Lucienne T.M. Blessing
Amaresh Chakrabarti

DRM, a Design Research Methodology

Springer

DRM, a Design Research Methodology

Lucienne T.M. Blessing • Amaresh Chakrabarti

DRM, a Design Research Methodology

Springer

Lucienne T.M. Blessing, PhD
Faculty of Science, Technology
and Communication
University of Luxembourg, Rectorat
Avenue de la Faïencerie
L-1511 Luxembourg
lucienne.blessing@uni.lu

Amaresh Chakrabarti, PhD
Centre for Product Design
and Manufacturing
Indian Institute of Science
560012 Bangalore
India
ac123@cpdm.iisc.ernet.in

ISBN 978-1-84882-586-4 e-ISBN 978-1-84882-587-1
DOI 10.1007/978-1-84882-587-1
Springer Dordrecht Heidelberg London New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009929232

Cover design: eStudioCalamar, Figueres/Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The initial motivator for the development of *DRM, a Design Research Methodology*, and the subsequent writing of this book was our frustration about the lack of a common terminology, benchmarked research methods, and above all, a common research methodology in design. A shared view of the goals and framework for doing design research was missing. Design is a multidisciplinary activity occurring in multiple application areas and involving multiple stakeholders. As a consequence, design research emerges in a variety of disciplines for a variety of applications with a variety of subjects. This makes it particularly difficult to review its literature, relate various pieces of work, find common ground, and validate and share results that are so essential for sustained progress in a research community. Above all, design research needs to be successful not only in an academic sense, but also in a practical sense. How could we help the community develop knowledge that is both academically and practically worthwhile?

Each of us had our individual ideas of how this situation could be improved. Lucienne Blessing, while finishing her thesis that involved studying and improving the design process, developed valuable insights about the importance and relationship of empirical studies in developing and evaluating these improvements. Amaresh Chakrabarti, while finishing his thesis on developing and evaluating computational tools for improving products, had developed valuable insights about integrating and improving the processes of building and evaluating tools. Many discussions took place with various researchers, in particular with Ken Wallace, who had particularly useful thoughts and insights based on his many years of supervising PhD students and involvement in the design research community. As background, several pieces of work were available: the extensive review of design research literature by Finger and Dixon (1989a; 1989b) categorising literature into descriptive, prescriptive and computational studies; the classical research approaches in natural and social sciences of creating and evaluating models and theories of reality; the approaches of research in disciplines such as economics and management studies, where observations are used to develop interventions to improve reality; and last but not least, the many theses and other publications describing many interesting approaches to tackle the challenges of doing design research.

As the literature showed, different methods can be, and have been, used to address the various issues involved in design research, and many areas of research have developed, focusing on various research questions. Based on our own experience, insights and analyses of these research questions, we aimed at putting the research areas together into one framework. The result is DRM, a generic design research methodology that links the research questions together and provides support to address these in a systematic way.

A preliminary version of DRM was developed as early as 1991 by us and Ken Wallace and published in Blessing *et al.* (1992). At that stage, however, only the major research questions and the DRM framework for addressing these questions were available, along with some examples of how to interpret and use this framework in research. An expanded version, with more examples, was published in 1995 (Blessing *et al.* 1995).

We started applying the framework for structuring the research of our students, which met with some success. However, it was clear that substantial further development had to take place to support each individual stage of the methodology. This was the precursor to a long period of joint research for over ten years. It involved creation, evaluation and improvement of various specific methods through our own research projects and those of our Masters and PhD students, the analysis of a large number of research projects in design, and the feedback from those outside our own research groups. DRM has been taught in a Summer School on Engineering Design Research in Europe since 1999 and as a Graduate level course in the Indian Institute of Science, Bangalore since 2002 (see Chapter 8 for some of the experiences). Several of the students we taught continue to use the methodology in their work. The feedback we received, and still receive, has been invaluable. Together with our own experiences, this has led to a clearer focus and greater substance and validity of DRM.

One of the consistent observations of our students is the lack of material for researchers on design research methodology: also on DRM. The papers and lecture notes we produced have been found helpful, but inadequate for understanding and using DRM in detail. The lack of detailed publications will have been a reason why some aspects and terms were misinterpreted, although the sources of some misinterpretations and even quotes are unclear: in particular the misinterpretation that DRM by emphasising Measurable Success Criteria would focus only on a quantitative approach to design research and devaluate qualitative methods; that the DRM process would be linear, narrowly focused on process aspects of design only; that DRM would only be applicable to individual research projects rather than research programmes. As the following paragraphs will explain, these interpretations are in direct contradiction with how we view design research, our own research projects, and what we have taught and written in our publications.

The adjective ‘Measurable’ in Measurable Success Criteria refers to the need to assess whether the criterion has been realised. The criterion as well as the methods used can be qualitative and quantitative. Design research in many instances needs a combination of qualitative and quantitative research in order to be able to answer the research questions. As we pointed out in Blessing *et al.* (1995), “methods from a variety of disciplines are needed for carrying out various aspects of design research" and our own research and that of our students show a clear combination of qualitative and quantitative research. In 2002 (Blessing and Chakrabarti 2002) we defined a Measurable Criterion as "the measure against which the results of the project will be judged". Note that we have now returned to using the full term 'Measurable Success Criteria' rather than the shortened version 'Measurable Criteria', as the latter caused some confusion. The emphasis in the book on qualitative, more inductive approaches to research is not because we consider these approaches more relevant, but because we assume the reader is more familiar with the quantitative research approaches and methods common in engineering.

DRM has never been intended to be a linear process, as should be clear from the DRM framework in which arrows link back to earlier stages of DRM. In comparison, the circular process models proposed as a better alternative tend to show a far stronger linear sequence of steps: returning to a stage can only be done in the next round. Similar to other 'linear' representations, our representation was chosen to emphasise the need to carry out the research in a systematic way connecting all stages. In particular, the representation is intended to indicate that one should: not start support development unless there is enough understanding and evidence that the need is real and no support currently exists; not evaluate support before carrying out adequate development that ensures that the support can indeed be evaluated for its goals; not only consider improving the support after its evaluation, but also reconsider the understanding upon which the support is based. In our papers, we always emphasised the non-linear, iterative nature of the research process and the fact that some stages may run in parallel.

Contrary to focusing only on process-related aspects of design, DRM is intended to address all facets of the phenomenon of design. As we wrote in Blessing and Chakrabarti (2002), "Design is a complex activity, involving artefacts, people, tools, processes, organisations and the environment in which this takes place. Design research aims at increasing our understanding of the phenomena of design in all its complexity". "Each of these facets is dealt with in specific disciplines [...]. Each discipline has its specific research methods and, equally important, underlying paradigms and assumptions". A design research methodology "should help in identifying research areas and projects, and in selecting suitable research methods to address the issues".

Finally, contrary to being applicable only to individual research projects, DRM is meant to be used both at project and programme levels, as emphasised in the design research types discussed in Chapter 2. "It cannot be expected that each of the stages of the methodology will be executed in depth in every single project [...] a research project may address only one stage because it is part of a larger project" (Blessing and Chakrabarti 2002). DRM, in fact, has already been used as the basic methodological structure of a product platform for computational design tool research for "more loosely related but still potentially complementary projects often steered by different investigators" (Bracewell *et al.* 2001). We have also found the DRM outline used in proposals for research programmes.

This book presents, for the first time, the DRM methodology and associated methods and guidelines in its entirety. Those who were involved in the process or read earlier publications will particularly notice the change in terminology. We decided to use the opportunity of writing the book to put all the feedback together and reassess the terminology as a whole. This has most clearly affected our overview figures, Figures 2.1 and 2.9, which display most of the key terms. For reasons of continuity we kept the terms used for the three main stages: Descriptive Study I, Prescriptive Study and Descriptive Study II, although we also discussed these extensively. We hope that the new terminology introduced in this book now more clearly expresses the underlying concepts and the messages we wish to convey. However, we realize that there is still much work to be done and continue to welcome suggestions for improvement.

The primary aim of our methodology and its related guidelines is to help engineering and industrial design research to become more relevant, effective and efficient. In addition, we believe that much of the content of the book should be useful for research in other design domains as well. This book is intended to be a practical handbook for teachers, students and researchers in design. The central objective is to help researchers and research groups to rigorously and efficiently plan, implement and communicate their research. This, we hope, should help make design research more creditable to the academic community at large as well as to product development practice and society where our contribution as a useful discipline counts most.

A large number of people have contributed to the development of our ideas and the writing of the book. First and foremost, we acknowledge the sustained encouragement from Ken Wallace, as an initial contributor to DRM, as the Director of and colleague at the Cambridge EDC where much of the theoretical development of DRM took place, and beyond our Cambridge days, as a keen friend and well-wisher who tried to ensure that we did not lose sight and hope in this lengthy exercise.

We also thank Mogens Myrup Andreasen for giving us the opportunity to join and extend the Summer School on Engineering Design Research that he initiated, so that DRM can be taught to, and feedback received from, PhD students across Europe. We are also thankful for his critical comments and encouragement in the many discussions that followed.

We thank our own research and PhD students and those we taught over the years at the Summer School on Engineering Design Research and the Methodology for Design Research course at the Indian Institute of Science, Bangalore, for trying out our framework and the various methods we proposed, as well as their ever so helpful criticism and suggestions for improvement. We specially thank Mattias Bergström, Åsa Ericson, Thomas Flanagan and Judith Jänsch who provided us with careful analyses of DRM compared to other approaches.

We, furthermore, would like to thank all contributors to the book, who have been so kind to provide a summary of their research or that of their PhD students, which we used as examples in Appendix C of this book: Eckart Frankenberger, Ade Mabogunje, Mogens Myrup Andreasen, David C. Brown, and Ken Wallace. In particular, we would like to thank them for their patience during the long and difficult pre-natal period of this book! We are also grateful to Springer, in particular Nicholas Pinfield, Oliver Jackson and Aislinn Bunning, who have been patient with us during long periods of lull during the writing of the book and supported us with the final editing.

A large number of colleagues and students have variously helped in the creation, development and dissemination of DRM. Our thanks go to all of them, with special thanks to: AV Gokula Vijaykumar and Srinivas Kota for helping us connect via Yugma™ to work together remotely during the last few months of writing the book, Ivan Yates for his continuous encouragement, and Steve Culley and John Clarkson for ordering the book well before it was published.

We are particularly grateful to Mogens Myrup Andreasen and Norbert Roozenburg for reading an early manuscript of this book and the many suggestions they made for its improvement. We are also thankful to Pavan Sridharan, Madhusudanan N. and Ujjwal Pal for reading through the copy-edited version of the manuscript and spotting some vital errors.

Last but not least, we acknowledge the contribution of our spouses – Peter Stomph and Anuradha Chakrabarti – for their continued patience and kind encouragement during this difficult period, and apologize to them and our children – Koen, Saskia and Apala – for the many hours that we were away from home to work on the book.

Enjoy your research.

Luxembourg *Lucienne Blessing*
Bangalore *Amaresh Chakrabarti*
December 2008

Contents

1

Introduction

The importance of design, in particular as an industrial activity and the increasingly complex and dynamic context in which it takes place, has led to the wish to improve the effectiveness and efficiency of design practice as well as education. In this chapter, we discuss the nature of design, review the current state of design research, and explain the need for a common methodology for design research.

1.1 Design

Many definitions of design exist, very much depending on the culture and background of the author. When we speak in this book about *design,* we refer to those activities that actually generate and develop a product from a need, product idea or technology to the full documentation needed to realise the product and to fulfil the perceived needs of the user and other stakeholders. The perceived need may be social (*e.g.*, transportation means) as much as economic (*e.g.*, manufacturing systems for mass production). The impulse to start such a process can come from: the market, such as needs of customers and competing products; internal needs of product development enterprises, such as new developments, cost reduction, production automation and diversification goals; and from other sources, such as research results, legislation, environment, society and politics.

The design process can be undertaken by individuals, industry, or a community, using various product development practices. The solutions resulting from these practices can be of both engineering and non-engineering nature, and may be physical (such as a coffee-maker or an aircraft), virtual (such as computer software a plan or a process) or a combination of both (*e.g.*, a mechatronic system or a product-service-system – a combination of an artefact and a process (service) – such as an aircraft leasing scheme or a mobile phone service). In this sense, the term product, as used in this book, signifies a much broader concept than conjured up by its commonly perceived image – an industrial, often mass-produced artefact created by industry. Design is an activity that has an effect on nearly every sphere of human life (Pahl and Beitz 2007).

In this book, our primary focus is on supporting engineering and industrial design research, although the content of the book, we believe, can also support research into other types of design in a wider variety of disciplines.

Design requires not only knowledge of the stakeholder goals and the product, but also about its life cycle, *i.e.*, how it is to be produced, transported, installed, used, maintained and retired, and about the process of design, *i.e.*, how to proceed in an effective and efficient way. This implies that in order to design, designers have to draw on knowledge from areas as diverse as physics, chemistry, mathematics, engineering sciences, economics, aesthetics, ergonomics, psychology and sociology, as well as on methods and tools to support the application of this knowledge. If knowledge is not available, which is often the case, designers have to rely on assumptions while minimising the risks, or undertake research to generate this knowledge.

Design is not only a knowledge-intensive activity, but also a purposeful, social and cognitive activity undertaken in a dynamic context, aimed "at changing existing situations into preferred ones" (Simon 1981) Design is a complex, multi-faceted phenomenon, involving: *people*, a developing *product*, a *process* involving a multitude of activities and procedures; a wide variety of *knowledge, tools and methods*; an *organisation*; as well as a *micro-economic* and *macro-economic* context. All of these facets have their own goals, structures and cultures that put demands and constraints on design, all of which have to be dealt with and balanced. Any of these facets might change during the course of a design project: needs become obsolete or change, competitors introduce new products, legislation becomes tighter, team members leave, and new technologies are introduced all the time. To complicate matters, "design is governed by natural laws as well as human purpose" (Simon 1981). Furthermore, every design project is by definition unique: the aim of a project is to create a product that does not exist yet (the uniqueness may relate to a particular detail as well as to the overall concept); the tools, methods, resources and context in which the project takes place will differ; and the knowledge and experience of the team members is growing all the time. The many facets, the dynamics and the uniqueness challenge not only those who plan and undertake design, but also those who plan and undertake research into design: how to investigate the strongly interrelated influences of all the mentioned facets; how to investigate such a dynamic process and its effects on the product and the process; how to support designers to enable them to better deal with these facets and dynamics; and how to cope as a researcher with the uniqueness of the investigated processes?

1.2 Design Research

Although people have undertaken and attempted to improve design processes for centuries, it was not until well into the second half of the 20th century that researchers became interested in *design* as a topic of research (see Pahl and Beitz (2007) for an overview and Heymann (2005) for a detailed historical account). *Engineering science* research, which includes research, *e.g.*, into thermodynamics, mechanics and materials, has a much longer tradition, as can be seen from the establishment of many technical universities in the second half of the 19th century. Key to the change of attention toward design research, was the increased complexity of design, its economic importance,[1] and the realisation that design is more than applying scientific findings; while “it (design) stands on scientific foundations, there is a big gap between scientific research and the engineering product, which has to be bridged by the art of the engineer..” (Gibbons and Johnson 1982). Gradually, design came to be regarded as a topic of scientific research with its own body of knowledge, related but not identical to other sciences (including engineering science).

Design research can be considered to have passed through three overlapping phases: the Experiential, Intellectual, and Experimental (Wallace and Blessing 2000). In the Experiential phase, which lasted up to the late 1950s, senior designers wrote about their experiences of the design process and the resulting products. However, their observations were not placed within any theoretical framework and were generally very specific to one technical domain. The Intellectual phase started during the 1960s and lasted about 20 years. During this phase a great deal of thought went into providing a logical and consistent basis for design, and many methodologies, principles and methods were proposed. In the Empirical phase, which slowly started in the 1980s and gained momentum in the 1990s, empirical studies were undertaken to gather data, both in the laboratory and in practice, in order to understand more fully how designers and design teams actually design, and what impact new methods and tools had on the design process.

Simon described the Science of Design in his book on the Science of the Artificial (Simon 1981) as the study of “the way in which the adaptation of means to environments is brought about”. This should result in “a body of intellectually tough, analytic, partly formalisable, partly empirical, teachable doctrines about the design process”. However, despite fifty years of design research, a theoretical framework for design is still missing, despite some notable attempts: Altschuller’s TRIZ (Altschuller 1984), Andreasen’s Domain Theory (Andreasen 1980), Braha’s Mathematical Theory of Design (Braha and Maimon 1998), GEMS of SAPPhIRE Model (Srinivasan and Chakrabarti 2008a; Srinivasan and Chakrabarti 2008b), Gero’s Function-Behaviour-Structure Framework (Gero 1990; Gero and Kannengiesser 2002), Hatchuel and Weil’s CK-Theory (Hatchuel and Weil 2003; Hatchuel and Weil 2009), Hubka’s and Eder’s Theory of Technical Systems (Hubka and Eder 1988), Lossack’s Domain-Independent Design Theory (Lossack 2006) based on Grabowski’s Universal Design Theory (Grabowski *et al.* 1998), Roozenburg’s and Eekels’ Logic of Design (Roozenburg and Eekels 1995), Smithers’ $K_L D_E$-Theory (Smithers 1998), Axiomatic Design (Suh 1998), Tomiyama’s Theory of Synthesis (Takeda *et al.* 1999; Tomiyama *et al.* 2002), Yoshikawa’s General Design Theory (Yoshikawa 1980) and Weber’s CPM-Theory (Weber 2005). The publication of Lossack (2006), provides an in-depth discussion

[1] Companies that develop their own products have enormous economic leverage, because they have far greater potential freedom to influence the added value of their operations than other companies, and so increase profitability and generate the wealth required to fuel growth (Yates 1998).

of several of the theories. Unfortunately, most of these theories and frameworks are not well known within the research community, while those that are known are not very widely used.

