysed. Second, the documentation of the design processes of the subsystems was analysed to determine the product data available as the subsystems evolved. This enabled the clarity, simplicity and unity to be assessed at various points during the process and thus determine at which points these assessments could be made and how well these reflected the actual reliability obtained from the warranty data.
  - Apart from document analysis, the designers of the subsystems were consulted in those cases where the documentation was not clear to the researcher. All these meetings were documented.
- To use this data to develop a theory about the relationship between reliability and the three measures.
- To verify the theory using additional cases.
- To modify and verify again if necessary.
- To provide suggestions for the development of support.

4.6.4 Pilot Study

The aim of a pilot study is to try out the research approach to identify potential problems that may affect the quality and validity of the results. A pilot study is not the same as an exploratory study (see the beginning of Chapter 4) which is a proper study with the aim to study a phenomenon, albeit in an exploratory way. The need to do a pilot study before undertaking an empirical study cannot be overemphasised. Actually trying out the research as planned – including data processing, analysis, and drawing conclusions – and requesting feedback from the participants involved in the pilot study, will reveal that several changes are required if the study is to be effective and efficient. Examples of changes are: formulating less ambiguous questions in a questionnaire, changing to better quality recording equipment that has the right resolution, finding an easier way of recording that interferes less with the observed process, or adding other methods to capture aspects not yet captured or not in sufficient detail.

A pilot study usually involves only one or two cases. The setup should be as close as possible to the setup of the intended study. If the availability of participants or products is limited, one should try to avoid using the most important ones. This is particularly important when designers are involved. Most of them have little time available; their involvement therefore should be limited to the actual study where possible. Often, the opportunity to involve students or colleagues in a pilot study exists because the emphasis is on trying out the method and related procedures rather than on the actual data obtained. However, care should be taken that the collected data is relevant, because the pilot study should not only cover data collection, but also all subsequent steps. Participants in a pilot study should be asked to be particularly critical and requested for feedback on their experiences. Sufficient time should be planned between pilot study and actual study. A second pilot study with the modified research plan may be necessary.

The importance of good instrumentation and clearly defined data-collecting procedures should not be underestimated and should be tested carefully to ensure their applicability under the conditions given by the context in which the method is to be used. Poor equipment, poor conditions, and the lack of clear procedures can make subsequent analysis of the data very difficult, if not impossible. In particular, when several researchers are involved, they should be trained carefully and take part in the pilot study. No time should be saved on the preparation of an empirical study.

4.7 Undertaking an Empirical Study

In this section, we provide some guidelines for undertaking empirical studies focusing on collecting data, processing data, analysing and interpreting data, verifying results and drawing conclusions. The literature on the chosen methods should be consulted for more detail. This section ends with some guidelines for updating the Reference Model and determining further empirical studies.

4.7.1 Collecting Data

How data is collected is determined by the method(s) that have been chosen. Continuous reflection on the data-collection process is necessary and documentation is recommended.

The pre-requisite for reliable data collection is a good operational definition, and careful execution is required so that the collected data is unambiguous and can be easily processed and analysed. The recording procedure and the set-up should be realised as planned and followed throughout each data-collection activity. The preparation time required before each data-collection activity should be included in the schedule. This can include the time needed to set up recording equipment, to arrange the furniture in the room, to prepare documents, to welcome the participants, or just to concentrate oneself on the task ahead. Starting without being fully prepared, *e.g.*, because it may annoy the participant, can render the data useless. To hope that 'we'll sort this out later' may turn out to be unrealistic.

Similarly important is to plan time *after* each data-collection activity to reflect and make notes. When using recording equipment, note taking is still necessary. Apart from an overview of the process, the notes should contain reflections on the research process as it progresses, including potentially relevant events that occurred during this process, new questions and hypotheses that emerged, and descriptions of modifications to the research methods applied. When multiple researchers are involved, these notes are particularly valuable. The notes aid the interpretation of the findings, help reduce bias, and support the process of writing up the results. Appendix A4.1 discusses the types of notes that can be distinguished, many of which can be used in conjunction with methods other than observation.

Although strongly debated in the research community, we suggest starting processing and analysing at least part of the data as soon as these become available in order to verify that the methods are applicable in the context in which they are used.

Data Validity

To obtain valid data, two types of problems have to be avoided. The first type of problems are errors that occur for all cases, *i.e.*, they systematically affect the data in a particular direction. This is called *bias* (Cook and Campbell 1979) or *systematic error* (Frankfort-Nachmias and Nachmias 1996). The second type of problems are errors that affect each case in a different way, *i.e.*, they increase variability and therefore decrease the chance of obtaining statistically significant effects. The above authors call these *error* and *random error,* respectively.

Frankfort-Nachmias and Nachmias (1996) mention 3 types of what they call bias during observation.

- *Demand characteristic*: this occurs when the participants are aware of being studied and try to behave in a way that they think is expected of them (see also the earlier discussion). Their expectations may be right or wrong.
- *Experimenter bias*: this occurs when a researcher unintentionally communicates his or her expectations to participants. These expectations can be based on earlier observations, for which reason some researchers

argue for collecting all data before starting data analysis, rather than doing these activities in parallel – as we recommend earlier to be able to verify whether the research methods are appropriate. When the researcher coincidentally informs only some of the participants, this would be called error.

- *Measurement artefacts bias*: this occurs when the research methods and equipment used give participants hints as to what the researcher is after, or when the use of a measurement device does not fit the behaviour of the observed, whether participant or product. The latter happens, for example, when participants are asked to use a software programme in order to ease data collection, where they normally use pen and paper.

Systematic and random errors can occur in all stages of research. Systematic errors can be caused by the chosen theoretical perspective, the selected method, the data sources, the researcher – in particular his or her point of view – and the way the method is applied. Random error is more often caused by the way in which the method is applied, the behaviour of the researcher and inconsistencies in the data sources, all at the time of application. An interesting discussion about bias can be found in Hammersley and Gomm (1997), although we do not fully agree with their view that "researchers should resist active commitment to other goals than the production of knowledge, such as practical causes, because they are sources of motivated bias". Our methodology is based on the assumption that design research is motivated by practical causes. Awareness of potential problems based on this motivation should of course be raised and mitigation encouraged. Denzin (1978) suggests triangulation, that is, the combination of multiple data sources and research methods, application of different theoretical perspectives, and use of multiple observers to reduce or at least detect bias and error.

4.7.2 Processing Data

Before data can be analysed, it has to be processed. This may involve tasks such as transcribing tapes or hand-written notes (it is wise to do this as soon as possible after the data has been collected), putting data in spreadsheets, tagging segments of interview data or video sequences, labelling photographs, or identifying elements in graphical representations. The careful selection of data collecting and processing equipment can save much time. Data processing can be very time consuming: a detailed transcription of a think-aloud session recorded on video will require around 8 hours per recorded hour. A detailed transcription of a meeting of two or more people will considerably increase this effort. Talking to other researchers about their experiences and the equipment used is very worthwhile.

Coding Schemes

Processing data often involves coding the data to abstract or index the collected data in order to facilitate retrieval, organisation and analysis. Codes that *abstract* the data are intended to be used for analysis instead of the original data. Codes that *index* the data, as is often the case in qualitative research, are mainly intended to facilitate retrieval and organisation of data elements that can then be analysed together. Coding has to be done carefully as details will be lost.