Discussions about what constitutes design research and how it is distinct from or similar to other disciplines are still very much ongoing, see, *e.g.*, Blessing *et al.* (1992); Blessing (2002); Cantamessa (2001); Frey and Dym (2006); Galle (2002); Galle (2006); Horváth (2001); Horváth (2004); Lossack (2006); Love (2000); Love (2002); Reich (1994a; 1995). The effects on design practice and education have so far been limited, but it is expected that "the most significant changes in design practice will occur when the field of design is fully endowed with a firm science base" (Suh 1998). However, the term 'design science' is still treated with scepticism. Some do not believe design to be a topic suitable for scientific investigations and point at the differences between design and science.[2] Others point to the fact that the term science is often used to refer to the natural sciences, thus leading to comparisons that fail to recognise the specific characteristics of design and its research.

At the moment, "it is no simple matter to define the contents, the research approach or the community behind research in engineering design" (Cantamessa 2001). The reasons Cantamessa mentions are: the relative youth of the discipline; the involvement of researchers of different disciplinary backgrounds; and the fact that there is no existing academic discipline of which it can be viewed as a natural offspring and from which research methods and tools could have been inherited. In addition, the complexity of the phenomenon of design, as outlined in Section 1.1, is a factor.

There is no common view as to what design research attempts to investigate, what its aims are and how it should be investigated. Many different methods are applied, many different aims pursued and many different aspects investigated. Some, such as Buchanan (2004) consider the existence of different views a strength; others are worried that this might prevent coherent theory development or, as expressed by Galle (2006) cause the Problem of Disintegration.

Many definitions of design research exist, often including the aim to improve design in practice. Initially, this practical aim was the main focus of design research, rather than the aim to better understand design. This resulted in an exceedingly large number of different means of support,[3] many of which remained at a conceptual level; few were implemented in practice. This focus is surprising, because the development of support that is intended to improve design is likely to

---

[2] A difference indeed exists between the activities of design and science, but this difference is not a valid argument against a science of design. Crime is not the same as science either, but this does not mean that criminology cannot be a field of scientific research.

[3] The term *support* is used to cover the possible means, aids and measures that can be used to improve design. This includes strategies, methodologies, procedures, methods, techniques, software tools, guidelines, information sources, *etc*., addressing one or more aspects of design. Support thus covers a spectrum as diverse as: checklists for identifying requirements, software for calculating stresses, drawing aids, guidelines for embodiment design, tools for product life-cycle assessment, project management tools, procedures for introducing methods, plans for new organisational structures, standards and regulations.

be far more efficient and effective if design is better understood. When, in the Experimental Phase mentioned earlier in this section, research started to focus on understanding design, this unfortunately happened rather independent of the research focused on improving design; the increased understanding was rarely used to inform the development of design support.

Our definition of design research integrates these two main strands of research: the development of *understanding* and the development of *support*. These strands are closely linked and should therefore be considered together to achieve the overall aim of design research: to make design more effective and efficient, in order to enable design practice to develop more successful[4] products. Accordingly, design research has two objectives:

- the formulation and validation of models and theories about the phenomenon of design with all its facets (people, product, knowledge/methods/tools, organisation, micro-economy and macro-economy, see Figure 1.1); and
- the development and validation of support founded on these models and theories, in order to improve design practice, including education, and its outcomes.

In the terms used by Horváth, design research is “generating knowledge about design and for design” (Horváth 2001).

Figure 1.1 illustrates the overall aim and objectives, and highlights the facets of design, discussed earlier in detail in Section 1.1.

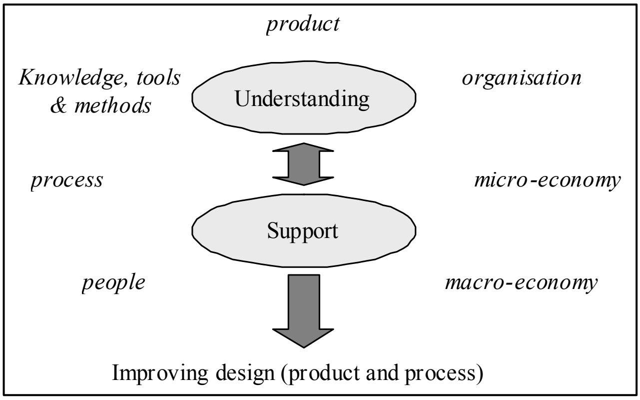


Figure 1.1 Design research: aim, objectives and facets of design

These facets are inherent parts of the phenomenon of design: they are to be studied or at least taken into account in design research. Typical for design research is that usually a combination of facets is studied. Hence, the research methods and

[4] The term *success* is used to mean a variety of achievements: including the achievement of the expected levels of quality, cost and time to market and other company goals, and the fulfilment of the perceived technical, aesthetic, ergonomic or societal needs.

paradigms of different disciplines have to be addressed. This view is shared by researchers, such as Fulcher and Hills (1996); Samuel and Lewis (2001), who argue for an all-embracing approach to design research, because the application and take-up in industry has had "undue emphasis", rather than research driven by a need to generate fundamental design knowledge. Schregenberger (1997) too pleads for a holistic approach.

It has to be noted that the term 'design research', and related terms such as 'design research method' and 'design research methodology' are also used in the context of design practice to refer to approaches and methods *for designers*, rather than researchers, to acquire knowledge for improving their work by doing research, *e.g.*, as part of the task-clarification stage of a particular design project.

1.3 Main Issues

The current status of design research gives rise to three related issues that need addressing (Blessing and Chakrabarti 2002):

- the lack of overview of existing research;
- the lack of use of results in practice;
- the lack of scientific rigour.

1.3.1 Lack of Overview of Existing Research

Design research has experienced an exponential growth. The number of researchers and research groups has grown rapidly, the number of papers published over the last 10 years has outnumbered that of all decades before, and an increasing variety of disciplines is involved. The downside of this 'wild growth' is that (Blessing 2002; Blessing 2003):

- Many loosely coupled strands of research have emerged.
- The strands are neither established nor clearly defined.
- It is no longer possible to obtain an overview of the results.
- Many 'referencing islands' exist: these are groups of researchers linked through a journal, conference, language or region, referring mainly to work published within the group.
- There is no agreed terminology, not even for basic terms such as 'function' and 'design'; the authors of the 390 papers published at a major engineering design conference used 1462 key terms, 1049 of which were unique, 880 used only once, and less than 100 used twice (Lowe *et al.* 2001).
- There is little contradiction among the findings of empirical research: not because findings are confirmed (little verification and validation takes place) but because all address something different and few attempts are made to bring results together.

Samuel and Lewis too, complain that design research is highly fragmented and focused streams of activity are lacking (Samuel and Lewis 2001), and Horváth

(2001) observes that design research "has grown to a significant complexity", for which reason "it is not easy to see the trends of evolution, to identify landmarks of development, to judge the scientific significance of the various approaches, and to decide on the target fields for investments".

To find some "orientation in the jungle" (Horváth 2001), there have been attempts to create an overview of design research, such as Finger and Dixon (1989a); Finger and Dixon (1989b); Horváth (2001). More and more researchers point out the need to go further and set up a comprehensive, archival database of research results to allow these to be compared and integrated, *e.g.*, Samuel and Lewis (2001); Blessing (2003). Such an overview could support the necessary identification of subsets of research (Blessing 1995).

The stage, in which we currently are, is often referred to as pre-theoretical, pre-paradigmatic (Cantamessa 2001) or pre-hypothesis (Horváth 2001). Bringing the results together is one of the pre-requisites for the development of comprehensive models and theories. The other important pre-requisite, namely scientific rigour, is discussed in Section 1.3.3.

1.3.2 Lack of Use of Results in Practice

The second issue that needs addressing in design research is the limited implementation and impact of research results: a situation discussed for many years in journals and at conferences, *e.g.*, Reich (1994a); Upton and Yates (2001), but up until now with little effect. Most results end up in scientific publications only and rarely in practice. If the aim of design research is to improve design, this research should have some effect on practice, directly or indirectly.

However, Cantamessa (2001) observed in his analysis of the papers of two large conferences in engineering design that the issues of implementation in industrial settings are only dealt with in 37% of the 331 papers on support. In 68% of cases, the relation of the tool to the current state-of-the-art in commercially available tools was not addressed. Publications of research in which industry had been involved, however, did usually address these issues.

Already in their 1996 report on research opportunities in engineering design (Shah and Hazelrigg 1996) the US National Science Foundation recommended that there should be greater and more effective interaction between industry and researchers in order to (1) assess industry needs for tools and technologies, (2) transfer research results into industrially usable methods; and (3) benchmark technologies in industrial settings. In their 2004 report (Shah *et al.* 2004) they reiterate the need for "major coordinated multi-disciplinary collaborative efforts including academia, industry and government agencies/labs".

The need for a better link between academia and industry was also found in a survey of UK industry (Upton and Yates 2001). According to this study, the most serious shortcomings of design research are that it "does not match industry's needs: in most cases the results of design research are not directly applicable and research is incorrectly focused" and that "there is a lack of mutual understanding between industry and academia". Furthermore, "engineering design researchers lack knowledge of industrial processes" and there is "low industry awareness of engineering design research activities". The authors propose a set of corrective actions, in particular: to improve working together; to increase dissemination of results; and to change the academic recognition and reward system.

These recommendations are useful and necessary, but not sufficient. The research approach itself has to be addressed as a cause of the observed shortcomings. Many guidelines, methods and tools have weak foundations: empirical data – as far as available – are hardly used to inform and drive support development; evaluation is poor; and implementation issues are rarely addressed. A more rigorous approach to research is required (see next section) in order to provide a sound foundation for the effective and efficient realisation of the recommendations proposed above and hence the successful implementation of research results into practice.

1.3.3 Lack of Scientific Rigour

The previous sections point at the third important issue that needs addressing: the observed lack of scientific rigour, in particular with respect to the application of research methods, the interpretation of findings, the development of support, and the validation and documentation of results.

The multi-faceted nature of design is one of the reasons for the diversity of research topics and methods. Diversity is not a problem, provided that an overview of research results exists and subsets of research have been identified (as discussed in Section 1.3.1). The problem is that "while variety has the potential of delivering value, this is not a certainty. If left to itself, there is a risk that research may end up in a set of unconnected streams and in a sort of methodological anarchy where anyone can come along and claim the scientific validity of his work" (Cantamessa 2001). This is exactly what seems to have happened in design research.

A reason might be that design researchers "are yet to properly grapple with the overwhelming complexity of the discipline" (Samuel and Lewis 2001), which requires a variety of methods to be applied, often from disciplines unfamiliar to design researchers, such as cognitive psychology. The consequence is that the literature contains many examples of projects in which research methods from other areas seem to have been chosen without knowledge of the underlying paradigms, that is, without careful consideration of their suitability for the project. The unfamiliarity with many of the methods also leads to incorrect use, resulting – unknowingly – in biased and useless data.

Literature on 'how to do research' focuses on other disciplines and provides little help. We have clearly noticed how ill-fitted these books are for design research, which has – as most disciplines do – its own set of characteristics and constraints that only partly overlap with other disciplines. We ourselves struggled with this issue and observe the same with the PhD students we encounter and supervise. Illustrative is the outcome of several discussions on design research with groups of PhD students, who had been involved in design research for one to two years: they were not clear about what constitutes design research and how to go about doing it (Blessing and Andreasen 2005). In our opinion, this at least partially explains the lack of methodological rigour that can be observed. Even though many research projects are successful, this is often at the expense of an inefficient research process.

1.4 Need for a Design Research Methodology

The prime emphasis of this book is on the third issue: to help achieve more rigour in design research. We believe that improving the status of research in this respect will also have a significant impact on both the other issues, *i.e.*, the need for an overview of design research and the transfer of results into practice. If we make explicit the questions that we as a community want to address, and make explicit the methods that are appropriate for answering these questions, we will be better able to evaluate our research and advance the field of design and design research.

Design research must be scientific in order for the results to have validity in some generic, theoretical as well as practical sense. For this, design research has to develop and validate knowledge systematically. This requires a research methodology: "It is the methodology that makes a topic of investigation scientific" (Frankfort-Nachmias and Nachmias 1996).

A methodology for design research should guide the selection and application of a suitable approach and appropriate methods, and encourage reflection on the approach and methods to be used. The call for a design research methodology can be found in several publications, *e.g.*, Cross *et al.* (1991; Fulcher and Hills (1996); Reich (1995). Sadly, the status of design research into its own research methodology is, with a few exceptions, poor (Reich 1995).

In this book, we propose a design research methodology called **DRM** to support a more rigorous approach in order for design research to become more effective and efficient. A *design research methodology* is defined here as an approach and a set of supporting methods and guidelines to be used as a framework for doing design research. The *overall objectives* are those mentioned for design research in Section 1.2: the formulation and validation of models and theories about the phenomenon of design, as well as the development and validation of support founded on these models and theories, in order to improve design practice, management, education and their outcomes. The relationship between design research methodology, design research and design is illustrated in Figure 1.2. This raises the question as to what type of methodology is necessary, given the typical characteristics of design research.

One distinctive feature of design research is the complexity of the phenomenon of design. Each of the facets shown in Figure 1.1 is the focus of a particular discipline with its own research methodology and methods, such as engineering, sociology, psychology, computer science, philosophy, history, management, and economics. Individual research methods from the various disciplines can be used – and should be used – depending on the specific research questions and hypotheses. A design research methodology should allow and support this.

Another distinctive feature of design research is reflected in the two objectives. Design research not only aims at understanding, but also at improving design. This requires (1) a model or theory of the existing situation, (2) a vision (model or theory) of the desired situation, and (3) a vision of the support that is likely to change the existing situation into the desired situation, and maintain this. That is, *design research itself involves design*, namely the creation and evaluation of a model or theory of the desired situation and of the support.

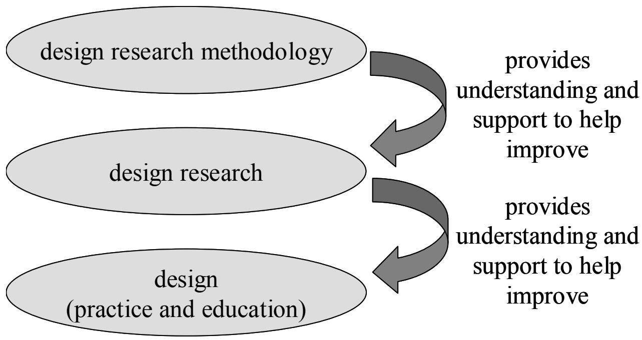


Figure 1.2 Relationships between design, design research and design research methodology

In most cases, the desired situation cannot be extrapolated from the existing situation; neither can the required support be derived directly from the difference between the two. The existing situation represents an undesired, possibly negative situation (although positive aspects may be known too), and it does not include the support with all its effects and side-effects. The model or theory of the desired situation, as well as of the support, is therefore mainly based on assumptions. The desired situation describes the 'ought' rather than the 'is'. As Roozenburg and Eekels (1995) write: "In philosophy the 'is-ought' transition is a source of controversy, as in a formal logical sense something is wrong". A design research methodology should explicitly support these activities.

Unfortunately, the objective of developing support, *i.e.*, improvement rather than explanation and prediction, is largely absent, ignored, or given little emphasis in existing research methodologies, even in those disciplines that have improvement as one of their explicit aims, such as management sciences, business administration and public administration (see also Reich (1994b); Reich (1995)). The few methodologies that do include, what they call intervention, as a step to change a situation once understanding has been gained, do not support the development of this intervention; the view seems to be that the most suitable intervention can be directly derived from the model or theory of the existing situation. An expectation is Action Research, discussed in Section 2.7.

Finally, the specific features of design as object of investigation, are a major challenge for evaluation in design research: "how does one validate design research in general, and design methods in particular, given that many proposed designs will never be realised and that it is often infeasible to follow the realised designs through their complete life cycles" (Seepersad *et al.* 2006).

1.5 Objectives

This book offers design researchers an overall framework and a stepwise, hands-on approach to design research. The objectives of the book are:

- to define design research by presenting our view on the various issues involved and how they link together;

- to present DRM, a design research methodology that:
    - links the various types of design research together;
    - helps plan design research by providing process steps; and
    - improves the chances of obtaining more valid and useful outcomes;
- to propose a set of methods and guidelines to help carry out design research;
- to provide pointers to existing approaches, methods and guidelines that are useful for design research;
- to help document and disseminate research results effectively.

This, we hope, should make researchers better prepared and equipped for research into design, making design research more rigorous, effective and efficient and its outcomes academically and practically more worthwhile.

1.6 Structure of this Book

Chapter 1 starts with an introduction to design and design research, and highlights the need for a research methodology.

Chapter 2 introduces the proposed design research methodology DRM with an example, describes the objectives of each of its stages, and presents how different types of design research fit within the framework provided by DRM. This chapter also introduces the main concepts that are used.

Chapter 3 discusses the first stage of the methodology: Research Clarification. Guidelines are given for clarifying the current state-of-the-art, and for determining the research goals, the type of research and the overall research plan.

Chapters 4 to 6 provide detailed discussions, methods and guidelines on how to plan, carry out and document research related to the three main stages of the methodology. Several examples from real projects are included to illustrate specific issues and solutions.

Chapter 7 provides guidelines on writing research publications.

Chapter 8 summarises the main contributions of the book and presents some of the experiences of users of DRM. The chapter concludes with a discussion of the outstanding issues in design research and research methodology and the need for further work.

Appendix A presents some of the most commonly used methods for doing empirical design studies. Appendix B presents various methodologies and methods for support development. Appendix C contains seven case studies of PhD projects written by the researcher or his/her supervisor. These case studies are intended to demonstrate the wide variety of problems addressed and approaches used in design research, and to allow the reader an insider's view of how design research is carried out. We have classified these projects using DRM to demonstrate its use as a framework for positioning a wide variety of design research projects.

Two examples are used through the chapters to illustrate the methodology. The *reliability example* is used to introduce the methodology. The description of its research process is kept relatively linear for the sake of clarity; in reality the process involved many more iterations. The *synthesis example* is used in the later chapters after many of the concepts of the methodology are explained, to illustrate the iterations between stages of the methodology. The examples are inspired by the work described in Appendices C.8 and C.3, respectively.

1.7 Main Points

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

- Design is the process through which one identifies a need, and develops a solution – a product – to fulfil the need. Design affects nearly every sphere of human life.
- Design is not applied science – it applies knowledge from engineering, natural, human and cultural sciences, and, if this knowledge is not available, makes assumptions to minimise risk or takes up research.
- Design is a dynamic, complex, multi-faceted phenomenon, involving people, processes, knowledge, methods and tools within an organisational, micro-economic and macro-economic context. Each design is, in some sense, unique.
- The importance and complexity of design are major motivations for improving the effectiveness and efficiency of design practice, management and education.
- Design as a topic of research with its own body of knowledge has been quite recent compared to other topics. Currently, there is no common view on the aims, objectives and methodology for design research.
- The overall aim of design research: to make design more effective and efficient, in order to enable design practice to develop more successful products.
- Design research has two, related objectives: the formulation and validation of models and theories about the phenomenon of design, and the development and validation of support founded on these models and theories, in order to improve design practice, including education, and its outcomes.
- The current status of design research lacks: an overview of existing research; use of its results in practice; and in its scientific rigour. The prime emphasis of this book is on the third issue.
- A design research methodology is an approach and a set of supporting methods and guidelines to be used as a framework for doing design research.
- The status of research into design research methodology is relatively poor, and little guidance exists as to how to do design research. Existing methodologies do not support the specific features of design research.
- The proposed design research methodology (DRM) and its methods are intended to support a more rigorous research approach by helping to plan and implement design research. The methodology, used flexibly, should help make design research more effective and efficient.

2

DRM: A Design Research Methodology

This chapter presents the outline of our methodology and introduces the main stages and concepts. At the end of the chapter, a comparison is made with the few other methodologies that have a similar purpose.

2.1 Introduction

The previous chapter described the two overall objectives of design research as formulating and validating models and theories about the phenomenon of design as well as developing and validating knowledge, methods and tools founded on these models and theories with the aim to improve design, that is, to improve the chances of producing a successful product. This raises a number of important questions:

- What do we mean by a *successful* product?
- How is a successful (or unsuccessful) product *created*?
- How do we *improve* the chances of being successful?

The first question leads to issues such as what the *goals* are and, derived from these goals, what *criteria* should be used to judge success, as these can be used to determine whether our research has been successful. The second question leads to issues such as the identification of the *influences* on success, how these influences interact, and how they can be assessed. Investigating these issues will increase our understanding of design, which is needed to improve it. The third question gives rise to issues related to how this understanding can be used to develop *support* and how this can be evaluated. Evaluation is needed to determine whether the *application* of the proposed support indeed leads to more success as determined by the criteria, *i.e.*, whether our goals have been achieved. Our research methodology intends to address these issues in an integrated and systematic way.

While a methodology should help realise a better planned and smoother research process, thereby increasing the *chances* of obtaining valid and useful results, such outcomes cannot be guaranteed: the nature of a methodology is heuristic, rather than algorithmic. Each researcher has his or her personal background and interests, making each research process unique. A methodology can only *support* this process. The outcome may be better and the topic may be more evenly researched, more rigorous and more reliable, but of course a good solution can be achieved without a methodology (usually at some cost) and a poor result can still be achieved when a methodology is applied (*e.g.*, because of a lack of specialist knowledge in the field of study or of a lack of reflection).

A methodology should be used in a flexible and opportunistic way to be able to adapt to the specifics of the research topic and any interesting avenues that may emerge (see also Section 3.8: general guidelines on doing research).

As stated in Chapter 1, the *aim* of DRM is to help design research become more effective and efficient. The *specific objectives* of DRM are:

- to provide a framework for design research for individual researchers as well as teams;
- to help identify research areas, projects and programmes that are most likely to be academically and practically worthwhile and realistic;
- to allow a variety of research approaches and methods;
- to provide guidelines for systematic planning of research;
- to provide guidelines for more rigorous research;
- to help develop a solid line of argumentation;
- to provide new methods and pointers to existing methods to carry out the stages of the research process;
- to help select suitable methods and combinations of methods;
- to provide a context for positioning research projects and programmes relative to other design research;
- to encourage reflection on the applied approach.

2.2 Methodological Framework

DRM consists of four stages: Research Clarification, DS I, Prescriptive Study (PS) and Descriptive Study II (Blessing *et al.* 1992; Blessing *et al.* 1995). Figure 2.1 shows the links between these stages, the basic means used in each stage and the main outcomes. The bold arrows between the stages illustrate the main process flow, the light arrows the many iterations.

A simple example is used to describe the framework. In the example, the stages are executed in a linear fashion for reasons of clarity. The example is followed by a discussion of the many variations of research that are possible within this framework. Section 2.6 provides a description of the objectives and main concepts of each stage. Details on how to execute each stage and suggestions for methods that can be used can be found in Chapters 4 to 6.

Example

Imagine a research project that starts with the aim of improving the way in which the early stages of the design process are executed, in particular the task-clarification stage. The underlying assumptions of the researchers (partly based on their understanding of design and partly on beliefs) are the following: task clarification is a crucial activity; improving the quality of task clarification will improve the design process; this in turn will result in a better and thus more successful product. Furthermore, they consider the currently available design support ineffective. The researchers decide not to immediately concentrate on their initial idea – the development of a requirement management tool – but to apply a systematic research approach, following the DRM framework shown in Figure 2.1.

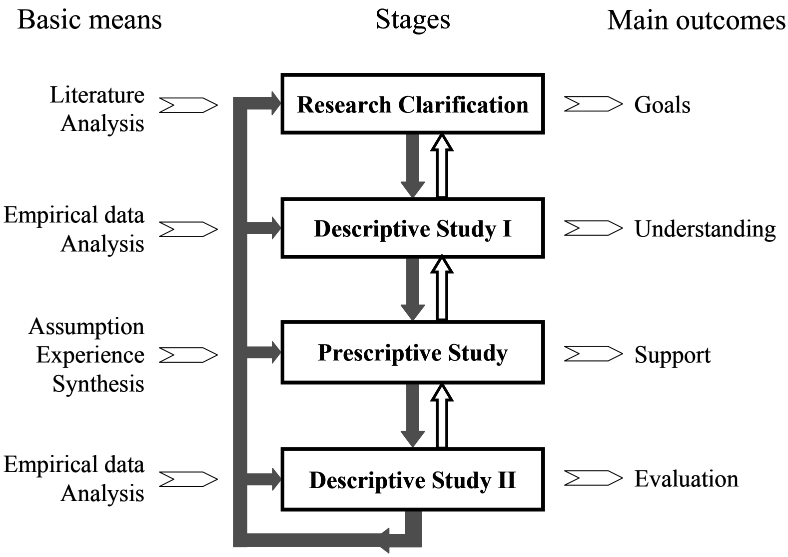


Figure 2.1 DRM framework[5]

In the **Research Clarification (RC)** stage the researchers try to find some evidence or at least indications that support their assumptions in order to formulate a realistic and worthwhile research goal. They do so mainly by searching the literature for factors that influence task clarification and product success, in particular those factors that link the two together. Based on the findings, an initial description of the existing situation is developed, as well as a description of the desired situation, in order to make the assumptions underlying each of the descriptions explicit. The researchers continue to formulate some criteria that could be used as measures against which the outcome of the research, *i.e.*, the support for task clarification, could be evaluated. It becomes clear that criteria for product success, such as ‘increase in profit’, cannot be used as a measure given the timeframe of the research project, but that ‘reduction in time-to-market’ could be a possible useful proxy.

In the **Descriptive Study I (DS-I)** stage, the researchers, now having a clear goal and focus, review the literature for more influencing factors to elaborate the

[5] Note that the terminology has changed since the inception of this framework in 1992 (see Preface).

initial description of the existing situation. The intention is to make the description detailed enough to determine which factor(s) should be addressed to improve task clarification as effectively and efficiently as possible. However, they do not find enough evidence in literature to clearly determine these crucial factors, and decide to observe and interview designers at work to obtain a better understanding of the existing situation, before moving on to the next stage and start developing support to address these factors. The analysis of the empirical data reveals the typical characteristics of insufficient problem definition and shows that insufficient problem definition in the task-clarification stage is related to a high percentage of time spent on modifications in later stages of the process. No evidence is found that more time spent on modifications increases time-to-market, but logical reasoning supported by other findings in the literature suggests that this is a plausible assumption. They decide that their understanding, reflected in the description of the existing situation, is sufficient for them to proceed to the Prescriptive Study stage.

In the **Prescriptive Study (PS)** stage, the researchers use their increased understanding of the existing situation to correct and elaborate on their initial description of the desired situation. This description represents their vision on how addressing one or more factors in the existing situation would lead to the realisation of the desired, improved situation. They develop various possible scenarios by varying the targeted factor(s). The researchers decide to focus on improving the quality of the problem definition as the most promising factor to address. Their argument is that this should reduce the number of modifications, which in turn should reduce design time, which eventually should shorten time-to-market and increase product success through increased profit. They now have enough confidence to start the systematic development of a support to improve the quality of problem definition. They use their understanding of the various interconnected influencing factors obtained in the DS-I stage; the well-developed description of the desired situation; as well as their experience in developing design support. To help them develop the support in a systematic way, the researchers choose to follow a design methodology. After a task clarification and conceptual design stage, they have the concept of a software tool (the intended support) that is expected to encourage and support problem definition as intended. They decide to focus their realisation efforts on the core of this support, as this should be sufficient to be able to evaluate the concept and verify the underlying assumptions. A first evaluation of this actual support shows that it has been developed correctly. Whether the support has the desired effects, however, is not clear yet, because of the many assumptions upon which the description of the desired situation and the development of the support have been based.

The researchers proceed to the **Descriptive Study II (DS-II)** stage to investigate the impact of the support and its ability to realise the desired situation. They undertake two empirical studies to gain an understanding of the actual use of the support. The first study is used to evaluate the applicability of the support. The main question is whether the software can be used to encourage and support high-quality problem definition. The second study is used to evaluate the usefulness, *i.e.*, success of the software, based on the criteria developed earlier. The main questions are whether less time was spent on modifications, and whether this eventually reduced time-to-market. The studies show that the support is applicable, but that the usefulness is less than expected. The researchers find that this is partly caused by the fact that the support actually developed includes only part of the support intended. They observe several effects they had not anticipated, such as the large amount of time needed to keep the problem definition up-to-date. The researchers conclude that their concept is promising, but that further investigations of the existing situation are needed and that the picture of the desired situation needs to be adapted accordingly before the tool can be improved and recommend a revisit of the DS-I stage.

Iterations and Variations

As we indicated at the beginning of this section, the example does not show the many iterations and the parallel execution of stages that are part of reality. Neither does it show that the starting point can be in any of the stages, and that it is possible, in an individual project, to concentrate on one or two stages only. The example is simplified and only intended to clarify the main flow of the process.

DRM is not to be interpreted as a set of stages and supporting methods to be executed rigidly and linearly. The negative effects of doing so are well known from the application of design methodologies. Fricke, *e.g*., observed that designers who tried to follow a design methodology step by step in a rigid fashion, produced designs of a lesser quality than those following a goal-directed but flexible approach (Fricke 1993a). The design process and the application of its methods are to a certain extent opportunistic (Bender 2004) and have to be adapted to the situation at hand (Zanker 1999). Iterations take place to increase understanding, as well as when understanding has increased (Chakrabarti *et al.* 2004) and stages are executed in parallel for a more efficient process (known as Concurrent or Simultaneous Engineering).

The same is true for the research process. As discussed in Antonsson (1987); Reich (1995), science does not often proceed in the linear, logical fashion suggested by its methodologies, although reports often suggest this. Iterations are commonplace within each stage. The results of an empirical study in the DS-I stage may reveal the need for further, erstwhile unplanned, studies, each enriched by the knowledge gained in the previous studies. Iterations are also common between stages. In the RC stage, it might be necessary to carry out some exploratory study (DS-I) to clarify the research goals and to develop a research plan, when little is known about the phenomenon of interest. While developing support (PS stage) an additional DS-I might be necessary to obtain more information about certain aspects of the context in which the support is to be implemented. And the results of the DS-II stage will usually warrant a revisit of one of the earlier stages.

To avoid too many unexpected iterations between stages, it is useful to plan stages to be partly executed in parallel. For example, it is necessary to start planning the evaluation of a support (DS-II) during and not after the development of this support (PS) in order to be able to determine which parts of the support need to be realised in order to do the desired evaluation. An example of parallel execution of stages can be found in Bracewell and Shea (2001) shown in Figure 5.14. The number and extent of iterations and the degree to which stages are run in parallel depend on the focus and constraints of a particular research project or programme.

2.3 Types of Research Within the DRM Framework

DRM as presented in this chapter is essentially comprehensive. As discussed earlier, it is not assumed, however, that a specific research project will necessarily include each stage, or undertake each stage in equal depth. In some cases, the literature provides sufficient material for a particular stage; in other cases, a research project may focus on only one stage for an in-depth study, because of time restrictions or because the project is part of a larger programme.

Figure 2.2 lists what we believe are the seven possible types of design research based on whether the state-of-the-art with respect to a particular stage requires a comprehensive study or whether a **review-based study** is sufficient. The research questions and hypotheses, and the available time and resources will determine the type of research to be undertaken. A review-based study is based only on the review of the literature. A **comprehensive study** includes a literature review, as well as a study in which the results are produced by the researcher, *i.e.*, the researcher undertakes an empirical study, develops support, or evaluates support. An **initial study** closes a project and involves the first few steps of a particular stage to show the consequences of the results and prepare the results for use by others. Each of the seven types will be discussed in more detail in Section 3.5.

Research Clarification	Descriptive Study I	Prescriptive Study	Descriptive Study II
1. Review-based	Comprehensive		
2. Review-based	Comprehensive	Initial	
3. Review-based	Review-based	Comprehensive	Initial
4. Review-based	Review-based	Review-based Initial/ Comprehensive	Comprehensive
5. Review-based	Comprehensive	Comprehensive	Initial
6. Review-based	Review-based	Comprehensive	Comprehensive
7. Review-based	Comprehensive	Comprehensive	Comprehensive

Figure 2.2 Types of design research projects and their main focus. (Iterations omitted)

The following assumptions were behind the selection of these seven types of research:

- Each project should start with a clarification of the research (RC stage) by reviewing the literature, to determine the aim, focus and scope of the research project.

- Any Comprehensive DS-I should be followed by an Initial PS to at least suggest how the findings could be used to improve design. An exception is Type I, in which the focus of the DS-I is on identifying criteria of success that can be used in design research. This type of research is followed by any of the other types of research.
- The comprehensive development of support (Comprehensive PS) should at least be based on a review of descriptive literature (Review-based DS-I), and be followed by an Initial DS-II to evaluate the resulting support. Many research projects we have seen end with the realisation rather than an evaluation of the support.
- A Comprehensive DS-II (evaluation) should be based on a Comprehensive PS or a Review-based PS to identify the background of the support to be evaluated, and at least be followed by an indication of how the support is to be improved (Initial PS).
- Each project should take into account all the stages of DRM, *i.e.*, past and future research have to be considered. Research projects and programmes are always contributions to a larger knowledge domain. It is therefore important that existing knowledge (results of past research) is referred to and used where appropriate, and that the results of one's own research are prepared in a way that allows others to use the gained knowledge in future projects.

The first four types of research project in Figure 2.2 focus on one particular stage, and are very suitable for PhD projects, although we have seen very few projects of Types 1 and 4. Types 5 and 6 cover two stages in-depth. The initial plans of PhD projects often aim for these types of research, but, as we observed, the time and resources required are often underestimated and the projects mostly end as Types 2 and 3, respectively. Type 7 requires three stages to be undertaken in-depth. This is more common for the work of a research group or when a problem with a very specific scope is addressed.

Our example represented a research project of Type 7. Had the literature review provided enough evidence to support the assumptions of the researchers and sufficient understanding to directly focus on the development of the support (PS stage), there would have been no need to do an empirical study in DS-I. This would then have been a project of Type 6. Had the researchers known that existing support was ineffective, but not known the exact problems, they could have decided to focus on a systematic evaluation of the use and usefulness of existing support (DS-II) and the development of suggestions for improvement (PS). This would result in a research project of Type 4. Other considerations, such as time constraints, would have resulted in other types of research.

2.4 Representing Existing and Desired Situations

As illustrated in the example, descriptions of the existing and the desired situation play a central role in DRM. We propose the use of – what we call – **networks of influencing factors** to describe the situations. We distinguish two types of networks of influencing factors, to describe the two situations relevant for DRM. The **Reference Model** represents the existing situation in design and is the reference – hence its name – against which the intended improvements are benchmarked. The **Impact Model** represents the desired situation and shows the assumed impact of the support to be developed. The models developed in the RC stage (see our example) describe the initial image of these situations and hence are called **Initial Reference Model** and **Initial Impact Model** (see Chapter 3 for details of developing these Initial Models). A full Reference Model is developed in the DS-I stage (see Chapter 4 for details) and a full Impact Model in the PS stage (see Chapter 5 for details).