Codes are categories, usually derived from research questions, hypotheses, key concepts or important themes (Miles and Huberman 1984). Categories can come from the researcher as well as the participants involved, also referred to as *outsider* or *insider* approach (Patton 1987). Categories can be *pre-defined* (also called pre-set or deductive coding) or *post-defined* (also called emergent or inductive coding), *i.e.*, the coding scheme can be developed before or after data collection. Pre-defined coding is typical for a theory-driven approach. In design research, an often used pre-defined categorisation for studying design processes are the main steps of the design processes proposed in methodologies such as in Pahl and Beitz (2007) or VDI-2221 (VDI 1993). Examples can be found in Hales (1987) and Fricke (1993b). Post-defined coding is typical for a data-driven approach; the codes emerge during data analysis. Examples of this type of coding in design research can be found in Ahmed *et al.* (2003; Sarkar (2007).

Quantitative data can be used directly or coded into categories. For example, when assessing the reliability of a product on a scale from 1 to 10, these values can be used directly or coded using ranges such as '< 3', '3–6', and '> 6' or descriptions such as 'low', 'medium' and 'high' reliability. If coding is used, we recommend to always collect data as detailed as seems necessary and always to keep the original data in order to be able to go back to this data during data analysis. Note that the more descriptive categorisation contains an interpretation: the values are translated into an assessment of the reliability as being high, medium or low. This is not the case in the categorisation using ranges, where no assessment is made of whether a particular range is low or high. Which categorisation is more suitable depends on the research question, the way in which the data is collected and further available information, which may allow an interpretation, such as the one discussed above. Categories can also be based on a combination of data, *e.g.*, using 'low-medium-high' levels of experience to replace the two data sets 'actual number of years designers have been working in their job' and 'experience with the particular type of product'. In all cases, the categories and how these have been derived have to be described in sufficient detail for others to understand.

Qualitative data is often categorised or labelled, using easy-to-remember abbreviations that are then used to retrieve related data. However, bringing together related data has the disadvantage that it takes this data out of its contexts. This "does not facilitate an accurate documenting of [observed] processes taking into account both temporal sequencing and group interaction" (Catterall and Maclaran 1997). This could be a major disadvantage when studying design where temporal sequencing and group interaction are important. Opinions differ as to the most appropriate method to prepare qualitative data for analysis. Where early analysis software only allowed retrieval based on coding, new developments allow text and image retrieval, text management, conceptual network builders, *etc*.

When qualitative data is coded, the data has to be explored and interpreted sensitively to avoid pre-emptively reducing the data to numbers and losing the richness of the data. Qualitative data can be quantified, *e.g.*, by classifying and ranking the data, but whether this is appropriate or not, depends entirely on the issue that is being addressed and the setup of the study.

When qualitative data is quantified, extreme care has to be taken that once numerical values are assigned, these are analysed in accordance with the type of data. The fact that a number can be assigned to a category, *e.g.*, 1 = male: 2 = female, or 1 = low; 2 = medium; 3 = high complexity, does not imply that mathematical operations can be applied. This is most clear in the male-female example: doing calculations obviously does not make sense; the numbers 1 and 2 are labels, not real numbers. This is less obvious for the complexity example (low-medium-high). Calculating the average complexity seems to make sense. However, the calculation of an average is based on the premise that the distances between the numbers are equal and that there is a natural zero (see below). This implies that $3 = 3 \times 1$, $2 = 2 \times 1$ and $3 = 1.5 \times 2$, *i.e.*, high quality is 3 times as high as low quality, but only 1.5 times as high as medium quality, and medium quality only 2 times as high as low quality. The category labels can obviously not be used for this calculation. The scales or levels of measurement have to be considered.

Scales or Levels of Measurement

Data that has been coded can be analysed for such features as dependencies between variables and strengths of relationships. To select the right analysis method, the way data has been measured and coded is important. Four scales can be distinguished.

- Nominal scale.
  - This non-metric or topological scale represents qualitative properties, the order of which does not play a role. For example, gender (female; male), profession (1 = mechanical engineering; 2 = civil engineering; 3 = software engineering) or lubrication (none; grease; oil).
  - Relations can only be defined in terms of equalities (= , ≠).
  - Calculations other than frequency counts for each category and the mode (category with highest frequency) are not allowed, even if the categories are given numerical values such as in the example above.
  - Typical representations are bar charts and pie charts.
- Ordinal scale.
  - This is the second non-metric or topological scale and represents qualitative properties that can be *ranked*, but the distance between the categories cannot be said to be equal, if known at all. Furthermore, the numbers do not represent absolute quantities. Examples are experience level (novice; intermediate; expert), or growth rate (1 = below sector average; 2 = average; 3 = above average; 4 = leading).
  - Relations can be defined in terms of equalities (=, ≠) as well as inequalities (< , >).
  - Apart from frequency counts, the median and centiles can be calculated.
  - Bar charts are more suitable than pie charts, because tendency can be observed more easily.

- Interval scale.
  - On this metric scale, the distances between the categories are known and equal; the numerical codes do have a meaning *relative* to each other, but the scale does not have a *natural* zero point. For example, ratings of quality based on degree of fulfilment of a particular set of requirements: 0–25; 26–50; 51–75; 76–100.
  - Relations can be defined in terms of equalities (= , ≠), inequalities (< , >), as well as using addition and subtraction.
  - Apart from the operations mentioned above, the values can be added together or subtracted, and the average (arithmetic mean) can be calculated. Any change in numbers must preserve the relative difference, for example by changing from absolute into percentages.
  - Line graphs can be usefully applied here.
- Ratio scale.
  - This is a metric scale with equal distances between categories and a natural zero. Examples are cost, number of design staff, and many of the physical properties.
  - Relations can be defined in terms of equalities (= , ≠), inequalities (< , >), addition and subtraction, as well as multiplication and division.
  - This data allows all sorts of calculations such as geometric mean, variance and ratio.
  - All types of representations can be used.

Data can always be coded at a lower scale (the nominal scale being the lowest) than the scale at which it was originally measured, and can always be represented at a lower scale than the scale at which it was coded. For example, reliability values can be grouped into low reliability, medium reliability and high reliability, resulting in an ordinal scale. Obviously this will reduce the amount of detail. As discussed earlier, representing and coding data on a higher scale then measured should be avoided.

According to Frankfort-Nachmias and Nachmias (1996):

- coding categories must be mutually exclusive (note that some literature allow dual coding);
- the coding scheme must be exhaustive, *i.e.*, able to categorise all data;
- categories must be specific enough to capture differences using the smallest possible number of categories.

Miles and Huberman add the following suggestions for qualitative data (Miles and Huberman 1984):

- codes should have some structural and conceptual order, *i.e.*, there should be some logic behind the categories and the order in which they are listed: this will help coding and determine the exhaustiveness of the categories;
- definitions should be given;
- abbreviations are easier to use than numbers;
- codes should be able to be put on 1 single sheet.

In our opinion, coding schemes should also be developed such that a distribution of the data over the categories can be expected: 90% in one category and the remaining 10% distributed over several other categories suggests that the variable is not distinctive enough, the categories are chosen incorrectly, or the one large category covers too many different aspects. If the categories were pre-defined, a more detailed category scheme may have to be set up after the first data analysis.

Coding Process

It is necessary to document the coding process in detail, *e.g.*, by adding examples to the definitions of the categories. This is particularly important when data elements are found difficult to categorise, for example because they seem to fit in two categories. Adding the example to the definition, this can act as a reminder when coding similar data elements, thus preventing these from being categorised differently. Analysing the definitions and the examples will also help sharpen the definitions of the categories. Dual coding, that is when two or more codes per data element are being used, can seem a solution while coding the data, but can make data analysis far more difficult. Instead of dual coding, it would be more useful to a coding scheme, or modify the scheme to include a new code covering the dual code. To avoid forcing an element into a particular category, a category named 'other' can be included for those data elements that cannot be coded, or for which it is not clear which category is most appropriate. The 'other' category should then be analysed later and the elements be re-categorised where appropriate. Every change to the category scheme requires the already coded data to be checked to ensure that the categorisation of the data elements is still correct.