A *model* is a likeness of something that exists in reality, but restricted to some particular aspects of this reality. Which aspects are represented depends on the purpose of the model, *i.e.*, on its intended use. Models are used in science to provide conceptual organisation. They show the significant relationships between the concepts or attributes, and thus highlight the aspects that are the focus of the research. Models are not theories, but they can be used to represent a theory.

2.4.1 Graphical Representation

This section summarises the main characteristics of the graphical representation we developed to present these models, using a Reference Model (see Figure 2.3) developed for the example discussed earlier. This Reference Model represents the level of understanding of the existing situation the researchers had at the end of their DS-I stage.

Figure 2.3 has to be interpreted as follows. The nodes represent influencing factors. An **influencing factor** (or **factor** for short) is an aspect of the existing situation (or the desired situation in the case of an Impact Model) that influences other aspects of this situation, *e.g.*, ‘the quality of the product’ or ‘the satisfaction of the customers’ influence ‘market share’. Influencing factors can cover all of the facets of design shown in Figure 1.1 and can come from the literature or other sources, such as assumptions, experience, research goals, focus, questions and hypotheses. A particular situation is represented by the factors that influence this situation and the links between these factors.

An influencing factor is formulated as an **attribute** of an **element** that is considered relevant and that can be observed, measured or assessed, *i.e.*, for which a so-called operational definition can be formulated (see Section 4.5.2). An example is ‘quality *(attribute)* of problem definition *(element)*’, see Figure 2.4.

The addition of the attribute is essential. ‘Problem definition’, *e.g.*, cannot be an influencing factor as it only describes the element. This introduces ambiguity: the researcher could mean ‘time spent on problem definition’, ‘quality of the problem definition’, ‘knowledge about the source of the problem definition’, *etc*. Each of these would be linked differently in the network: ‘time spent on problem definition’ influences the ‘overall design time’; ‘quality of problem definition’ influences ‘reliability of the product’; *etc*. The attribute thus determines the link to other factors and hence has to be made explicit.

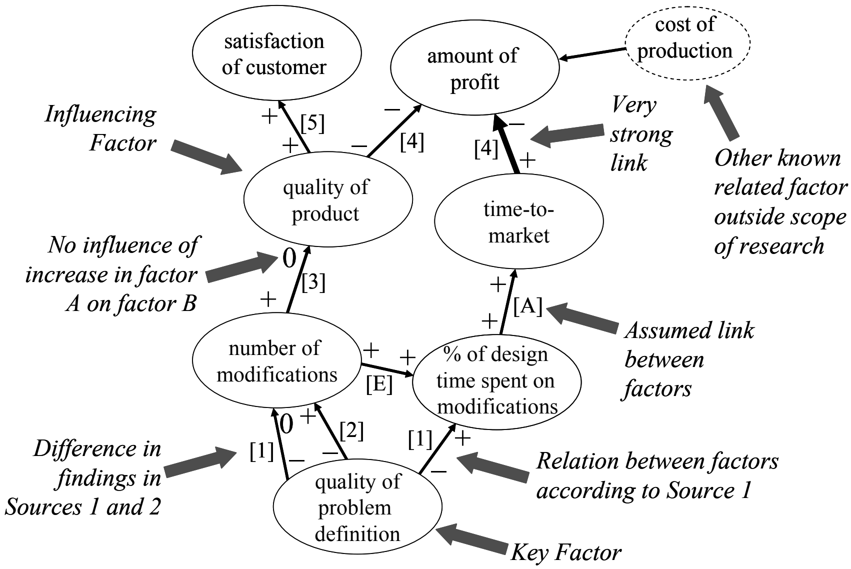


Figure 2.3 A Reference Model representing the – partly assumed – existing situation

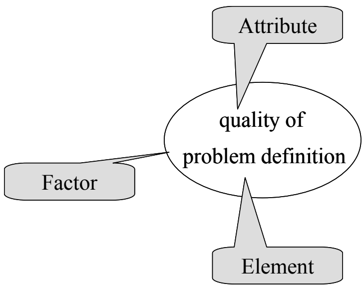


Figure 2.4 Factor, attribute and element

Key Factors are those influencing factors that seem to be the most useful factors to address in order to improve an existing situation. These are considered the core factors or the root causes. The Key Factors are addressed directly by the support. In our example, the researchers decided on the basis of their investigations that addressing the factor 'quality of problem definition' would be the most promising approach to improve 'product quality'.

Attributes have **values**:

- Values can be qualitative, such as 'high' and 'poor', or quantitative, *e.g.*, '20', 'larger than 20', 'between 10 and 30', depending on the definition used in a particular study.
- The existing or desired value of an attribute is attached to a link between factors by means of a '+', '–', or '0' sign. For example, a '–' next to the link and near the factor 'amount of profit' indicates 'low' profit or, depending on the definition, profit 'less than 20%'.

- Values should not be included in the description of the factor in the node. Thus ‘poor problem definition’ is not a correct formulation for a factor. The first reason is that including the value (poor) rather than the attribute (quality) introduces ambiguity: it is not clear whether poor refers to poorly written, poor contents, *etc*. The second reason is that including the value in the node does not allow multiple, differing statements to be represented using the same node. An example are the statements related to ‘quality of product’ in Figure 2.3; using ‘poor quality of product’ as factor, based on the link to ‘amount of profit’ (statement [4]) would not have allowed the statement labelled [5] to be represented as this refers to ‘high product quality’.

The **links** between factors show how the factors influence or are desired to influences each other, *i.e.*, they represent explicit statements about the existing or desired situation.

- The combination of ‘+’, ‘−’ and ‘0’ signs at the ends of a link describe how the value of the attribute of the factor at one end relates to the value of the attribute of the factor at the other end. Figure 2.3 represents, *e.g.*, that a poor quality of problem definition (−) relates to high percentages of time spent on modifications (+), and that a large number of modifications (+) was found to have no effect on the quality of the product (0).
- If the link is known or assumed to be a causal link, this is indicated with an arrow (→) from cause to effect.
- If a link exists between three or more nodes, *e.g.*, two factors *together* affect another factor, the links near the affected factor are connected and a single value is placed near the connection, see Figure 3.6 for an example.
- If certain factors are known to influence a factor in the network, but are themselves outside the scope of the research project, these factors are drawn differently, as illustrated by the factor ‘cost of production’ in Figure 2.3. Acknowledging the effects of such factors indicates awareness of the researcher of other possible influencing factors and supports the search for alternative explanations for research findings.
- Statements that are found in the literature cannot simply be reversed: if high costs lead to reduced sales, this does not imply that low costs lead to high sales. The latter would be an assumption. It is important to base the Reference Model on the original statements, even if this implies a non-continuous line of argumentation (as, *e.g.*, shown by the in- and outgoing links of the factor ‘quality of product’ in Figure 2.3). The Reference Model represents the current understanding as-is. Assumptions can be added that differ from the original statements, as long as they are labelled as such. The Impact Model, in describing the desired situation, provides the freedom to change statements, but here too, these have to be marked as assumptions.
- It is useful to place the nodes such that the main cause and effect chains are easily seen, for example by placing these from bottom to top, as in Figure 2.3, or from left to right.

Every link is labelled with the source(s) of the statement(s) it represents, using the following abbreviations:

- [X]: the statement was published in the reference numbered X;
- [A]: the statement is an assumption;
- [E]: the statement is based on experience of the stakeholders;
- [O]: the statement is based on own investigations;
- [?] : it is not known whether a link exists.

If contradicting or differing sources are found, these can be represented by drawing a link for each source, each with its own sign-combination as shown in Figure 2.3 between ‘quality of problem definition’ and ‘number of modifications’.

If the literature provides statements that differ from what was assumed or experienced, it might still be useful to keep the links labelled [A] and [E], and add a new link reflecting the statement from the literature. This is particularly true for statements based on experience. The difference between the statement in the literature and experience might be due to a difference in context in which the statements were obtained. There could be difference in batch size, type of company, novelty of the product, subjects involved in the study, *etc*. If such a factor is causing a difference, it can be considered a relevant influencing factor and thus added as a node. The model can be enriched by giving each link an appropriate width to represent the amount of evidence available or the relative strength of one link compared to another. Figure 2.3 shows that according to reference 4 ‘time-to-market’ has a much stronger influence than ‘quality of the product’ on ‘amount of profit’.

Summarising the above, a statement about the existing or desired situation can thus be modelled as two or more nodes representing the factors involved, connected by a line that is marked at either end with the values of the attributes of the factors to represent the details of their relationship. In the case of a causal link, the line becomes an arrow, pointing at the effect. The link is labelled with the source of the statement. Figure 2.5 shows our modelling terminology using the graphical representation of a statement from the literature source 1 stating that a “high product quality has a positive effect on customer satisfaction”.

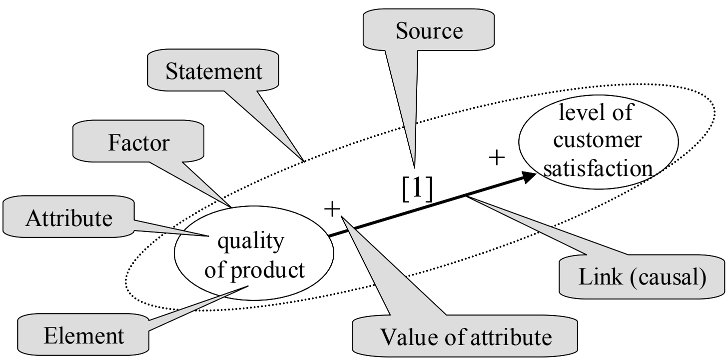


Figure 2.5 Graphical representation of a statement and associated modelling terminology

Note that the models, as we use them in this book, represent statements in a rather qualitative way. Some sources provide statements that are quantitative, including *e.g.*, mathematical equations linking two factors. These details can be added to the model, but care has to be taken not to overload the figures and obscure the overview they are intended to provide. A combination of overview model and partial models can be a useful way to convey details.

In the course of the research process, nodes and links might have to be added, removed, modified or 'opened up' as understanding grows. For example, when the factors that constitute 'level of creativity' have been identified as 'level of novelty" and 'degree of usefulness' (Chakrabarti 2006), or when the factors that cause the link between the "quality of problem definition" and "number of modifications" used in Figure 2.3 have been identified. The opposite might also be useful: to aggregate nodes and links into (new) higher level factors to support explanation or to 'collapse' the model temporarily to provide an overview.

2.4.2 From Reference Model to Impact Model

The Reference Model can be very similar to the Impact Model. An example is a Reference Model of the behaviour of designers in successful projects. Without much editing, it might be possible to create the Impact Model as a basis for developing a set of guidelines for good practice. In most instances, however, the Impact Model cannot be derived directly from the Reference Model. As discussed earlier, the existing situation usually represents a problematic situation that we wish to understand and then improve through the introduction of support. The desired situation is supposed to be different. Hence the model of the desired situation, *i.e.*, the Impact Model, has to be *generated* on the basis of the Reference Model.

Compared to the Reference Model, the Impact Model includes the support and the desired, expected, effects. This may require the introduction of new nodes and links, *e.g.*, auxiliary effects of the use of the support; the removal of existing ones, *e.g.*, those that are no longer relevant, once the support has been introduced; and the changes to the values of certain attributes.

These modifications usually require the introduction of assumptions, because there may be no available evidence of their validity. It is very important to make these assumptions explicit, so that the reasoning behind the Impact Model can be traced and judged. For example, even if a poor quality of the product reduces the amount of profit (the existing situation, shown in Figure 2.3), this does not necessarily imply that high product quality results in large profit (the desired situation). The latter remains an assumption and the corresponding link should be indicated as such in the Impact Model.

Figure 2.6 shows the Impact Model developed for our example on the basis of the Reference Model shown in Figure 2.3. The links have been modified to represent the desired effects. The links that do not have an effect in the current situation have been removed and replaced with links that are assumed to be brought into existence when the support is used. For example, the link between 'number of modifications' and 'quality of product' is removed. Instead, the latter is assumed to be influenced by the 'quality of design evaluation', which is influenced by the 'quality of problem definition'.

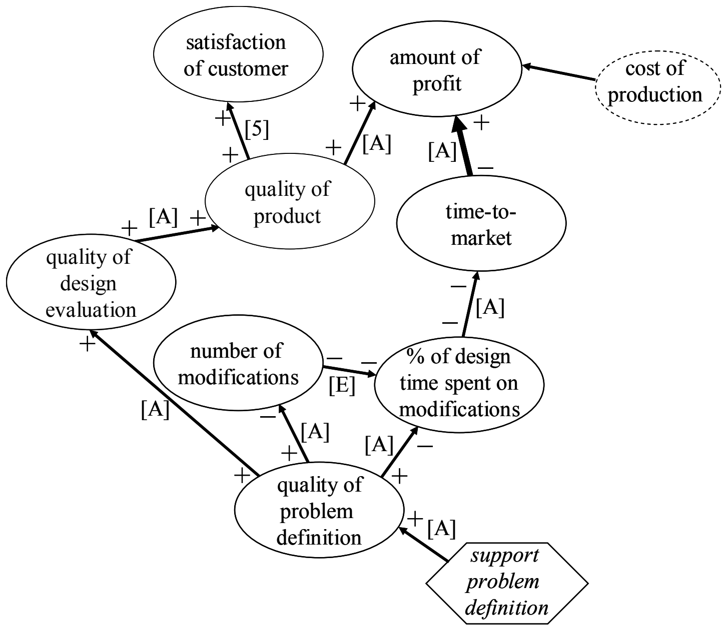


Figure 2.6 An Impact Model, representing the – partly assumed – desired situation after the introduction of the support (represented as an hexagonal element)

In Figure 2.6 the *support* is represented as an hexagonal element to distinguish it from the factors shown in ovals. The label used is the function of the support, *e.g.*, ‘support problem definition’. The support is currently linked to the Key Factor, ‘quality of problem definition’ in the above figure, by a causal link that has no sign at the support end. This means that at this stage, the details of the support are unknown: all we envisage is that the desired effect of the support is to increase the quality of problem definition. As the support is developed further, the Impact Model will be elaborated based on the support’s functionality, concept, implementation, introduction, customisation, use and maintenance, which may introduce new factors and links or modify existing ones.

Setting up a graphical representation of the existing and the expected situation in the form of Reference and Impact Models structures findings and clarifies thoughts. The resulting models help:

- to improve understanding by linking various findings and making explicit for which links evidence exists and which ones are based on assumptions;
- to identify realistic research areas and goals, and suitable criteria for judging the results;
- to illustrate and clarify the line of argumentation that shows the relevance of the research and the research approach;
- to determine whether the level of understanding is sufficient to develop support for improvement or whether too many assumptions are involved;
- to identify the factors to be addressed by the support (Key Factors);

- to illustrate one's vision by making explicit the expectations about the desired situation;
- to illustrate and clarify the line of argumentation for developing specific support;
- to encourage discussion and reflection on the existing and desired situation.

Unfortunately, the literature thus far shows very few attempts to draw up networks of influencing factors. A notable exception is Frankenberger (1997), see also Appendix C.4, Figure C.13. Our students found developing Reference and Impact Models to be a powerful method to clarify their thoughts, structure their understanding and reveal their assumptions. In our opinion, developing such models as a research community would reveal our current understanding of design and could act as an important basis for future research (Blessing 2003).

2.5 Success Criteria and Measurable Success Criteria

For a research area, such as design research, with the ultimate aim of improving a situation, formulating criteria for success is essential to be able to determine whether the results help achieve this aim.

Criteria are used to be able to focus the investigation of the existing situation; to assess the contribution of the findings of such investigations to the research goal; to focus the development of support on the most relevant factors; to plan the appropriate evaluation; to focus the realisation of the support on this evaluation; and to assess the evaluation results. That is, criteria are needed to be able to judge the outcome of the research against the research goals.

We define a **criterion** as the desired value of the factor the research project sets out to understand and/or influence as described in the research goal. In our example, the goal was to develop support to reduce time-to-market: the Criterion with which to judge the support resulting from this research project is thus 'short time-to-market'. A criterion can be relative or absolute, qualitative or quantitative. If a research goal refers to several factors, several criteria have to be formulated. Note that the research goals and criteria we discuss here only relate to one of the possible research outcomes, namely the support. Other criteria are needed to judge the scientific quality of the research.

A **preliminary set of criteria** has to be defined during the RC stage, since the choice of criteria will strongly influence the research approach and methods. We found it important to distinguish between Success and Measurable Success Criteria, although it might not be possible to make this distinction until more understanding has been obtained in the DS-I stage.

Success Criteria relate to the ultimate goal to which the research project or programme intends to contribute. These criteria usually reveal the purpose of the research and the eventual, expected contribution to practice. In our example, this was an 'increased amount of profit'.

In the Reference and Impact Models, Success Criteria relate to the **Success Factors**. These are the factors at 'the top' of the network, *i.e.*, at the end of the cause–effect chains that provide the justification of the research. The desired values of the Success Factors are taken as Success Criteria. In Figure 2.3 the potential Success Factors are ‘satisfaction of customer’ and ‘amount of profit’, of which the latter was chosen. Had the reason for that project been that customers were not satisfied because products were unreliable, ‘satisfaction of customer’ would have been the more suitable Success Factor. These choices clearly affect the focus of the research and the means of evaluation.

Success Criteria can relate to any of the facets of design (see Figure 1.1), but usually refer to long-term effects of the research, most of which can only be observed after the product has been produced and introduced into the market. Success Criteria we found in the literature include criteria as varied as: increased sales volume, return on investment, improved company image, optimal exploitation of company competences, increased competitive strength, sustainable development, improved team performance, reduced lead-time and improved product development process.