It is, furthermore, useful to mark data elements that are particularly interesting, for example because they differ from what was expected or illustrate a particular point very clearly. During coding, ideas about patterns in the data will emerge. It is important to write these down with reference to the relevant data elements. These ideas will have to be verified by finding sufficient evidence, and one has to accept that many have to be rejected, or at least reformulated.

The traditional preference for quantitative data is based on the availability of mathematical methods for processing. The availability of software tools further eases the processing of large quantities of data. Nowadays, powerful software is also available to support qualitative data analysis. This software assists with indexing of text and video data, searching data with the same index, documenting emerging interpretations, and building concept trees. Some software is able to identify the hypothesised links between categories and concepts in the data. It is important to be informed about the methodological basis of the software, because different methodologies require different ways of handling and interpreting data.

Inter-encoder Reliability

In particular when coding qualitative data, it is important to start with double coding. *Double coding* involves coding of at least a part of the data by two different people or by the same person twice but with a time delay in between. This will help sharpen the definition of the codes. Double coding also allows the calculation of the *inter-encoder reliability*, which should be higher than 70% to be acceptable:

$$\text{inter-encoder reliability} = \frac{\text{number of agreements}}{\text{number of agreements + number of disagreements}}$$

From our experience we suggest:

- to check the inter-encoder reliability relatively early in the coding process to avoid having to recode too much data when changes or redefinitions in the coding scheme are required;
- to ask each encoder to mark where coding was difficult or unclear, *e.g.*, when a data element could not be categorised, when doubts arose about the correct category, or when a data element could be coded using multiple categories;
- to discuss after double coding the differences between the assigned codes and the marked elements, in order to better define the categories, to merge or split categories, to create new categories or even new category schemes.

4.7.3 Analysing and Interpreting Data

Summarising, organising and presenting data in graphical, tabular or matrix form provides an overview of the data and a good starting point for analysis. Analysis will often start with simple enumeration, some descriptive statistics, or summaries of the data. This is followed by a more detailed analysis linking the findings, identifying relationships and possible correlations or even causal relationships, findings explanations for the findings and drawing inferences. If inferences are to be drawn beyond the cases involved in the study, inductive statistics is required (see Appendix A.7.1)

Quantitative data allows statistics that makes "summaries, comparisons, and generalisations quite easy and precise" whereas qualitative data are "typically meant to provide a forum for elaborations, explanations, meanings and new ideas" (Patton 1987). Note that qualitative data can be the source of quantitative analysis: counts of key categories and measurements of the amounts of variables are possible if coding has taken place. Irrespective of the type of data collected – quantitative or qualitative – an appropriate representation is needed to support the analysis. Most of us are familiar with representing numerical data, and standard software packages, such as spreadsheets, are able to produce a wide range of graphical representations from quantitative data. The aim to maintain the richness of the original data makes analysis of qualitative data a complex, and potentially very subjective task. Miles and Huberman (1984) is one of the most extensive publications on the possible ways of representing qualitative data to support analysis, in particular to draw meaning.

Many books about quantitative data analysis exist, and packages to support statistical analysis, such as SPSS have been around for many years. They usually depend on large data sets, normal distribution and coded data. The data resulting from design research, however, is often different. The number of cases can be very low, the data may not be numerical, data may be missing, and the distribution may be unknown. In Appendix A.7 some relevant terms are introduced and guidance is given for the selection of suitable statistical methods for quantitative or quantified data that has such characteristics.

Although "qualitative data are attractive, [..and..] are a source of well-grounded, rich descriptions and explanations of processes occurring in local contexts [..] the most serious and central difficulty in the use of qualitative data is that methods of analysis are not well formulated" (Miles and Huberman 1984). The situation has improved, but no generally accepted methods of analysing or even representing qualitative data exist: many methods are specifically linked to particular paradigms and heavily debated by those following other paradigms.

The development of specialist software packages for qualitative data analysis has had a major impact, in particular reducing the amount of effort in analysing the data. Nevertheless, Lee and Fielding (1996) warn about the use of computer software for qualitative data analysis because there is an issue "about what background one might need to produce meaningful interpretations" from such software. "Faced with an apparently smooth and user-friendly resource offering all manner of subsidiary and supporting information, the naïve user may feel that it contains 'all there is to know' about the topic at hand". Obviously, the same warning is applicable to quantitative data analysis using statistical packages. These too can produce impressive results based on data that was unsuitable for the statistical method used. Overall, however, the use of the available software packages for data analysis has been a great help for handling large quantities of data such that these can be analysed.

Miles and Huberman (1984) suggest 12 specific tactics for drawing meaning from a particular representation of data, grouped into:

- to see what is there;
- to see what goes with what, to integrate and differentiate data;
- to see things and their relationships more abstractly; and
- to assemble a coherent understanding of the data.[14]

The aim of data analysis is to draw valid inferences about what has been observed and to avoid any *spurious relationships*. The term spurious relationship is used when an observed relationship is actually caused by a factor other than those described in the relationship. Sometimes spurious relationships are easy identified, because the finding is not plausible, for example when a significant correlation is found between the amount of grey-haired designers have and the quality of their designs. Obviously experience is the underlying cause, affecting both variables in the observed relationship. In many cases, however, the spurious variable and hence the spurious relationships may be very hard to detect.

Drawing inferences is a process that needs careful consideration and detailed attention. King *et al.* (1994) argue that although it is usually easier to draw inferences from quantitative data, both qualitative and quantitative research can use the same underlying logic of inference. They emphasise that the rules of scientific

---

[14] Note that these five labels indicate the successive steps in data analysis and interpretation.

inference can and should be applied in both qualitative and quantitative research. Using these rules should improve the reliability, validity, certainty, and honesty of the conclusions.

Inferences can be

- *descriptive*; using observations to learn about other unobserved facts, such as motivation, which can not be observed directly but only through a combination of other observations;
- *causal*; learning about causal effects from the data observed (King *et al.* 1994).

Causal relationships are very important, if one is interested in improving a particular situation. Identifying *causality* requires evidence of:

- time order between concepts: the cause has to happen before the effect;
- covariance: a high degree of relationship between the concepts has to exist;
- the exclusion of rival factors: spurious variables should not be the cause of the observed relationship.

To infer causality can be particularly problematic in situations, such as in design, that cannot be controlled by the researcher or only to a limited extent, or in cases where multiple causation occurs. The more open the system, the more fallible the causal inference will be (Cook and Campbell 1979). In their book, Cook and Campbell discuss at length the problems with causality and propose ways to improve the validity of inferences that can be drawn through appropriate planning of the empirical studies. They focus on what they call quasi-experimentation for those situations where true experiments are not possible or not suitable (see Appendix A.4.3 for more details). Their book provides an interesting overview of the concepts of cause in several paradigms.

Ericsson, as well as Miles and Huberman, discuss a number of problems in interpreting data, specifically related to qualitative research. Bias in encoding of protocol analysis is discussed in Ericsson and Simon (1996), who distinguishes between bias resulting from the encoders having prior knowledge of the hypothesis being tested, and bias in the inferences made, resulting from the encoders assuming that subjects will think in the same ways they do. Miles and Huberman (1984) list three archetypical *sources of bias* in qualitative research:

- the holistic fallacy: interpreting events as more patterned and congruent than they really are;
- elite bias: over-weighting data from articulate, well-informed, usually high-status informants;
- going native: losing one's perspectives and being co-opted into the perceptions and explanations of local informants.