The definition of success is still a topic of research, as many factors influence success. As a consequence, there are no established metrics to measure success. Furthermore, even if the above-mentioned criteria were established as metrics, they would generally be difficult to apply within the timeframe of a research project. In our example, the duration of the research project makes it impossible to observe an ‘increase of the amount of profit’, even if the support were able to generate such an effect.

What is needed in such cases are **Measurable Success Criteria**, *i.e.*, criteria that are linked to the chosen Success Criteria and can be applied to judge the outcomes of the research, given the resources available within the project or programme. The factors whose desired values are taken as Measurable Success Criteria are the **Measurable Success Factors**. It is important to note that the term *measurable* refers to the possibility of measuring the criteria during the project, and not to the nature of the methods to assess the fulfilment of the criteria, *i.e.*, both qualitative and quantitative research methods can be used.

When it is not possible to use Success Criteria as Measurable Success Criteria, Measurable Success Criteria should be chosen such that they can serve as reliable *indicators* (also called *proxies*) for the Success Criteria. The link between Measurable Success Criteria and Success Criteria is assumed to exist (preferably based on existing evidence, otherwise on reasoning) and is therefore not evaluated in the research project. Therefore, it is important that the Measurable Success Factors are chosen such that these are as close as possible to the Success Factors, *i.e.* the link should be as direct and strong as possible. In this way, the likelihood, that the Success Criteria are fulfilled when the Measurable Success Criteria are fulfilled, is high. In our example, ‘time-to-market’ was chosen as the Measurable Success Factor, as this factor has the strongest link to the Success Factor ‘amount of profit’ (see Figure 2.7).

More than one Success Factor and one Measurable Success Factor may be chosen. In the example, ‘quality of product’ could have been a second Measurable Success Factor. The reasons for not choosing this factor were that it was considered too difficult to assess within the timeframe of the project, that its link with ‘amount of profit’ was not very strong, and that the literature had shown no link with the number of modifications.

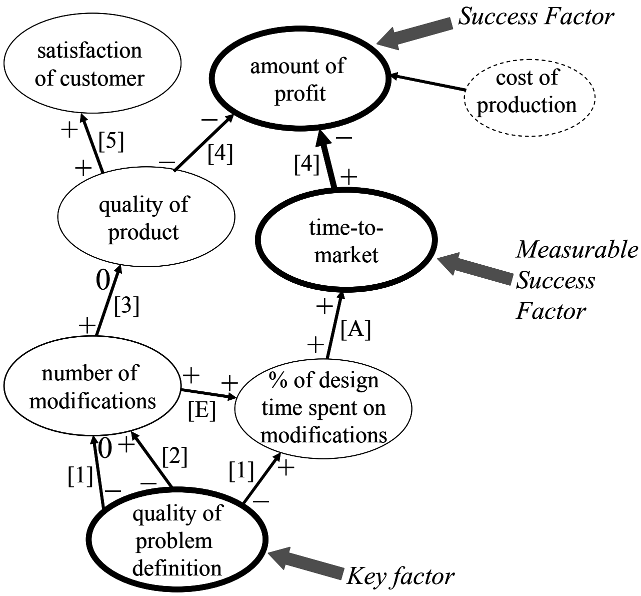


Figure 2.7 The chosen Success and Measurable Success Criteria as well as Key Factors for the Reference Model shown in Figure 2.3

The chosen criteria are transferred to the Impact Model shown in Figure 2.6, to identify the factors and links on which to focus the PS stage and in particular the evaluation in DS-II. During the course of the research project, it may be necessary or desired to select other criteria. In general, the Success Criteria remain relatively static. In many cases, the Measurable Success Criteria are redefined as understanding increases and specific support is developed. In our example, it was decided not to change or add criteria, but to leave out one part of the network in the evaluation as this part was considered not sufficiently influential. The result is shown in Figure 2.8.

The links between the criteria and the factors that the research is addressing can be very complex and the definition of success and suitable performance measures is still a topic of debate and investigation, see, *e.g*., Duffy (1998). However, these should not be reasons for not making these links explicit: assumptions can be introduced where evidence is missing. It is often necessary to piece together bits and pieces found in separate studies to form the overall argument linking the two (sets of) criteria. In some cases, it may be necessary to investigate the links between Success Criteria and Measurable Success Criteria, or part of it, as a study in its own right. This would be a project of Type 1 (see Figure 2.2).

We found that ‘amount of sales’, ‘amount of profit’ and ‘return on investment’ are the commonly used Success Factors in design research that used interviews and surveys in an industrial context. Success and Measurable Success Factors are (nearly) identical. In laboratory research, common Measurable Success Factors were ‘product quality’ – defined using the level of fulfilment of technical requirements – and ‘design time’ – defined using the ‘time spent to solve the given design task'. The links between these Measurable Success Factors and the Success Factors, *e.g.*, that 'product quality' influences 'market success', were partly derived from the literature and partly assumed. In these studies, the distance between the Measurable Success Factors and the Success Factors is large and the link much weaker. In such cases, it is particularly important to make the assumptions explicit and ensure that the claims are realistic.

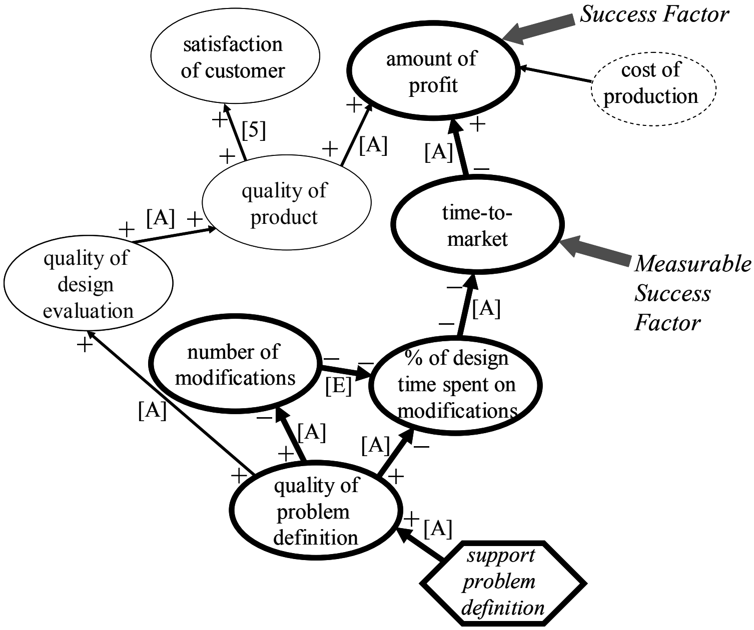


Figure 2.8 The Impact Model indicating the focus of the evaluation of the support

2.6 The Main Stages

In this section, the status of design research with respect to each of the four stages in DRM is presented, as well as their objectives and main deliverables. As to how to proceed in each of the stages is the subject of Chapters 4 to 6.

2.6.1 Research Clarification (RC)

In many research publications, the goals of the research projects refer to the improvement of design practice, *e.g.*, reducing lead-time or improving product quality. However, we found very few publications that provided evidence of a link between the stated goals and the actual focus of the research project, *e.g.*, improving communication between project members. The line of argumentation from the factors that are studied or addressed, to the factors mentioned in the goal, shows big gaps and is full of assumptions. In general, the goal is not used as the criterion against which the results of the research are judged: the relevant factors are often not even mentioned in the final section or chapter. As a consequence, little evidence exists that the goal has indeed been achieved. Often the reason is a practical one: the timescale of a research project does not allow an improvement of the effectiveness to be measured. Another reason is that the goals are often unrealistic for a single project: only few of the large number of interconnected factors can be addressed and therefore only a limited effect can be achieved.

Another issue is that if the research outcome is judged, the criteria are not made explicit and often seems to be based on assumption rather than evidence. Samuel and Lewis (2001) commented in a similar way on the lack of performance metrics in many studies. Take for example the oft-mentioned goal 'to improve the effectiveness of the design process'. In most cases, the terms are not defined to such a level that they could be used as criteria. What is effectiveness? What is a 'measure' of effectiveness? What type of design process is considered? What part of the design process is of interest? What is improvement: relative to what or with how much? For example, the term 'improve' in a goal has a considerable effect on the evaluation of the support: 'improve' is a relative term. In order to determine whether an improvement has been made, it is necessary to be able to compare the situation before and after the support has been introduced, or to compare situations with and without the support.

We have met many research students that aim, even for some years, at goals and criteria that are too abstract or too long-term, resulting in objectives, research questions and a project plan that are unrealistic. The RC stage intends to support researchers formulating a clear, challenging but realistic Overall Research Plan.

The objectives of the RC stage are:

- to identify the goals that the research is expected to realise; the focus of the research project; the main research problems, questions and hypotheses; the relevant disciplines and areas to be reviewed, and the area in which the contribution is expected;
- to develop Initial Reference and Impact Models, *i.e.*, an initial picture of the existing and of the desired situation;
- to identify a preliminary set of Success Criteria and Measurable Success Criteria against which to evaluate the outcome of the research;
- to provide a focus for the DS-I stage in finding the factors that contribute to, hinder or prohibit success;
- to help focus the PS stage on developing support that addresses those factors that are likely to have the strongest influence on success;
- to provide a focus for the DS-II stage for evaluating the effects of the developed support against the goals of the research.

The deliverables of the RC stage are:

- current understanding and expectations:
  - Initial Reference Model;
  - Initial Impact Model;
  - Preliminary Criteria.

- Overall Research plan:
  - research focus and goals;
  - research problems, main research questions and hypotheses;
  - relevant areas to be consulted;
  - approach (type of research, main stages and methods);
  - expected (area of) contribution and deliverables;
  - time schedule.

The approach and methods used in this stage are described in Chapter 3.

2.6.2 Descriptive Study I (DS-I)

The DS-I stage aims at increasing our understanding of design and its Success Factors by investigating the phenomenon of design through reviewing the literature about empirical research, undertaking empirical research, and, in addition, through reasoning.[6] The starting point is the Initial Reference Model drawn up during RC and the preliminary Criteria.

Investigating the phenomenon of design has been a very rapidly growing research area in the past decade. However, the current status is far from satisfactory, as we discussed in Section 1.3. The large variety of influencing factors studied and the variety of aims not only emphasise the complexity and extent of design as a research area, but also reveals, as a consequence of the low number of studies that deal with the same topic and the small number of cases in each study, the limited understanding we still have of design.

Most important in view of our methodology, is the fact that few studies focus on the explicit link between success and the influencing factors investigated. The usual focus is on links between pairs of influencing factors, occasionally linking these together, but without attempting to combine all together into a network of influencing factors that can form the basis for a comprehensive model or a theory of design and for the development of effective design support. The availability of results related to success or failure is particularly important for the development of design support. For example, a finding such as '20% of designers do X, 40% do Y and the rest do Z' can be useful for developing support, in the sense that these types of behaviour have to be taken into account. However, it does not provide any information for *improving* the situation, *i.e.*, to determine what to support and what to discourage to affect success. The link between this finding and success is needed.

This problem is aggravated by the fact that many publications do not provide sufficient details of the research approach to be able to compare different studies and to determine whether the findings are a suitable basis for one's own research. Most publications describe how data was collected, but not always detailed enough to determine the circumstances under which the data was collected. Very few

---

[6] Our use of the term 'descriptive study' is broader than its commonly used meaning in Social Sciences, and covers all types of study to investigate a particular phenomenon (for details see Section 4.1).

publications describe the data processing and analysis methods in sufficient detail to be able to verify the results. Moreover, the basic assumptions of the researchers that guided their interpretation of the data is not made explicit, and there is no evidence that the findings were validated. The analysis of Cantamessa (2001) of the 718 papers published in two major conferences on engineering design confirms our observations, showing that even basic information is lacking. He found that of the 111 papers describing empirical studies, 41% did not declare the sample size, 25% did not present the implications of the findings, 22% did not give the unit of analysis (the factor that was studied) and 10% did not state the research approach. As far as the research approach was described, he found little or no reflection on the methods and the approach that had been used.

We further noticed in many empirical studies inconsistencies between aim (criteria), research questions and hypotheses, data-collection method, data-analysis method, interpretations and conclusions. We found conclusions that cannot be drawn on the basis of the collected data, methods that are unsuitable to answer the research questions, *etc*. Moreover, findings, assumptions, interpretations and conclusions are often not clearly separated, thus providing a problematic or even unsound basis for further use.

There is often a tendency to use research methods that are most popular, rather than most suitable for the research goals and questions. A central reason for this is that many design researchers have an engineering background; in contrast to many other disciplines, research methods are usually not part of their curriculum. Where research training has been provided, this is likely to have covered quantitative methods for conducting natural science experiments. To investigate the phenomenon of design, a much wider variety of research methods, both qualitative and quantitative, from various disciplines has to be used to investigate the facets and aspects involved. Most design researchers will have heard of the more common of these methods, but are usually unaware of the underlying paradigms and lack knowledge about the pre-requisites for applying these methods.

In Chapter 4 on DS-I and in Appendix A, we attempt to address the above issues by providing a research approach, guidelines and methods, as well as summaries of existing methods from other disciplines, the main concepts to be familiar with, and pointers to the relevant literature. However, this information will never replace the importance of consulting experts in the relevant disciplines to ensure the most suitable methods are chosen and applied correctly.

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 the 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.

The approach and methods used in this stage are described in Chapter 4.

2.6.3 Prescriptive Study (PS)

Ultimately, design research is about developing support for improving design, even though this might not be the focus of an individual design project. The development of design support has a long tradition and is still a dominant research theme. However, there is little evidence of extensive use of valid empirical data: development relies on single findings, on assumptions and sometimes on experience. Many of the empirical results seem unknown to those developing support. Possibly because the research communities developed relatively independent of each other (see Section 1.3.1), and because many empirical studies do not establish links between influencing factors and success (see previous section).

We have also encountered the argument that it is not necessary to look into the existing situation in design, if the intention is to automate a particular task, rather than assisting the designer in executing this task, and if the support is not intended to mimic the human design process, but is to be based on another approach. In our view, it is always relevant to understand the existing situation, because this is the context in which the support has to be introduced and used, in order to address a particular problem or need.

Increasingly, we observe PhD projects starting with a small investigation of the current situation. Such investigations are important, but often unfortunately used as the only source. The following example shows how using all available understanding of the existing situation, rather than relying upon single findings, influences the potential success of the developed support. Several studies show that designers spend a large amount of time on collecting information, such as Beitz (1979) and Hales (1987). Based on this understanding, developing a computer tool to more easily access information seems to be a suitable solution. A more recent study, however, shows that although large amounts of information about past designs are available in digital form, personal contact is still the most frequently used source for information; the information designers need is often not contained in such databases (Marsh 1997). As a consequence of these findings, a more promising solution would be to develop a support that captures this information, rather than focusing only on supporting search. That this solution still might only solve part of the problem, is revealed by two other findings: in searching for solutions, successful designers restructure and summarise information (Fricke and Pahl 1991), and experts often rephrase the question when asked for information (Ahmed 2001). Capturing and storing information 'as given' is obviously insufficient. Another type of support has to be developed.

We see some evidence that the increasing number of empirical studies starts having an effect and expect that this will give new impulses to the development, improvement and implementation of support.

With regard to the support that is developed, we have observed that most publications do not reveal the view on design underlying the support, *i.e.*, the vision of the researcher about the desired situation and the role of the support. The assumptions upon which the support is based are often not made explicit or are presented as facts. The earlier mentioned analysis of publications on design research (Cantamessa 2001) showed that in 47% of the 331 papers on support, motivations are absent: only in 33% of papers were they defined precisely. Making the views and assumptions explicit is important, because these influence the development of the support and its likelihood of success.

We have further observed, that a considerable amount of time is spent on details of the support – in particular if this involves software development – rather than on its concept, although the core research contribution often lies in this concept. The aim of a research project is rarely to develop a commercially viable support. The aim, usually, is to define the envisaged support, the **Intended Support,** and realise this to such an extent that its core concept can be demonstrated and the effects evaluated. That is, the support that is actually realised, the **Actual Support**, might differ from the Intended Support. However, little help exists to develop demonstrators, prototypes or drafts that are sufficient to evaluate the concept.

Regarding the approach applied to develop the support, little is published and reflections on the approach are rare. Interestingly, much support aims at aiding a more systematic design process, but in developing the support (which is a design process in its own right) some of the basic principles of systematic product development, such as a thorough problem definition and the generation of variants, do not seem to have been followed. Support can take any form (guidelines, checklists, methods, equations, procedures, reorganisation proposals, *etc*., see also Footnote 3 (page 4), and medium (paper, software, models, workshops, *etc*.). The support can combine several forms and media, *e.g*., a checklist to collect 'the voice of the customer', a software programme to process this data, and guidelines on how to incorporate the results in a product. Unfortunately, few of these possibilities seem to be considered when developing support.

We believe that a more systematic way of developing design support in a research project can address the above issues, if this approach includes: the use of empirical data; the development of a model of the desired situation to reveal the underlying vision and assumptions; the distinction between the envisaged, Intended Support and the Actual Support developed for evaluation; and the use of the basic principles of systematic product development.

The objectives of the PS stage are:

- to use the understanding obtained in DS-I or DS-II to determine the most suitable factors to be addressed in PS (the Key Factors) in order to improve the existing situation;
- to develop an Impact Model, based on the Reference Model and the Initial Impact Model, describing the desired, improved situation that is expected as a consequence of addressing the selected Key Factors;
- to select the part of the Impact Model to address and to determine the related Success and Measurable Success Criteria;
- to develop the Intended Support, that addresses the Key Factors in a systematic way, and to realise this to such a level of detail that an evaluation of its effects can take place against the Measurable Success Criteria;
- to evaluate the Actual Support with respect to its in-built functionality, consistency, *etc.*, – the **Support Evaluation** – in order to determine whether to proceed to DS-II to evaluate the effects of the support;
- to develop an **Outline Evaluation Plan** to be used as a starting point for the evaluation in DS-II.