In general, all possible evidence from the collected data as well as the literature should be used to answer the research questions and test the hypotheses. As many *rival or alternative explanations* as possible should be generated. These explanations can be based on existing evidence and on reasoning, *e.g.*, related to the inherent limitations of the study. As discussed in Section 4.5.2 the existence of possible alternative explanations should already be considered during the development of the research plan in order to collect data to verify these alternatives. However, this is not sufficient. When analysing the data, it is important to consider alternative explanations for all findings, whether they confirm one's expectations or not. Discussing findings with others is very helpful. Different viewpoints will lead to different possible explanations. To choose the most likely explanation(s), different findings, possibly from using different methods, may have to be combined, or further data may have to be collected. If any plausible, alternative explanations remain, the original explanation cannot be accepted other than as a *possible* explanation. Usually, the available resources will not allow verification of every plausible alternative explanation in a given research project. These explanations should be documented as the input for further research. Thus, the result of a project may contain sets of possible explanations. As long as this set is smaller than at the start of the research, our understanding of what has been observed has increased and the study has made a contribution.

4.7.4 Verifying Results

Verifying results involves making judgements about the plausibility and credibility of evidence. Two types of problems can be distinguished: bias or systematic error, and error or random error as discussed in Section 4.7.1. Each type of problems can occur either due to those circumstances that the researcher cannot control or only with difficulty, and those circumstances that the researcher could control. Both influence the validity of the results.

Miles and Huberman (1984) suggest 12 tactics for verifying the results in order to confirm conclusions divided into four groups:

- assuring the basic quality of the data;
- checking findings by looking at differences;
- taking a sceptical, demanding approach to emerging explanations;
- getting feedback from informants.

Two aspects are important when verifying results: reliability and validity. *Reliability* is the reproducibility of measurement. *Validity* is the degree to which the measurements actually reflect the true variation in the outcome of interest. Apart from validating the individual statements that were made based on the findings, DS-I requires the validation of the Reference Model, bringing all findings together. According to the American Institute of Aeronautics and Astronautics, model validation is "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model" (AIAA 1998). Various publications exist in the different disciplines about methods to validate models. In this section, we will focus on the term validity and its different types.

The following paragraphs are based on the very informative discussion about validity of Cook and Campbell (1979), which we have adapted for design research. Their discussion focuses on experimental and quasi-experimental research investigating the influence of something (called a treatment), on finding causal relationships, and on the use of statistics. That is, the authors focus on hypothesis testing. The focus in design research is often different, but we consider their discussion still useful when descriptive inference is the aim and when qualitative data has been collected that can or cannot be coded into quantitative data. Some of the problems Cook and Campbell address relate more to the evaluative research discussed in Chapter 6. For reasons of clarity, the discussion about validity is kept together in this chapter rather than divided over Chapters 4 and 6.

Validity, according to Cook and Campbell, is the best available approximation to the truth or falsity of propositions, including propositions about cause. Validity is an approximation because we can never know for certain what is true. Cook and Campbell distinguish 4 types of validity and discuss the threats to these four types. Recognition of the threats can help to reduce or eliminate the threats, for which Cook and Campbell provide a number of suggestions. We have chosen to focus on those threats that seem most relevant for design research and to add examples from design research, in order to help generate a critical attitude towards the reader's own research approach and findings.

The four types of validity of Cook and Campbell are based on the four major decision questions for researchers looking for causal relationships:

- Is there a relationship between variables? (statistical conclusion validity)
- If so, is it plausibly causal? (internal validity)
- If so, what is involved in the relationship? (construct validity)
- Can this be generalised across persons, settings or times? (external validity)

Statistical Conclusion Validity

Covariation is a necessary condition for cause, that is, the first thing to determine is whether the variables are related.

Threats to statistical conclusion validity are, amongst others, the following:

- Most tests require that certain assumptions be met if the results of the data analysis are to be meaningfully interpreted. Examples are: normal distribution of the sample, a certain level of data, *e.g.*, ordinal, and a minimum number of cases. Some statistical packages do check the basic criteria, others do not.
- The implementation of the treatment, that is the way in which the experiment is conducted, can be unreliable, *e.g.*, variation between cases due to a lack of standardisation.
- The measures themselves could be unreliable.

Internal Validity

Internal validity is concerned with the issue of whether a relationship is causal. The essence is to account for alternative interpretations of a presumed relationship involving other variables. Furthermore, the relationship might not be existing or be quite different, for example, cause and effect could have been interchanged.

Threats to internal validity are, amongst others, the following:

- Something happens between the first point of measure and the second subsequent measure. For example, in an industrial setting, the organisational structure or the market situation of the company may change during the study.
- The participants mature. This is a serious problem in design: an experiment cannot be repeated using the same task, as the task will no longer be a design task, in the sense of creating something new.
- The instrumentation matures, for example, the observer or interviewer becomes more experienced.
- Groups that are compared are different in one or more aspects. Some of these may be known, others not. Randomisation may not always be a cure when using small sample sizes, as we found out (Blessing 1994), when after random allocation the designers in the control group turned out to be more experienced than those in the experimental group, making it difficult to verify one of the hypotheses.
- Different groups or cases experience the above threats in different ways: companies, *e.g.*, are different.
- Ambiguity exists about the direction of causal inference; does A lead to B or does B lead to A?
- Information about the study is passed on from one participant to another before the latter has participated.
- Cases are selected using a pre-test (see Appendix A.4.3) that is not reliable.
- Participants who are involved as a benchmark (see Control group, Appendix A.4.3) may be resentful, if they are aware of what the other participants are receiving. This may play a role in DS-II, when design support is introduced only to some of the participants to allow comparison.

Statistical conclusion and internal validity are both internal to the study, that is, they are based on avoiding drawing false positive or negative conclusions about causal hypotheses. These represent a more deductive process of inference. The following two types of validity are external, concerned with whether a presumed causal relationship can be generalised to and across alternative measures of cause and effect, and across different types of persons, settings and time. This represents a more inductive process of inference.

Construct Validity of Causes or Effects

This validity relates to the process of making generalisations about higher-order concepts or constructs from the findings that have been measured. These are generalisations across exemplars of particular causes and effects. Constructs, as mentioned earlier, are theoretical concepts that cannot be measured directly, but can be measured by measuring certain characteristics of behaviour and background. The question is whether the findings about these characteristics indeed say something about the construct itself. One of the problems is that these characteristics may relate to more than one construct. Thus, if we find cause and effect relationships between constructs and these have characteristics in common, it will be difficult to determine causal effects. Constructs should preferably be defined and measured such that generalisation is possible. For example, measuring experience using participants working between 10 and 20 years in a particular company, will allow fewer generalisation then when participants were involved that have a minimum of 5 years working experience and varying backgrounds.

Threats to construct validity are, amongst others, the following.

- Inadequate measures due to inadequate definition of constructs.
- Constructs based on a single characteristic. Instead, multiple characteristics should be used, and additional data gathered from alternative measures.
- Bias due to using only one method.
- Influence of the participants (see also Section 4.6.3), *e.g.*, when they try to behave as a good subject ('demand characteristic'), or when they are especially motivated because of the attention they receive ('Hawthorne effect'). We also found a negative effect when participants were afraid their design competence would be judged and this judgement used by others, or when they were not certain whether to tell the truth about design errors.
- The expectations of the researcher. These can directly bias data collection, analysis and in particular interpretation (see earlier example), but also affect data collection indirectly when these expectations are conveyed to the subjects. Examples of the latter are asking leading questions or empathising too strongly with the interviewees.
- The conditions under which the study takes place. The can make it difficult to generalise the findings across settings. Examples are generalisations from observations in a laboratory setting to a setting in practice, or from the working behaviour of individuals to that in teams.

External Validity

Generalisations can be (1) *to* particular target persons, settings and times, and (2) *across* these. The aim is to determine whether the results are person, setting and time independent.

The first type of generalisation is possible if the study is based on a well-drawn sample of a particular group that is randomly assigned. The groups are equivalent and represent a population with that characteristic, *e.g.*, SMEs. It is therefore possible to generalise *to* this part of the population, *i.e.*, to SMEs in general. This type of generalisation is often associated with large-scale experiments. When a questionnaire has been sent to a representative set of companies, it is necessary to verify whether the returned questionnaires are still representative. It is very well possible, that only a certain type of companies is interested in the topic of the questionnaire. Questions should therefore be added to identify the characteristics of the companies that react.

The second type of generalisation refers to sub-populations. Although generalisation can be made *to* a specific population, such as SMEs, it may not be possible to generalise the results across the subpopulations. For example, SMEs may differ on certain aspects from large companies, but certain types of SMEs may differ more from large companies than other types of SMEs.

In design research, the generalisation *to* target populations may be rare, in particular where data is collected in an industrial setting. Formal random sampling for representativeness is rare in field research and thus for this type of setting. "The practice is more one of generalising *across* haphazard instances where similar-appearing treatments are implemented" (Cook and Campbell 1979).

Threats to external validity are, amongst others, the following.

- Selection of subjects or objects that are participating is non-representative. The subjects may be the people that are volunteering, because they are interested, have time, *etc.*, or the objects may be the products the company is interested in. One way of counteracting this is to make it as convenient as possible to participate. Another way is to afterwards check against population statistics, although these may not capture the characteristics that made the participants participate.
- The setting has an effect. A common problem in design is the generalisation of findings from a laboratory setting to practice.
- The point in time at which the study takes place has an effect. It may not be possible to extrapolate the present findings into the future. This can be due to changing technologies and related different ways of working. Even mundane factors such as the mood of the subjects on a particular day, may have an effect. In some design research, questionnaires were used to gauge the mood.

The four types of validity are related. For example, carrying out randomised experiments may increase internal validity, but companies who are willing to participate in this type of study may not be representative, thus decreasing external validity. Which validity is most important depends on the research aims. Planning research always involves trade-offs, requiring prioritisation. For testing theories, internal and construct validation are likely to be the most important. In applied research, where the aim is, *e.g.*, to determine whether the situation has improved after the intervention, less interest in the causal details of the intervention may exist: the main thing is that it works. In general, however, both internal and external validity are important. Internal validity is always high on the priority list because it forms the basis for external validity. Internal validity is strongest in experiments. This may explain the emphasis of Cook and Campbell on quasi-experimentation, striving for a situation fulfilling as many of the premises of experimentation as possible, only releasing those premises that really cannot be met.

4.7.5 Drawing Conclusions

It is important to draw conclusions that are in line with the research questions and hypotheses, the data collection, processing and analysis methods, and the research setting. In general, it is better to err on the safe side: the number of cases used in design research does not usually justify wide generalisation or provide proof of the kind sometimes suggested. This is an issue of responsibility: one needs to realise that others will use the conclusions in their research. No one likes to base his or her research on a strongly formulated, but actually weak premise. Therefore, phrases such as 'the designers *observed spent* 30% of their time on gathering information', is usually better than the generalised ‘designers spend 30% of their time …’. Similarly, it is usually more appropriate to state that ‘the findings *support* the formulated hypothesis’, rather than ‘the findings *prove* the hypotheses’.

A distinction should be made between statistical significance and relevance. Statistical analysis can reveal the statistical significance of a finding, but not whether the finding is relevant or not. In one of our own studies, a statistically significant difference was found between two groups of designers regarding the time they spent on erasing their writings and sketches. However, the actual time they spent on erasing was less than one per cent of the time they spent designing, which – given the aims of this study – made the difference irrelevant. As John Dewey put it (Star 1997) “a difference that makes no difference is no difference”.

Reliability Example

For the reliability example the results can be summarised as follows:

- A significant positive correlation was found between the level of clarity and the level of reliability of the products investigated, provided that the levels of simplicity and unity were at least adequate.
- The correlation between the level of simplicity and the level of reliability was less strong but still significant, provided that the levels of clarity and unity were at least above average.
- For the correlation between the level of unity and the level of reliability of the products investigated only a tendency could be observed. In general, the level of unity (expressed by its mechanical strength) was high.
- Clarity, simplicity and unity of the products investigated can be assessed using documentation from the early embodiment design phase, even though clarity is currently assessed by experienced designers rather than a rule.
- In the cases in which the embodiment design of the products investigated scored high on clarity, simplicity and unity, the reliability of the product was not necessarily high. A possible reason was found using the Initial Reference Model: poor product reliability due to poor detail design. This was verified using the available interview data and found to be the case.
- Conclusion: For the products investigated, the combination of high levels of clarity, simplicity and unity (as defined in this study) correlated with a high product reliability of a product. Clarity had the largest effect. Simplicity had an effect but is not an absolute measure and therefore only relevant when comparing products. Unity is relevant but not found to be a problem in the investigated cases.

4.7.6 Updating the Initial Reference Model

After each empirical study the Initial Reference Model and if necessary the Initial Impact Model are updated to represent the level of understanding obtained. Assumptions may be confirmed or rejected, new influencing factors may have been identified, links may have to be added, removed or modified, Key Factors and

Measurable Success Criteria may have to be changed, and even the Success Criteria may have to be reconsidered.

4.7.7 Determining Further Empirical Studies

The outcomes of each empirical study and the status of the updated Initial Reference Model will give rise to new questions and hypotheses. When new factors emerged or the outcomes were rather unexpected, the questions and hypotheses might be quite different from the original ones. In many situations the new questions and hypotheses go more in depth: from an understanding of what, to an understanding of how and why. Whether a further empirical study to address these questions and hypotheses is required within the research project depends on whether the level of understanding that has been obtained is sufficient to proceed with a PS. When an empirical study was not successful in addressing the research questions or hypotheses, it might be necessary to repeat the study in a different form or select another set of methods.

Reliability Example

In the reliability example the first empirical study based on case studies showed that it was possible to *assess* clarity, simplicity and unity from the product documentation using the formulated operational definitions (see earlier discussion) and that the results of the assessment related to reliability. However, whether the rules as they are presented in the literature are applicable to *generate* high levels of clarity, simplicity and unity, had not been part of the first empirical study. The researchers decide that without this information it does not make sense to develop a method to support reliability assessment. An additional empirical study is required.

To develop this second study, the same approach as for the first study was taken (see ‘Example’ in Section 4.6.3). This resulted in an empirical study in which two groups of six designers were each given a sketch of a concept and asked to use this to produce a rough layout. The designers in the so-called experimental group are given a description of the basic design rules derived from the literature and asked to apply these while designing. The designers in the control group are only given the design task. The task is based on one of the cases in the empirical study. The concept sketches are directly taken from the documentation of that case. This task is chosen to obtain some indication of the expected reliability of the embodiments created by the designers. All designers work individually and are asked to think aloud. The processes are videotaped. No time constraints are imposed. A pilot study showed that a designer needs about 4 hours. Questionnaires before starting and after finishing the task are used to collect data about the designers and their opinion about the setting and the task. The experimental designers are also asked about their understanding and application of the given basic design rules.