The deliverables of the PS stage are:

- documentation of the Intended Support:
  - Intended Support Description: what it is and how it works;
  - Intended Introduction Plan: how to introduce, install, customise, use and maintain the support as well as organisational, technical, infrastructural pre-requisites;
  - Intended Impact Model;
- actual Support: workbook, checklist, software, *etc.*
- documentation of the Actual Support:
  - Actual Support Description;
  - Actual Introduction Plan;
  - Actual Impact Model;
- results of the Support Evaluation;
- Outline Evaluation Plan.

The approach and methods used in this stage are described in Chapter 5.

2.6.4 Descriptive Study II (DS-II)

The DS-II stage focuses on the evaluation of support. In many PhD dissertations we have found that the developed support is not really evaluated in a way that allows an assessment of its effects, although realising these effects is said to be the goal of the research project. In other words, what is evaluated is not in line with what is claimed.

In particular we have seen inappropriate generalisations, where 'generic methods' are developed based on the analysis of a specific problem and evaluated using the same problem. In many cases, statements are made about the use of the support, although the evaluation involved only the researcher. Moreover, the developed support is often evaluated using existing products or processes only, that is, products and processes that are already known. In order to see the effect of a support, it needs to be applied without knowing the outcome. Furthermore, design support is expected to be used, eventually, to address the needs and problems that triggered its development. The emphasis on *use* implies that the human factor and the actual introduction and maintenance of the support in the user environment have to be considered. Research projects rarely address these issues. Hence, most evaluations are unlikely to reveal the *real* issues of using the support for design.

Notwithstanding this criticism on current evaluation practice, these evaluations can be a useful starting point for a first identification of the major issues. However, a more detailed evaluation addressing the above issues is required if the evaluation results are to be used: to determine whether the goals have been achieved; to inform improvement of the support; to increase our understanding of design; and to suggest how introduction should take place.

One of the reasons for the observed situation is the lack of involvement of users and practice *throughout* design research projects: to understand the current situation, to inform support development, and to evaluate. Fortunately, more and more researchers do involve users. More persistent reasons we hear for the lack of detailed evaluation are: the lack of time and availability of users; the supposed impossibility of obtaining valid results using a small number of cases; and the limitations of the actual support. A detailed evaluation is indeed time consuming; finding users and settings to evaluate the support is often problematic and may indeed result in a very small number of cases; and we cannot expect the actual support to be complete – in many cases it is a prototype or demonstrator of the intended support with limited functionality, robustness and coverage. A detailed evaluation can therefore be very difficult, requiring careful thought. Even then, generalisation of the results may be limited. As a consequence, detailed evaluation of the developed support is often neglected, but without it we can say little about the success of the support, as one of the outcomes of research. In our view, this is one of the main reasons that much of the design support developed in academia is not taken up in practice (see Section 1.3.2), and often not even by other researchers.

Many ways exist in which design support can be evaluated, but creativity is required to set up a proper evaluation that fits the aims and constraints of the project, while at the same time provides enough confidence in the proposed support. This makes DS-II a challenging but not impossible task: when well thought out, it is possible to carry out empirical studies that provide useful evaluation data within the timeframe of a research project. For this, the evaluation should be kept in mind *in all stages* of the research project.

Unfortunately, little guidance exists for selecting suitable evaluation methods. The 1996 NSF Workshop on research opportunities in engineering design concluded that “methods for validating the results of research in design need to be developed. We need to have the means with which to determine the value added by a tool or method, its reliability, and its scalability to practical problems. Such tests are hard, but without them we are doing philosophy” (Shah and Hazelrigg 1996). Nothing much has changed since, but the issue of evaluation is increasingly being discussed in the design research community (Frey and Dym 2006; Seepersad *et al.* 2006). Looking into other areas and their ways of evaluating research results that are intended for practice, we found interesting approaches and guidance, which are addressed in Chapter 6.

Based on our observations, we consider it necessary to distinguish between two types of evaluation, in addition to the commonly applied Support Evaluation in PS. The first type of evaluation, the **Application Evaluation**, aims to identify whether the support can be used for the task for which it is intended and that is does address the factors that are directly influenced (the Key Factors) in the way they are supposed to be addressed, *i.e.*, the focus is on *usability* and *applicability*. Using our earlier example, we need to investigate whether users understand the support that has been developed, whether they can use it, and whether it indeed improves the quality of the problem definition. In terms of the Impact Model we need to investigate the effect of the support on the value of the Key Factor(s). In the example shown in Figure 2.8 this is the effect of the support on the ‘quality of problem definition’. Some research projects do address this type of evaluation.

However, whether a positive effect on the Key Factor(s) indeed contributes to success, *i.e.*, whether the support is useful is not certain. The second type of evaluation, the **Success Evaluation**, therefore aims to identify whether the support has the expected impact *i.e.*, whether the desired situation represented in the Impact Model has been realised, taking into account that unexpected side-effects may occur. The focus is on *usefulness*. Using our earlier example, (see Figure 2.8), we need to investigate whether the percentage of time spent on modifications and the number of modifications have been reduced, and, most importantly, whether this has reduced time-to-market, *i.e.*, we need to verify the links from the Key Factor to the Measurable Success Criteria.

To a certain extent, DS-II also validates the findings of DS-I: the understanding gained from evaluating the support enables the evaluation of the Impact Model, which in turn enables the evaluation of the Reference Model, as well as a reflection on the chosen Success and Measurable Success Criteria. In that sense, DS-II also contributes to our understanding of success and the definition of metrics of success.

Evaluation is an essential part of development of support, and could therefore have been included as an activity in the PS stage. However, the decision to separate development from evaluation was taken deliberately to highlight the importance of formal evaluation of support and to make explicit the difference between the approach and methods required.

The approach and methods used in DS-II are similar to those in DS-I, but the aims are different: the aim of DS-I is to understand design, the aim of DS-II is to understand the impact of a support.

The objectives of the DS-II stage are:

- to identify whether the support can be used for the task for which it is intended and has the expected effect on the Key Factors (Application Evaluation);
- to identify whether the support indeed contributes to success (Success Evaluation), *i.e.*, whether the expected impact, as represented in the Impact Model, has been realised;
- to identify necessary improvements to the concept, elaboration, realisation, introduction and context of the support;
- to evaluate the assumptions behind the current situation represented in the Reference Model, and the desired situation represented in the Impact Model.

The deliverables of the DS-II stage are:

- results of the Application Evaluation;
- results of the Success Evaluation;
- implications and suggestions for improvement for:
  - the Actual Support;
  - the Intended Support, its concept, elaboration and underlying assumptions;
  - the Actual and Intended Introduction Plan including introduction, installation, customisation, use and maintenance issues;
  - the Actual and Intended Impact Model;
  - the Reference Model;
  - the criteria used.

The approach and methods used in this stage are described in Chapter 6.

2.6.5 Summary

Figure 2.1 can now be extended to include the deliverables of each stage. The result is shown in Figure 2.9.

2.7 Comparison with Other Methodologies

Few publications exist on DRM, although the need for addressing the methodological issues has been discussed for some time, *e.g.*, in Antonsson (1987); Duffy and Andreasen (1995); Eckert *et al.* (2003); Reich (1994a); Reich (1994b); Reich (1995), and our own publications. Some proposals for a methodology have been made in the area of engineering design, notably Bracewell and Shea (2001; Duffy and Andreasen (1995); Eckert *et al.* (2003); Langdon *et al.* (2001); Stacey *et al.* (2002). Researchers in the more artistic design areas, such as industrial design, graphic design and sculpture, are involved in a large interesting debate about design as research, see, *e.g.*, Buchanan (2004); Dilnot (2004); Galle (2002); Love (2002) but no methodology has been proposed.

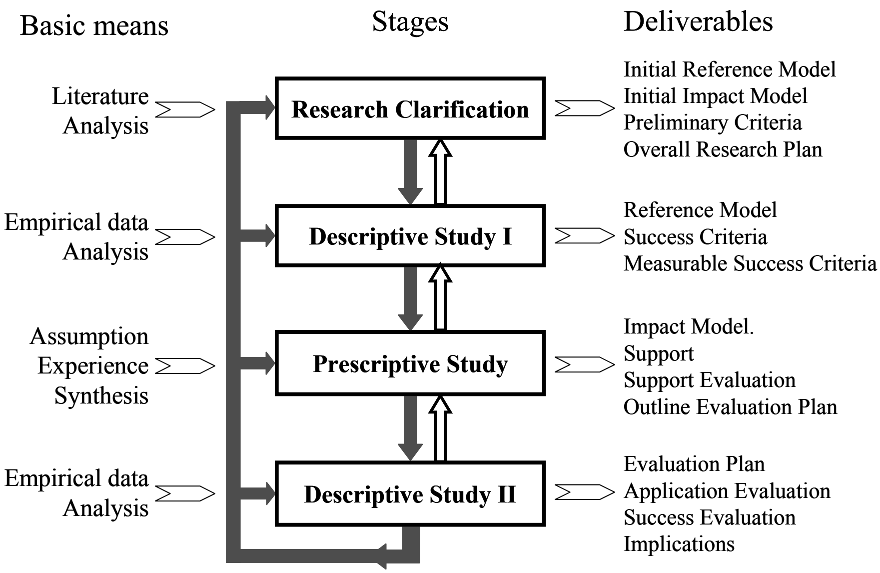


Figure 2.9 DRM framework: stages, basic means and deliverables

Two methodologies that are close to DRM are discussed below: the research framework and methodology of Duffy, Andreasen and O'Donnell, and the Soft Systems Methodology of Checkland.

The research framework developed by Duffy, Andreasen and O'Donnell (Duffy and Andreasen 1995; Duffy and O'Donnell 1999) stresses the need "to facilitate the research and development of appropriate means to support design […] and its management based upon a fundamental understanding of design." They develop the framework for conducting design research shown in Figure 2.10 "based upon the hypothesis that any developed tools (be they human or computationally based) will make an impact upon the design process itself when employed". Similar to DRM, they introduce criteria for evaluation based on reality and models: "The reality and models would act as the criteria upon which to base critical and objective evaluations of the consequent models, but when employed as tools would affect the 'reality' in which design is carried out" (Duffy and O'Donnell 1999). Because of the latter, they too distinguish between descriptive and prescriptive models: the former based on reality (our Reference Model) and the latter on "the envisaged or foreseen reality that would be considered as enhancing design practice" (our Impact Model). In contrast to DRM, their framework is characterised by three models: a phenomena model, a knowledge model and a computer model. This difference is in line with the focus of their framework: the development of computer support.

The general research methodology related to the framework consists of six steps: Design problem; Hypothesis; Research problem; Solution; Formal evaluation; and Documentation. The literature informs the first four steps; design practice informs the first and fifth step. Unfortunately, no details about the framework and methodology are available.

The methodology of Duffy and Andreasen shows overlap with the three-level model of evaluation described by Smithers in Donaldson (1991): (1) knowledge level: tests models and theories of the design process; (2) symbol level: tests the capability of knowledge representation and of control knowledge and its application; and (3) system engineering level: tests the implementation. Level 1 is similar to the validation of the Reference Model in DS-I in DRM; level 2 to the Support Evaluation in PS and part of the Application Evaluation in DS-II; level 3 relates to the evaluations in DS-II, although the distinction between Application and Success Evaluation is not made.

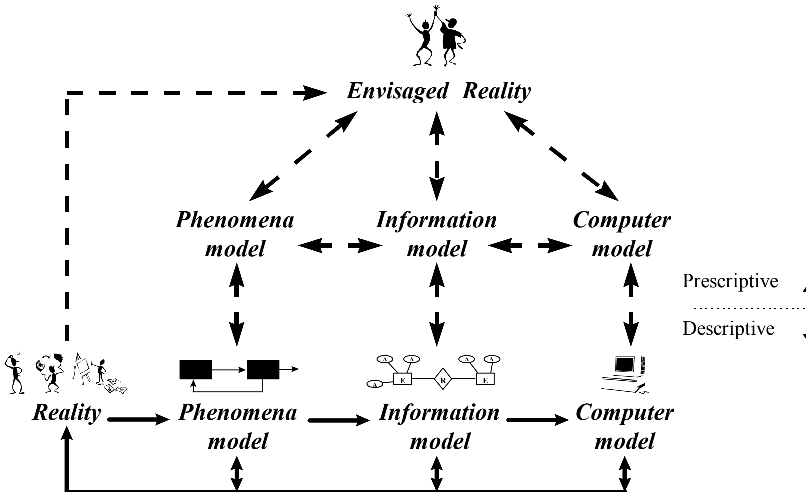


Figure 2.10 The DRM proposed by Duffy and Andreasen (1995)

The Soft Systems Methodology (SSM) of Checkland (1981; 1999) is the result of an Action Research programme, with the aim "to find ways of understanding and coping with the perplexing difficulties of taking action, both individually and in groups, to 'improve' the situations which day-to-day continuously creates and continually changes" (Checkland 1999). Action Research is an approach for introducing and evaluating change, originally in organisations and programmes, but increasingly within design research (see Appendix A.4.9 for more details). SSM is "concerned with problem situations, not with problems, in which there are felt to be unstructured problems" (Checkland 1981) and "explores the value of the […] ideas captured in the notion of 'system'" (Checkland 1999) to find solutions that are "feasible and desirable" (Checkland 1981).

The main stages of SSM show a strong similarity with DRM. First, reality is analysed and a description of the essence created. Based on this, a description of the ideal situation is created, which is compared with reality to generate proposals for improvement of reality. The proposals are introduced and the 'new' reality is analysed. This cycle is repeated until the results are satisfactory. The result of each cycle is not only an improvement of reality, but also a better understanding of reality and of the quality and effects of the proposed actions.

In contrast to DRM, the main focus of SSM, and other action research approaches, is on on-site evaluation of the newly developed support, which is prone to result in local solutions and gradual improvement using short-cycles between support generation and evaluation. SSM is very much embedded in practice, which is also its advantage. In DRM, the aim is to generate more generic solutions, evaluating the initial support in a realistic, but not necessarily the real situation, and to do these using fewer but longer cycles. One reason for the differences is the types of solutions for which the methodologies were initially developed. Organisational changes and welfare programmes, the original areas of action research, are more localised and cannot be evaluated off-site, in contrast to most of the design support on which DRM focuses. An overlap, however, exists and many of the methods described in SSM can be usefully applied as part of DRM.

2.8 Main Points

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

- This chapter presents the outline of our methodology for design research DRM and an overview of its stages and main concepts.
- The specific objectives of DRM are to provide: a framework for design research, help to identify research areas and develop the argumentation; guidelines for research planning; guidelines for rigorous research; help to select research methods; a context for positioning research projects and programmes; and, encouragement to reflect on the approach.
- While using DRM one must be flexible and opportunistic, pursuing promising, unexpected avenues that may lead to new solutions.
- The main stages of DRM are: Research Clarification (RC), DS-I, PS and DS-II.
- RC helps clarify the current understanding and the overall research aim, develop a research plan and provide a focus for the subsequent stages. The deliverables are: an Initial Reference Model, an Initial Impact Model, a preliminary set of Criteria, and an Overall Research Plan.
- DS-I aims at increasing the understanding of design and the factors that influence its success by investigating the phenomenon of design, to inform the development of support. The deliverables are: a Reference Model, an updated Impact Model, Success and Measurable Success Criteria, as well as implications of the findings for the development of support.
- PS aims at developing support in a systematic way, taking into account the results of DS-I, developing an Impact Model, developing support (distinguishing between Intended and Actual Support), and undertaking continuous Support Evaluation. The deliverables are: an Impact Model, descriptions of the Intended and Actual Support, the Actual Support, Support Evaluation results, and an Outline Evaluation Plan.
- DS-II focuses on evaluating the usability and applicability of the Actual Support (Application Evaluation) and its usefulness (Success Evaluation). The deliverables are: Application and Success Evaluation results and suggestion for improvement of the Actual and Intended Supports, as well as the Reference and Impact Models.
- Both DS-I and DS-II aim at developing an understanding of the phenomenon of design. While DS-I aims to understand design 'as-is', DS-II

aims to understand the impact of a support. DS-II involves intervention into the process – the introduction of the support.

- Based on the depth in which individual stages are executed, seven different types of design research are distinguished.
- DRM is not a set of stages and supporting methods to be executed rigidly and linearly. Multiple iterations within each stage and between stages are possible, as well as parallel execution of stages.
- A project can start in any of the stages of DRM, but the links to the other stages should be addressed, even if these are not executed within the project: research should both build upon existing and contribute to future research.
- In DRM, descriptions of the existing and the desired situation are modelled as networks of influencing factors. Factors are aspects of the situation under consideration that influence other aspects of this situation.
- The Reference Model represents the existing situation in design and acts as a reference for benchmarking intended improvements.
- The Impact Model represents the desired situation, and shows the envisaged impact of the support. The models develop as understanding grows.
- In most instances, the Impact Model cannot be derived directly from the Reference Model. The introduction of assumptions is necessary to represent the desired situation.
- An influencing factor is represented as an attribute of an element for which an operational definition can be formulated. Key Factors are those influencing factors addressed directly by the support. The links between factors represent how the factors (are desired to) influence each other.
- Criteria are the desired values of the factors a research project sets out to understand and/or influence, as described in the research goal. Criteria can be relative or absolute, qualitative or quantitative. These are used to judge the outcome of the research against the goals.
- Success Criteria relate to the ultimate goal to which the research project or programme intends to contribute.
- When Success Criteria cannot be used to judge the outcome of the research, given the resources available in the project, Measurable Success Criteria are selected that can serve as reliable indicators of the Success Criteria. The term measurable refers to the possibility to measure the criteria within the project.
- The support can take any form (guidelines, methods, equations, reorganisation proposals, *etc.*) and medium (paper, software, models, workshops, *etc.*).
- The Actual Support is a prototype or demonstrator of the Intended Support with limited functionality, coverage, and performance, but sufficiently developed to enable the evaluation of the core contribution of the researcher.
- Evaluation should be kept in mind in all research stages.