This second empirical study showed that the designers had problems with using the clarity rule. This rule was considered easy to understand but too ambiguous when really applied. In particular, the fact that no clear measures of clarity exist was considered problematic when the designers tried to use the clarity rule to improve the product. The simplicity rule was equally easy to understand but still difficult to use because of the lack of a benchmark. When variants had to be compared, the rule could be used. To apply the unity rule, the designers resorted to existing methods for calculating strength. In the questionnaires several designers commented upon the fact that they understood unity in a wider sense than the provided documentation suggested.

Overall, the expected reliability of the embodiments of the experimental designers tended to be higher than that of the control designers, but the difference was not significant. This suggests that the application of the rules does not have a significant effect. This seems in contradiction to the results of the first empirical study, which showed a clear correlation between the level of reliability and the levels of clarity, simplicity and unity of a product. A possible explanation can be found in the literature. Some studies suggest that designers do apply these rules, albeit without being aware of these. Hence, it is likely that the control designers in the second study also applied the rules, although they were not instructed to do so. This might have caused the observed lack of difference between the reliability of the solutions in the second study, in particular because the design tasks that were given were relatively easy so as to allow the process to be observed.

The researchers decide on the basis of the available evidence, that the current understanding is adequate enough to decide on the focus of the PS; there seems to be a need for a clearer method to assess and improve in particular clarity and simplicity. No further empirical studies are considered necessary.

4.8 Drawing Overall Conclusions

Once it has been decided that, at least for the project undertaken, no more empirical studies are necessary or that time does not allow further empirical studies, final conclusions have to be drawn for the DS-I stage. This requires various steps, which are described in this section.

4.8.1 Combining Results of Empirical Studies

If multiple studies have taken place, their results have to be compared and combined in the light of the goals, research questions and hypotheses of the project. Similarly, a comparison and combination with findings from the literature is required, to identify supporting evidence and possible contradictions. Any concluding statements should take all of these into account, as this will influence the strength of the statements and hence the formulation of the conclusions. Unfortunately, we regularly find statements generalised beyond what the findings allowed. More details about drawing conclusions were discussed earlier in Section 4.7.5.

Our reliability example described earlier showed how the combination of the findings of different empirical studies and the findings in the literature led to a conclusion that could not have been reached had only one study taken place. The studies focused on two different but complementary aspects. The first study focused on the link between the levels of clarity, unity and simplicity on the one hand, and product reliability on the other. The second study focused on the link between the use of the rules and product reliability. The combined results suggested that clarity, unity and simplicity are useful concepts, but that their assessment and improvement, by means of the rules, need to be addressed.

Summarising the findings using a table of the main statements not only supports the drawing of conclusions, but also provides an excellent overview for the research community as a whole. A comparison of the findings with the findings from the literature should provide concluding statements as well as possible explanations for any differences and similarities that were found. The checklist for reviewing empirical studies suggested earlier (Section 4.4.2) can be useful for finding alternative explanations caused by a different context, a different aim or different methods.

4.8.2 Completing the Reference Model and Updating the Initial Impact Model

During DS-I the Initial Reference Model (see Figure 3.10) is continuously updated using the findings from the literature and one's own empirical studies. Once the empirical studies have been completed, the Reference Model can be finalised and the Initial Impact Model updated, as will be illustrated using our reliability example.

Reliability Example

Figures 4.7 and 4.8 show the upper and lower part of the final Reference Model for the reliability example. Compared to the Initial Reference Model, the Key Factor has moved from 'product reliability' to 'embodiment reliability' due to an increased understanding of causes and effects. A possible link between early failure detection (upper part of the figure) and early assessment (in this case reliability of embodiment rather than of detail design) is not included for reasons of clarity and because this link was considered less relevant. The aim is to focus on the chain of causes and effects associated with improving product reliability product quality, customer satisfaction and market share, and not on the chain of effects associated with a reduction of iterations and lead time.

Based on the Reference Model, the Initial Impact Model can be further detailed. Figures 4.9 and 4.10 show the resulting upper and lower part of the Initial Impact Model for our reliability example. As discussed earlier, clarity had the largest effect, simplicity had an effect but is not an absolute measure, and unity is relevant but not found to be a problem in the cases investigated. The support should therefore focus on clarity and simplicity, with an emphasis on clarity. Furthermore, it is assumed that when the embodiment is more reliable, the number and size of modifications needed after the use of the Design for Reliability methods in the detailed design stage will be reduced and the amount of time left sufficient. The factors 'quality of components' and 'quality of production' were considered not to be problematic within the company, and 'motivation of use' as a factor that cannot be influenced by the researchers, so that those factors are not considered (shown with dashed lines). Note that at this stage, the Impact Model is still an Initial Impact Model, which will be finalised in the PS stage.