3

Research Clarification

This chapter focuses on the first stage of DRM, the RC stage. It discusses an approach and methods to support the initial stage of a design research project or programme. The aims are to identify and refine a research problem that is both academically and practically worthwhile and realistic. This involves obtaining an overview of the available understanding of the area of interest, so that it is possible to plan for the most suitable research to solve this problem. Both the understanding and the research plan will continue to evolve as the project or programme progresses through its various stages, in particular during the DS-I stage.

Referring back to Section 2.6.1, the objectives of the RC stage are:

- to identify the goals that the research is expected to realise; the focus of the research project; the main research problems, questions and hypotheses; the relevant disciplines and areas to be reviewed, and the area in which the contribution is expected;
- to develop Initial Reference and Impact Models, *i.e.*, an initial picture of the existing and of the desired situation;
- to identify a preliminary set of Success Criteria and Measurable Success Criteria against which to evaluate the outcome of the research;
- to provide a focus for DS-I in finding the factors that contribute to, hinder or prohibit success;
- to help focus the PS stage on developing support that addresses those factors that are likely to have the strongest influence on success;
- to provide a focus for DS-II for evaluating the effects of the developed support against the goals of the research.

The deliverables of the RC stage are

- current understanding and expectations:
  - Initial Reference Model;
  - Initial Impact Model;
  - Preliminary Criteria;

- Overall Research Plan:
  - research focus and goals;
  - research problems, main research questions and hypotheses;
  - relevant areas to be consulted;
  - approach (type of research, main stages and methods);
  - expected (area of) contribution and deliverables;
  - time schedule.

3.1 Research Clarification Process

We have divided this stage into six, iterative steps, shown in Figure 3.1.

1. Identifying the overall topics of interest. This involves identifying potential research goals and problems (see Section 3.2) using the initial understanding and expectations represented in the first Initial Reference and Impact Models.
2. Clarifying the current understanding and expectations. This involves developing further the Initial Reference and Impact Models using relevant literature to identify the state-of-the-art with respect to what problems are already solved and what remains to be solved (see Section 3.3).
3. Clarifying criteria, main questions and hypotheses. This involves identifying potential criteria against which to judge the results of the research; and formulating appropriate research questions and hypotheses, based on the Initial Reference and Impact Models (see Section 3.4).

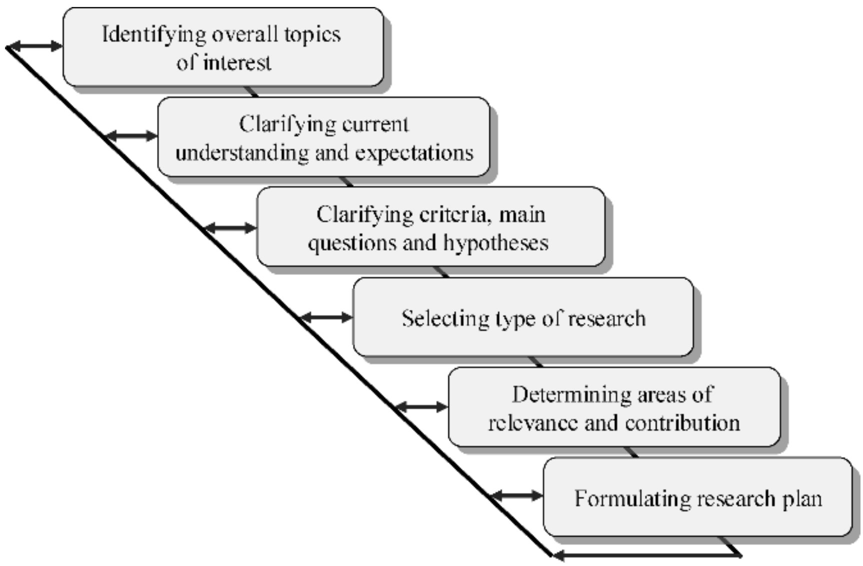


Figure 3.1 Main steps in the Research Clarification stage

4. Selecting type of research. This involves identifying the type of design research to be undertaken in order to solve the research problem (see Section 3.5).
5. Determining areas of relevance and contribution. This involves identifying the relevant knowledge areas and disciplines to be consulted to solve the research problem, and the areas and disciplines to which the research is intended to contribute (see Section 3.6).
6. Formulating Overall Research Plan (see Section 3.7).

The steps blend three strands: a gradual focusing on the research topic and the best approach; a gradual identification of the state-of-the-art and the possible contribution; and a gradual identification of the relevant criteria.

Reliability Example

In this and the next chapters, we will use the following example to illustrate the steps. A university research institute specialising in engineering design is approached by a manufacturer of large mechanical systems, who has serious problems with the reliability of these systems, despite the use of Design for Reliability methods. The company is not interested in a solution, such as the individual improvement of each existing system that has problems, but is looking for a generic solution, *e.g.*, a design support to prevent the reliability problems from occurring. This is in line with the interests of the research group and a meeting is set up to define the research project. A newly appointed researcher is given the assignment and the recommendation to follow DRM to guide his work. It is important to note that the entry point for a particular project can differ from the project used as example, and also the RC stage does not necessarily follow a linear process as this example may suggest (see Section 3.3).

3.2 Identifying Overall Topic of Interest

There are three central aspects that constitute an overall *topic of interest*:

- issue of interest (*e.g.*, reliability of designs);
- activity and/or stage of the design process (*e.g.*, evaluation, embodiment design); and
- area of application (*e.g.*, mechanical design).

The chosen issue, activity, stage and area may be given by an external partner, such as a company, government or a local interest group. The topic may also have been suggested by the researchers themselves, for example, based on the results of earlier projects, on the results of research and developments in other research groups, on personal interests, or on issues raised in the research community. In design research, the main reason why a particular issue in a particular area is considered of interest, is that the researcher, research group or research sponsors believe that this issue has an effect on design practice, although concrete evidence may not exist.

In research, as we see it, the goal is to identify and solve problems of interest that have a degree of generality, so that the fruits of the research contribute to our general understanding and – in design research – are applicable not just to a single product or a single practice, but possibly to a variety of these. Therefore, it has to be verified whether the focus area can qualify as a research area, *i.e.*, whether it is:

- *academically worthwhile*, that is, the problem is sufficiently challenging and generic and its solution is expected to contribute to our knowledge and understanding (as to what constitutes 'sufficient' varies with the available resources, *i.e.*, a three-month Bachelor's project will differ from a 5-year research programme for a group of researchers);
- *practically worthwhile*, that is, the problem has importance for practice beyond the practices of the stakeholders involved in the project, and the solution is expected to be sufficiently beneficial;
- *realistic,* that is, the research needed to address the problem is expected to be of a magnitude that can be tackled within the constraints of the project or programme.

Given our definition of design research, the problem should fulfil all three requirements.

The first suggestion for a topic may not necessarily be academically and practically worthwhile. Discussion with the stakeholders, *i.e.*, the researcher, research group, sponsors and/or practice should help to clarify the boundaries of the topic and to identify the relevant aspects and influencing factors, the most important problems and questions, as well as the criteria of success that are important for the stakeholders. The main aim is to make the beliefs and expectations of each of the stakeholders explicit, in order to obtain a first *shared* picture of the existing and the desired situations, and of the expected criteria against which to judge the research, *i.e.*, the Success Criteria.

The following checklist can be useful to guide the discussion:

- What problems/questions are important for each of the stakeholders to solve/answer? Note that a problem might have been recognised but it might not have been possible to define it.
- What benefits are solving/answering these problems/questions expected to bring to each of the stakeholders?
- What has already been tried to solve/answer these?
- How well did these solutions work? What are the reasons – known or believed – as to why the solutions did not have the expected impact? Which factors might have played a role?
- What (types of) solutions could possibly solve/answer the problems/questions?
- How could these solutions/answers be obtained?

The discussions should not only aim at gathering and documenting the available evidence, but also the underlying assumptions and beliefs. We found that many dissertations we reviewed are based on assumptions related to a particular worldview of design that is often not made explicit: *e.g.*, 'designing is information processing' and 'a more systematic approach is beneficial'. These assumptions have to be made explicit and analysed together with any other views (such as whether computers in design should support or automate) in order to identify their effect on the choice of research problem, criteria, research methods, *etc*

It is useful to start developing models of the existing and the desired situation (the Reference and Impact Models) to arrive at a shared view of the initial understanding and expectations. Discussions like the above result in:

- a preliminary set of influencing factors thought to be relevant;
- those factors that may be suitable Success Factors;
- the believed links between the factors in the existing as well as the desired situation, in particular those linked to the Success Factors;
- research problems worth investigating.

Reliability Example

In our example, a discussion between the company and the research group using this checklist, results in the following shared understanding, which is represented using networks of influencing factors.

- It is important to sort out the reliability problems in these systems.
- These problems are causing large maintenance cost, subsequently leading to loss of profit (for these systems maintenance costs are carried by the company), are putting the company's image at stake, and are believed to affect the market share – at least in the long term (see Figure 3.2, a first representation of the existing situation).

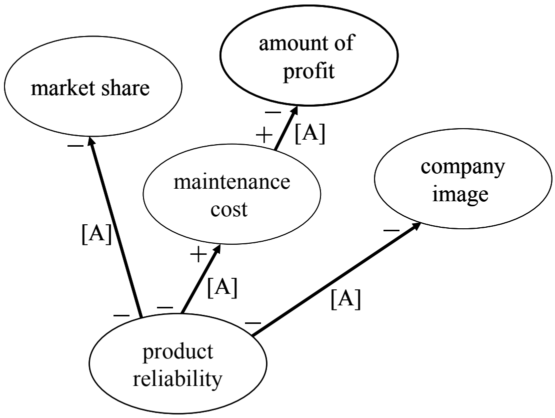


Figure 3.2 The shared understanding of the *existing* situation (all links still based on assumptions [A])

- Design for Reliability (DfR) methods have already been applied in the company, but in their experience, these did not improve reliability (see

Figure 3.3). Reliability is still considered less than what the company thinks necessary to be competitive.[7]

- It is believed that the company needs other methods for improving reliability of its products and that this will turn the existing situation around; one idea is to focus on better ways of *assessing reliability* of its products (see Figure 3.4).

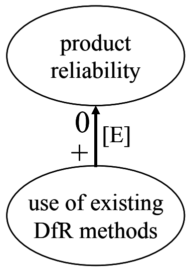


Figure 3.3 Partial Initial Reference Model: The use of DfR methods had no influence on product reliability (based on experience [E])

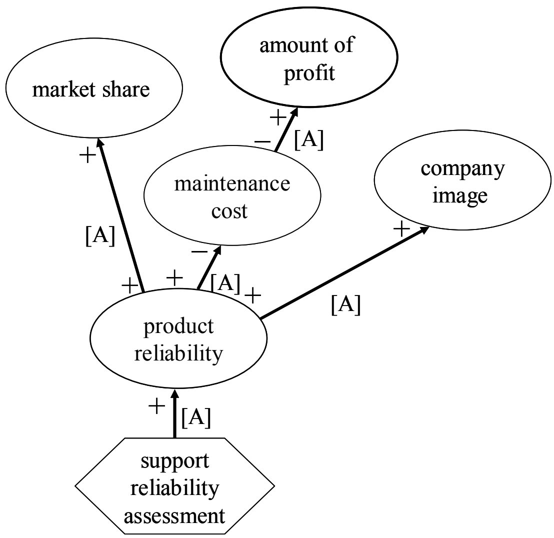


Figure 3.4 The shared view of the *desired* situation, showing the assumed impact of a support to assess reliability

---

[7] This example illustrates that statements about existing design support are not only made in DS-II, in which existing support is evaluated. Such statements can also be part of RC or DS-I, as long as the main focus is not on evaluation of the support. If it still remains unclear at the end of the RC or DS-I stage why existing support does not work, and this is considered a key issue, then a DS-II will be required, before new support is developed.

In order to obtain a better feeling for whether the desired situation is realistic and practically worthwhile, it is necessary to better understand the reasons behind the problems in the existing situation. This will also provide important information for the development of support that is supposed to address these reasons, and will help the research group decide whether the problem is sufficiently challenging and generic to be academically worthwhile.

In our example this raises the following questions: Is the desired situation as shown in Figure 3.4 realistic and practically worthwhile, *i.e.*, would developing support for assessing reliability be possible and solve the issues? Why did the applied DfR methods not have the desired effects?

A further discussion between the company and the research group resulted in the following additional understanding:

- If a method does not show effect, the inappropriate *use* of the method can be a reason. If the method is well established and applied with success elsewhere, the problem lies in the way the company uses the method and cannot count as a research problem (see Figure 3.5, left arrow). The methods used in the company are well established and are used correctly according to the company, but they saw little or no effect (see Figure 3.5, right arrow). The question is why these methods do not have an effect in this company.

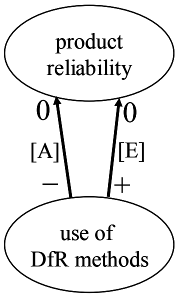


Figure 3.5 Partial Initial Reference Model: Possible reasons for the lack of reliability: existing method not applied correctly (*left arrow*, assumption) or method applied correctly but little or no effects (*right arrow*, according to the experience in the company)

- The company suggests two reasons as to why the application of existing DfR methods did not have an effect. First, the established methods can only be applied in the detail design stage. Discussions with the designers revealed that the methods sometimes *did* identify reliability problems, but that these problems could not be addressed to a satisfactory degree because of the advanced stage of the project. Secondly, the methods are not specifically developed for the types of machine system the company develops and it might be that they therefore do not apply well. The resulting partial model is shown in Figure 3.6. The lines connecting the edges indicate that the result – in this case ‘0’ or ‘no effect’ – occurs when both statements apply.

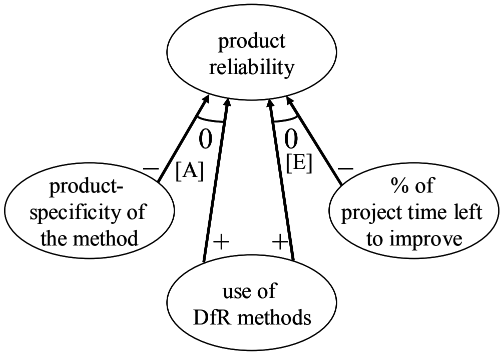


Figure 3.6 Partial Initial Reference Model: Reasons why correct application of a method may not have the desired effect

- The researcher sees an interesting research opportunity in the fact that it seems to be the current applicability of DfR methods in the *detail* design stage, which limits the effect. He suggests to focus on reliability assessment in *earlier design* stages, assuming that earlier detection of potentially unreliable product solutions is not only possible and more effective, but also more efficient. The earlier a problem is identified, the easier it might be to solve. The researcher illustrates this argumentation by expanding the initial description of the desired situation (shown in Figure 3.4) into the Initial Impact Model of which the relevant part is shown in Figure 3.7, indicating the assumptions made at this stage. Note that all possible influencing factors related to the support, such as quality of application, quality of introduction, time needed for application, *etc*., are not included yet.

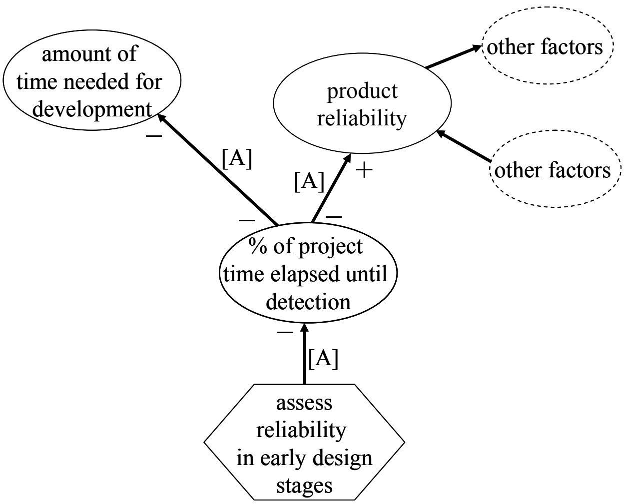


Figure 3.7 Part of the Initial Impact Model, introducing development time as a new factor

Investigating this problem seems very interesting from an academic point of view and the first formulation of the *research problem* is chosen to be 'How to assess reliability of mechanical systems in an early stage of the design process'. Several issues remain to be resolved, in particular: How early is this possible? And what has already been done to address this issue?

Choosing Area of Research

It is important to choose the area of research carefully. In our example, the choice of research area was strongly driven by a practical point of view and could thus be specified relatively early in the research project, but care has to be taken that the area is also academically worthwhile. Several authors give suggestions for selecting areas from an academic point of view.