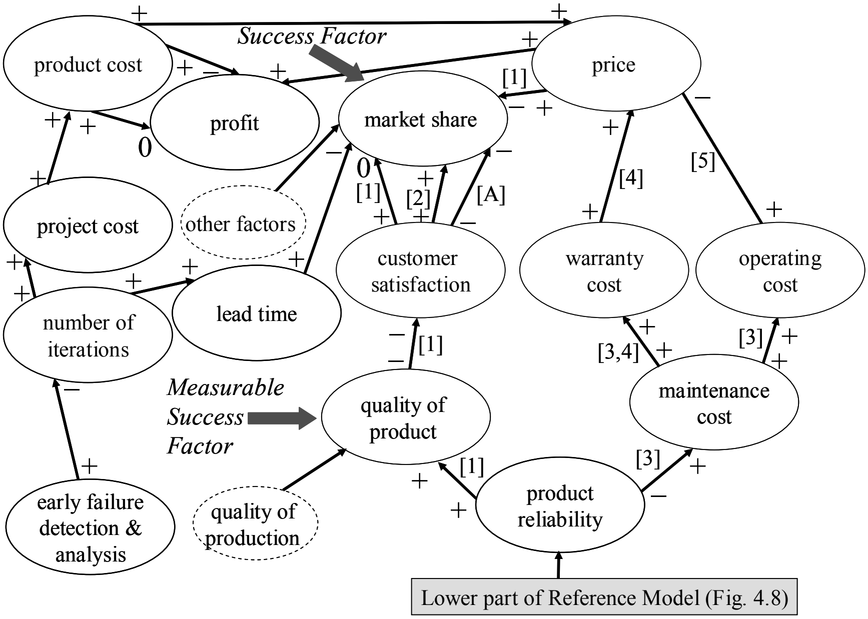

Figure 4.7 Upper part of the Reference Model resulting from the DS-I stage

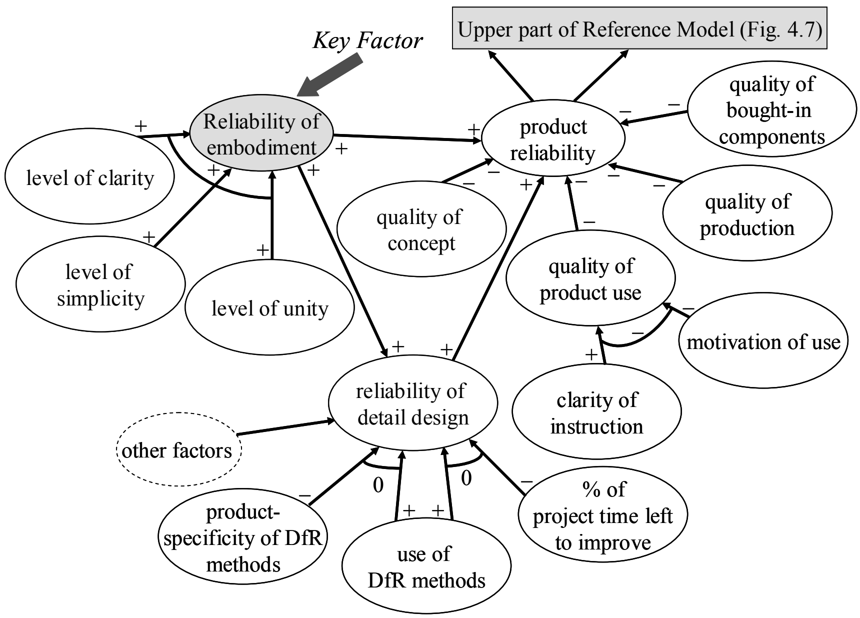

Figure 4.8 Lower part of the Reference Model resulting from DS-I

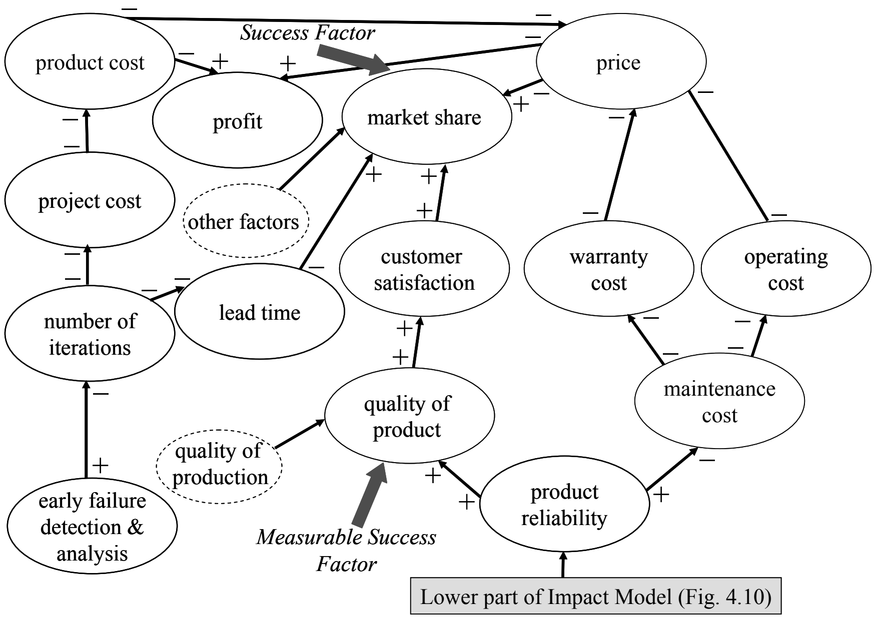

Figure 4.9 Upper part of the updated Initial Impact Model resulting from DS-I

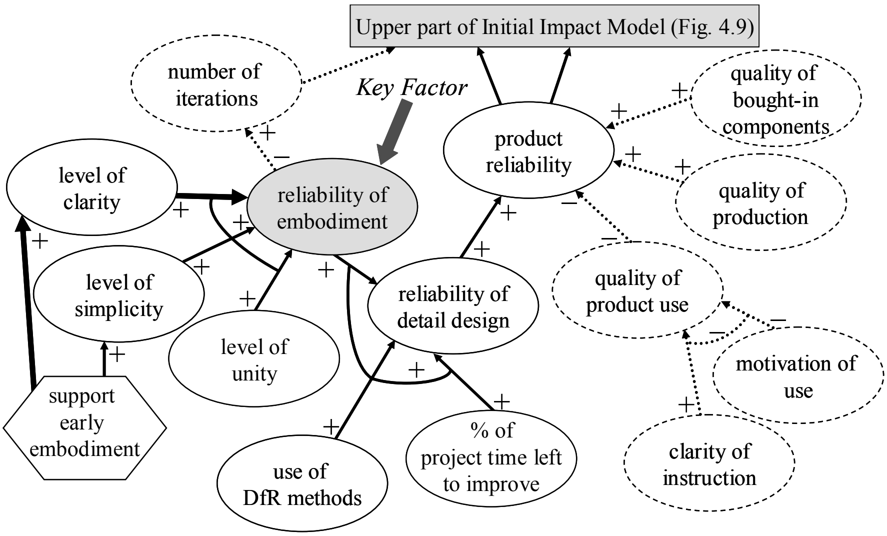

Figure 4.10 Lower part of the Initial Impact Model resulting from DS-I, focusing on improving the levels of clarity and simplicity, emphasising clarity as one of the main factors

4.8.3 Documenting Research

The documentation of the results of DS-I involves the compilation of the results of the separate empirical studies. Chapter 7 provides general guidelines on writing up. Some suggestions specific to this chapter are the following.

- Apart from the findings and conclusions, the methods of data collection, processing and analysis have to be documented in detail, *e.g.*, using the checklist in Table 4.1. as a guide.
- One's own viewpoint, assumptions and beliefs should be stated explicitly.
- Findings (the facts) should always be clearly separated from interpretations, either by using suitable terminology or through suitable formatting of the text.
- The findings should be presented as clearly and precisely as possible, without suggesting generalisations and significance beyond that which the method and findings allow.
- Where statistics have been used, the method, the significance levels and other data relevant to the specific method have to be mentioned along with the statements. For example, for a t-test result: 'the differences in activity sequence (t(46) = –2.08; $p < 0.05$) and the quality of the structure (t(46) = –2.49; $p < 0.05$) are significant', and for a Kendal τ-test result: 'An ex-post correlation of chronological age and prior experience with technical devices shows no significant relationship (Kendal $\tau = 0.13$, $p > 0.42$)'. Relevant the literature needs to be consulted for the correct formulation.
- It is useful to mark the significant findings in graphs and tables, for example as shown in Table 4.5, to help the reader to identify these more easily.
- Alternative explanations have to be given.
- The limitations of the research should be stated clearly, *e.g.*, by addressing the validity of the results.

Table 4.5 Example of the indication of significance (here the results of a regression analysis from Mahlke (2008))

Predictors	Overall product rating
Perceived usability	*0 .58****
Perceived aesthetics	*0.10*
Subjective feeling – valence	*0.30****
Subjective feeling – arousal	*0.09*
R^2	*62%*

*p< .05, **p< .01, ***p< .001

4.8.4 Consequences and Suggestions for the Intended Support

An important part of the conclusions is a description of how the increased understanding obtained through DS-I can help to improve the current situation, *i.e.*, the consequences for the development of support necessary to attain the desired situation. This will lead to suggestions for a means of support, which may consist of guidelines, a method, a computer tool, *etc.*, for designers, but also for other departments and stakeholders. Examples of the latter are new organisational structures, governmental regulations, improved information flows between customers and company, *etc*. These possibilities have to be considered, as the solution may not lie within the realm of the designer.

Reliability Example

In our reliability example, some of the conclusions are that the strength of components (unity) can be dealt with using existing methods, but that a method is needed to determine clarity and simplicity in an early stage. The suggestion is to base this method on the minimum amount of product data necessary because in the early stages, there is still only a description rather than a definition of the product.

4.8.5 Determining Next Stage

In general, the results of DS-I as a whole have to be used to determine the next stage. The possibilities are the following:

- the level of understanding is sufficient to suggest or develop realistic and effective types of design support ⇒ move to the PS stage;
- the level of understanding is still insufficient ⇒ carry out a Comprehensive DS-I to increase understanding;
- existing findings, models or theories seem incorrect or contradicting in the light of one's own findings ⇒ elaborate the literature review or carry out a comprehensive DS-I to verify these;
- existing design support seems ineffective, inefficient or is not used ⇒ undertake a PS (when the reasons for the identified problems are sufficiently well known to develop alternative support) or a Comprehensive DS-II (when it is unclear why existing support is not effective).

In our reliability example, the results of this Comprehensive DS-I stage are considered sufficient and it was decided that the next stage will be a PS stage to develop a reliability assessment method. The plan is further to evaluate the method in a DS-II stage, then implement necessary modifications in another PS stage and close the project with a final evaluation of the application of the improved method in a second DS-II stage.

4.8.6 Determining Future Work

The description of future work will discuss those questions and hypotheses that came up but that are not addressed, because they fall outside the scope of the project or because the current understanding seems sufficient to continue with the next stage. This, however, does not imply that full understanding has been achieved. From a practical point of view the understanding obtained may be sufficient to start developing the support, even though not all questions and hypotheses have been answered. Its development and in particular its evaluation will improve understanding and focus the possible questions and hypotheses that can or should be investigated further. The development of support can contribute to understanding in a way that a further empirical study, at least at the moment, cannot provide. In our example, the attempt to find a method that combines the assessment of the levels of clarity, unity and simplicity into one that represents the level of reliability may, *e.g.*, reveal the need for more details on the way in which experts assess clarity. Iterations between DS-I and PS will take place and are very useful, but should not result in a trial and error approach in the sense of 'let's just develop something and see if it works'.

4.9 Main Points

The main points of this chapter can be summarised as follows:

- The objectives of DS-I are: to obtain a better understanding of the existing situation by identifying and clarifying in more detail the factors that contribute to or are detrimental to the preliminary Criteria; to obtain a greater clarity of the expected situation by determining the factors that seem most suitable to address; to provide a basis for the effective development of support to improve design; and to provide detail that can be used for the evaluation of its effects.
- All design research types need a DS-I stage to complete the Reference Model. Depending on the research goal, DS-I will be limited to a Review-based Study involving a detailed review of the literature, or a Comprehensive study that includes a detailed literature review and an empirical study and takes place if the literature shows a lack of understanding of the topic.
- In this book the term 'Descriptive Study' refers to the particular stages of DRM in which all types of empirical study suitable for investigating design can be employed, including exploratory, descriptive and explanatory studies.
- When adopting research methods from other disciplines, the paradigms upon which the methods are based should be taken into account, as these can constrain their combination and application.
- Two main schools of thought exist: starting with a theory, developing hypotheses and testing these using empirical research (theory-driven); or using empirical research to develop theories and hypotheses (data-driven). In reality, neither occurs in their 'pure' form.
- Quantitative research is used to investigate the degree to which phenomena occur. Qualitative research is used to investigate the nature of phenomena. Their combination can obtain a richer picture of the phenomena.

- The steps of a Comprehensive DS-I process are: reviewing the literature; determining a research focus; developing research plan; undertaking an empirical study; drawing overall conclusions. The steps will involve much iteration.
- The literature review in DS-I, in particular of the literature on empirical studies, aims at updating the Initial Reference and Impact Models. Discussion with experts and stakeholders can be useful.
- Assessing an empirical study requires a detailed analysis of its publications. The checklist for reviewing descriptive studies can be used as an aid.
- In a proper empirical study, the aim, the research questions and/or hypotheses, the type of data collected, the way it is collected, processed and analysed, the interpretations and conclusions should all match.
- Not all factors and links identified in the Initial Reference and Impact Models can be investigated in detail, because of project-related constraints. Focusing is essential.
- To determine the research focus, factors and links of interest are identified and defined, research questions and hypotheses formulated and refined; and the final set chosen.
- While formulating and refining the set of questions and hypotheses, one should also consider: the research goal; possible effects of other factors on the phenomena; the methods used and setup of the study, project constraints beyond the researcher's control; the level of understanding that can be obtained from the literature.
- The analysis of questions and hypotheses may reveal assumptions, leading to further questions and hypotheses that require additional methods.
- We propose three techniques for refining research questions/hypotheses: Question and Hypothesis Analysis, Answer Analysis, and Question-Method Matrix Analysis.
- The concepts used in the research questions and hypotheses have to be given an operational definition to define 'what to do' to empirically establish the existence of a phenomenon described by the concepts. Validity tests are used to check whether a definition is suitable to measure a concept.
- For focusing and prioritising the set of research questions and hypotheses, it is useful to ask: What is the reason for including this question or hypothesis? How important is it? Do the questions and hypotheses relate to one another? Would the answers provide a coherent picture? What use can be made of the answer?
- A research plan of an empirical study defines: research goal and objectives for the study; research questions and hypotheses to be addressed; data-collection methods and setup; data-processing methods; data-analysis methods; data interpretation methods; and methods to validate the findings.
- Usually, data-collection methods are chosen first, but the other research methods should be considered simultaneously.
- Finding suitable research methods can start with reviewing the literature on studies in design, but consulting specialist literature on the research methods considered is essential.

- Real-time methods can produce more direct and rich descriptions of events and their context, but generally for few cases. Retrospective methods summarise events and use memory or documentation, but with the danger of post-rationalisation. These methods are suitable for large numbers of cases and when reflection is required.
- The Question-Method Matrix Analysis technique can be used to find a suitable set of data-collection methods. The detailed design, the combination of the methods, and the development of the necessary instrumentation determines their suitability for a study.
- Empirical studies focus on variables. Variables are characteristics of a situation or phenomena that can change in quantity or quality. Variables can be dependent (those the researcher wishes to explain), or independent (those that change the dependent variable). Control variables are those involved in alternative explanations of an observed relationship.
- To obtain valid data, two types of errors should be avoided: bias or systematic error, and error or random error. These can occur in all stages of research.
- The aim of a pilot study is to try out the whole research approach from data collection to drawing conclusions, to identify potential problems that may affect the quality and validity of the results, and to modify the approach as needed. Despite this, things can go wrong; planning for contingency is important.
- Analysing and interpreting the data requires the data to be processed, involving organising, abstracting or indexing the collected data using codes.
- Codes are categories, often derived from research questions, hypotheses, key concepts or themes, which can be pre-defined (deductive coding) or post-defined (inductive coding during analysis). The type of coding determines the analysis methods that are suitable and the possible results.
- Analysis and interpretation begins with simple enumeration or descriptive statistics and is followed by deeper analyses linking the findings, identifying correlations and possible causal relationships, finding explanations and drawing inferences. Inferences about causality require evidence of time order between concepts; covariance between concepts; and exclusion of rival factors (spurious relationships).
- The aim of data analysis is to draw valid inferences about the observation and to avoid spurious relationships.
- All possible evidence should be used, from data as well as the literature, to answer the questions and test the hypotheses. As many rival or alternative explanations as possible should be generated, taking different viewpoints.
- To choose the most likely explanations, different findings may have to be combined or further data collected. Not all plausible explanations can be verified in a single project.
- Verifying results involves judging the plausibility and credibility of evidence. Problems can occur due to circumstances beyond the control, or within the control of the researcher. Both influence the validity of the results.

- Two aspects are important: the reproducibility of measurement (reliability) and the degree to which the measurements actually reflect the true variation in the outcome (validity).
- Only if methods are carefully selected and correctly applied, and investigations carefully designed, is it possible to realise valid and reliable results.
- There are four, related types of validity: statistical conclusion validity (are the variables related?), internal validity (is the relationship causal?), construct validity (is the causal relationship valid for higher-level concepts?) and external validity (are the results person, setting or time-dependent?).
- A distinction should be made between statistical significance and relevance.
- Documenting results should be carried out during data collection, processing and analysis. The circumstances of the study, viewpoints and assumptions by the researcher should be made explicit. Findings should be separated from interpretation. The limitations of the research should be stated.
- The deliverables of DS-I are: a completed Reference Model, Key Factors, Success Criteria and Measurable Success Criteria, an updated Initial Impact Model, implications for support development and evaluation, *i.e.*, a description of how the understanding from DS-I can help improve the current situation.