- Do not follow the crowd (Oliver 1991). It may be beneficial to learn first what the crowd is doing in a partticular branch of science and spend substantial time visualising what is going on. It will then become clear what the crowd is not doing, and promising, unexplored directions may emerge.
- Take a long-term perspective. Often, it is helpful to see your area in long-term perspectives, and looking at a known area from a fresh perspective may lead to the emergence of new and exciting problems (Oliver 1991).
- It is important problems that lead to significant contributions, and the importance of a problem does not correlate with the difficulty in solving the problem (Thomson 1957).
- To locate problems, one can "play contradictions" – invent possible contradictions and see if they are true – or "play implications" – push the idea to its limits and see if it works (Root-Bernstein 1989). As Kuhn (1970) suggests, revolutions follow the recognition of anomalies.
- "Undertake a project manifestly important and nearly impossible. If it is manifestly important, then you don't have to worry about its significance. If it is nearly impossible, you know that no one else is likely to be doing it, so if you succeed, you will have created a whole domain for yourself." (Edwin Land the inventor of the Polaroid process and Land camera, quoted in Root-Bernstein 1989).
- It is important to "rebel but wisely" (Oliver 1991). Those seeking to make major contributions do not serve by adopting unsound positions no matter how unconventional or superficially appealing the position may be. It may be helpful to seek the non-questions – those questions that might have been asked but somehow forgotten or ignored.

The general idea is that the researcher should be able to question the dogmas and preconceptions of the field of enquiry, and challenge the existing boundaries.

3.3 Clarifying Current Understanding and Expectations

The results of the first step are topics and areas of interest shared by the stakeholders, and an initial understanding of the problem and solution directions, expressed in first, tentative models of the existing and desired situations. An exploratory review of the literature is necessary to clarify the current understanding and expectations, among others, by identifying the extent to which the problems are already solved in practice or academia, and what still remains to be solved. Such a 'reconnaissance survey' (Oliver 1991), helps develop an overall understanding, and avoids wasting time on details that have little importance in the overall scheme.

The models of the existing and desired situations are used to guide the literature search and the findings are in turn used to refine the models. The exploratory literature search will reveal the many dimensions of the topic of interest. New topics and problems might be identified that were not originally anticipated, and it might become clear which factors have most influence. The results of this step are an Initial Reference Model and an Initial Impact Model, sufficiently detailed to determine a suitable research plan.

To determine whether a publication is relevant or not, we suggest the following steps for a quick read:

- Read the abstract.
- If the abstract is interesting, then read the introduction and the conclusions.
- If these are interesting and relevant, then read the results.
- If these are relevant read the background, objectives and setup.

This quick read is aimed at determining:

- What (what is the objective)?
- Why (why have the authors done so)?
- How (which research methods were used)?
- Results (what are the findings)?
- How good (what is the quality of the research)?

For details about reviewing the literature in depth, see Section 4.4.

Descartes (Ramon y Cajal 1999) advises researchers not to acknowledge as true anything that is not obvious. This is echoed by others who advise to (1) avoid the false concept that the most important problems are already solved (Ramon y Cajal 1999), and (2) never fully accept any hypothesis, theory, law or doctrine (Oliver 1991). Arhenius goes even further and claims that things that are already said to be impossible are the most important to pursue for the progress of science (Root-Bernstein 1989). It is important to learn to consult work in other languages as this broadens the horizon of knowledge.

The literature is used to check each assumed or experienced link in the models to see the extent to which these have been shown to exist, or can be expected to exist using the available evidence (see Section 2.4 for a description of the symbols and their use). Even if statements in the literature seem obvious, it is important to check whether they have a sound basis or are based on assumptions. Preference should be given to statements that are based on clear evidence, in particular to those that have a similar context as one's own area of interest (see also Section 4.4.2)

The literature is also checked for additional influencing factors and links not considered earlier. Factors that are relevant but fall outside the scope of the research, are represented as nodes with dashed lines or aggregated as 'other factors', as shown in Figure 3.8.

Reliability Example

Taking our example, it is not clear yet whether a research project is required, as the understanding and expectations are based solely on available information and the interest within the research group and the company. For instance, it is possible that the problem has already been solved or has been addressed by other researchers unknown to the research group or company.

An exploratory review of the literature on existing support and support proposals reveals several solutions to address reliability problems: mathematical methods for calculating reliability, descriptions of guidelines and methods for assessing and improving reliability, and evaluations of the use of various support. However, these solutions can only be applied when details of the system are known, not in the early stages; the research problem identified seems unresolved.

The next step is to verify the factors and links within the models and modify these where necessary using an exploratory review of the literature and further discussions with the stakeholder, to determine the kind of research that is necessary to solve the problem. This results in the following changes to the Initial Reference Model, as illustrated in Figure 3.8, showing the relevant part of the model:

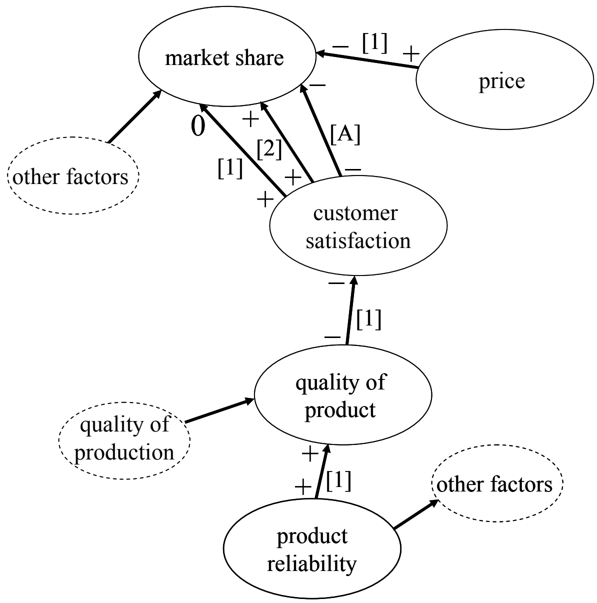


Figure 3.8 Part of the Initial Reference Model based on evidence from literature (all sources are ficticious)

- Earlier (Figure 3.2), it was assumed that low product reliability had a negative effect on the market share. However, no evidence is found for this direct link in the literature. Instead, according to one source, referred to as [1], a link *via* 'product quality' and 'customer satisfaction' exists. Two nodes are added accordingly.

- The statement in source [1] is that a reliable product was found to have a positive influence on how customers judged the quality of the product. Furthermore, a poor quality of the product resulted in customer dissatisfaction. The links in the Initial Reference Model are labelled accordingly.
- Source [1] found no clear evidence that customer satisfaction relates to market share, in contrast to source [2] where a positive effect was found. The latter seems more relevant for the research project, because the evidence found belongs to the same domain, namely machine systems. Note that choosing a link as most relevant, only because it confirms one's own assumptions, is not acceptable. Both links are added.
- Based on the strength and relevance of the evidence of [2], which only dealt with *high* customer satisfaction and its effect on market share, the researcher decides that it could be possible, that poor customer satisfaction has a detrimental effect on market share and adds this link as an *assumption*.
- Source [1] reveals further factors that influence market share, amongst which 'price' seems relevant, as 'maintenance cost' and the 'amount of profit' are influencing factors in the original model of the existing situation (see Figure 3.2).
- Other sources provide evidence that product reliability also affects other factors, and that 'quality of production' influences 'quality of the product', but all of these fall outside the scope of the project.

The partial Reference Model in Figure 3.8 shows a *complete* link based on evidence between 'reliability' and 'market share', *i.e.*, from the factor of interest to one of the factors that constitutes an important goal for the company. A complete link provides a strong basis for a research project. However, there is no further evidence yet about:

- the factors that affect product reliability, *i.e.*, what makes products (un)reliable;
- other, potentially more influential factors that affect the existing situation;
- the key assumptions in the Initial Impact Model (see Figure 3.7): does assessment in the early stages lead to earlier detection of problems and thus improves product reliability because improvements can be made in time. These assumptions (links) need verification if this is to be the basis for the support.

Resolving the above issues requires a detailed review of the literature. The researcher decides, however, that the current understanding and expectations, as represented in the Initial Reference and Impact Models, are adequate to determine the kind of research necessary to address the formulated research problem; the necessary detailed literature review will be the focus of the next stage, DS-I.

Reliability Example with Alternative Outcome

If the initial literature review would have not resulted in sources [1] and [2] but some other publications, such as [3, 4 and 5], the Initial Reference Model would have been different (see Figure 3.9) based on the following alternative understanding of the existing situation:

- No evidence is found about the link between product reliability and market share. The assumption remains as in Figure 3.2.
- One particularly detailed study [3] on reliability in the area of earth-moving equipment showed that low reliability causes high maintenance costs, and that high maintenance costs cause the warranty costs and the operating costs to increase.

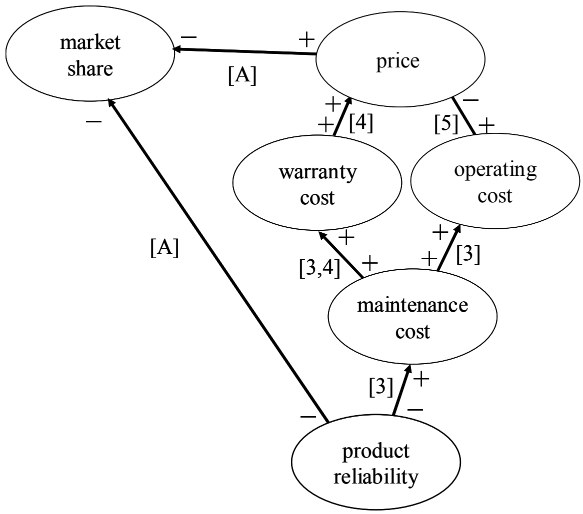


Figure 3.9 Alternative Partial Initial Reference Model

- A publication on warranty costs [4] confirmed the link with maintenance costs and found that warranty costs tend to be calculated into the price of the product, thus increasing the price.
- Publication [5] found a link between high operating costs and low price, but could not find a causal link: some products were sold at a low price to compensate for the high operating costs, other products had to have a low price for competitive reasons and the operating costs were made high, *e.g.*, through insurances or obligatory service contracts, to realise enough profit. In the Initial Reference Model a link without an arrow is added between ‘operating cost’ and ‘price’, to indicate that these factors are linked, but no cause could be established. The combined effect of the two factors ‘operating costs’ and ‘warranty costs’ on ‘price’ is not clear.
- It was assumed that ‘price’ affects ‘market share’, but no real evidence could be found related to mechanical systems.

The resulting Initial Reference Model, part of which is shown in Figure 3.9, shows an *incomplete* link between 'product reliability' and 'market share', both directly and indirectly (*via* price), in contrast to the model in Figure 3.8. This alternative model is too weak a basis to focus the research on developing a new method to support reliability assessment. If effects of 'product reliability' on 'market share' (taken as an important criterion for the stakeholder) are not clear, there is no reason to assume that a new method to assess product reliability could improve market share, even if it improved reliability. In this case, the kind of research needed is to obtain a better understanding through a Comprehensive DS-I.

A Note on Different Entry Points

In the example, the impulse for the RC stage was a problem experienced by a company and the wish of this company to focus on developing an effective tool. The research started with a prescriptive goal: the development of support. This is why early on a model of the desired situation could be drawn and some tentative criteria could be formulated (market share and profit). Nevertheless, as the example showed, obtaining a good understanding of the existing situation is crucial. Knowledge about the reasons for the problems experienced and about available support and support proposals helps identify whether there is indeed a need for developing a new support, and if so, which issues this support should address. Maybe a different type of solution than initially anticipated is required. As the example showed, the research plan might have to be changed, in this case, to include research to achieve the descriptive goal of improving the understanding before developing a support.

The impulse for doing research can start from a descriptive goal: improving our understanding of a particular situation, *e.g.*, the way in which requirements and solutions develop within a product development project. In this case, the starting point will be a preliminary model of the existing situation. We believe that most investigations are undertaken with a purpose based on the belief that the understanding gained can be used, ultimately, for addressing a particular problem. It is necessary to make this belief explicit; the initial model of the existing situation should therefore contain the links between the factors of interest and success factors. Only then is it possible to identify which improvements might be most effective and efficient and to develop a vision of the desired, improved situation.

A third impulse for doing research may be that support exists but the results are not known, but assumptions about its use and usefulness exist. The research starts with another type of descriptive goal, an evaluative goal: understanding the effects of a support through its evaluation. Information about the support, as well as its introduction, implementation, training, use, *etc.*, needs to be collected and a model of the desired situation (effects) has to be developed, if no Impact Model is already available. Developing an initial model of the existing situation without the support, if not already available, will show the problems the support was supposed to solve.

A fourth impulse for doing research may be that support exists but the results are unsatisfactory. The research starts with the descriptive goal: understanding the causes of the unsatisfactory results of the support. In this situation too, information about the support has to be collected and a model of the desired situation developed, if not already available. The RC stage focuses on the development of an initial model of the existing situation, *i.e.*, the unsatisfactory situation with the support.

In summary, irrespective of the original research goal, initial models of the existing situation and of the desired situation including the preliminary success factors are necessary to clarify understanding and expectations, and select the type of research.

3.4 Clarifying Criteria, Main Questions and Hypotheses

To determine the focus of the research, it is necessary to identify the criteria that are considered essential to determine whether the results help achieve the aim and to determine the main questions and hypotheses.

3.4.1 Criteria

To be able to judge the existing situation and suggest efficient and effective ways of improvement, our understanding needs to involve a link to success, *i.e.*, there should be a *complete link* between the factors that are of interest – the Key Factors – and the factors (Success Factors) the research project sets out to understand and/or influence as described in the research goal.

In Section 2.5 we defined Criteria as the desired values of the Success Factors. A distinction was made between Success Criteria and Measurable Success Criteria (as well as Success and Measurable Success Factors). Success Criteria relate to the ultimate goal to which the research project or programme intends to contribute and usually reveal the purpose of the research and the eventual, desired influence on practice. Measurable Success Criteria were introduced as those criteria that are linked to the chosen Success Criteria and that can be applied to judge the outcomes of the research given the available resources. They should serve as reliable proxies for the Success Criteria. As to which factors can act as Measurable Success Criteria depends to a large extent on the constraints of the project. The factors and criteria chosen in the RC stage are called *preliminary* to indicate their tentative nature at this stage.

In our example several potential Success Factors were mentioned: 'market share', 'amount of profit' and 'company image', for instance. A discussion with the company reveals that their interest is primarily in 'improving market share'. This focus seems acceptable from the understanding gained from the literature, and reflected in the Initial Reference Model in Figure 3.8, which showed a complete link between the preliminary Key Factor 'product reliability' and this preliminary Success Factor. The outcome of the alternative literature review, shown in Figure 3.9, illustrates that it might not be clear whether such a link exists until further research is undertaken.

The chosen preliminary Success Criterion 'increased market share', however, can only be assessed once the product is out in the market, which is outside the timeframe of the project. A factor directly linked to the preliminary Success Factor 'market share' is 'customer satisfaction'. Assessing the value of this factor requires at least a functioning prototype of the product. Assuming that this too will not be possible within the duration of the research project, ‘product quality’ is chosen as the preliminary Measurable Success Factor, and ‘high product quality’ as preliminary Measurable Success Criterion. This criterion needs further operationalisation in order to be able to be used, *i.e.*, the terms ‘high’ and ‘product quality’ have to be defined in such a way that this can be assessed within the project. For further details on formulating operational definitions, see Section 4.5.2.

Figure 3.10 shows the Initial Reference Model based on the literature sources of the example and the alternative example. The model includes the preliminary Key Factor and the preliminary Success and Measurable Success Factors. Note that this model is simplified. In a research project, a Reference Model is likely to be more complex and to have more than one Success and Measurable Success Criterion.

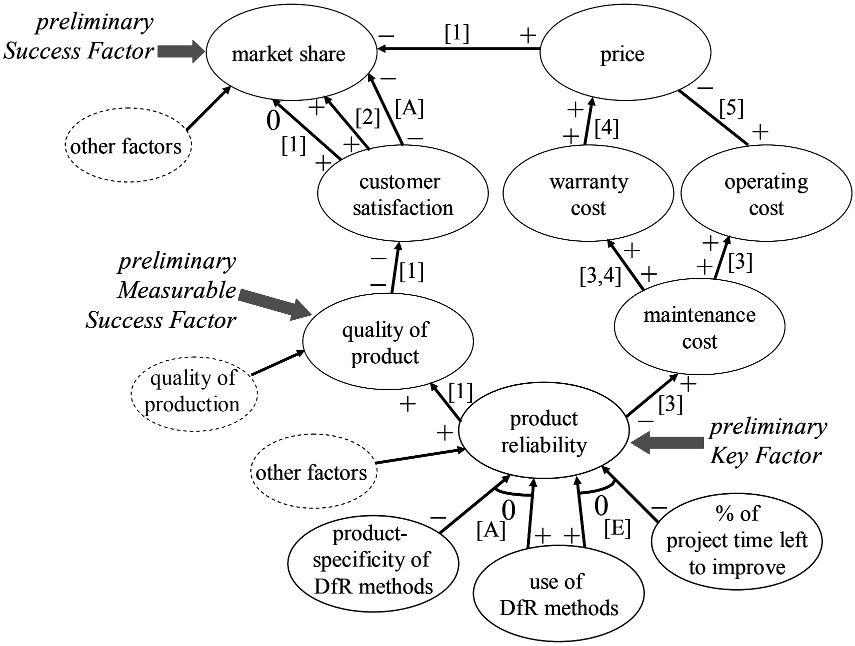


Figure 3.10 Initial Reference Model including preliminary criteria

Based on the Initial Reference Model and the expectations, the Initial Impact Model is updated, and the preliminary Key Factor and preliminary Success and Measurable Success Criteria added, see Figure 3.11.

As the research project progresses, Measurable Success Criteria become more precisely defined and may change. Success Criteria usually do not change. For instance, the support developed may introduce new influencing factors, which may require new Measurable Success Criteria, or the development of a support may prove to be more time consuming than expected. As a consequence less time is available for evaluation, and fewer or even different criteria may have to be chosen. When alternative Measurable Success Criteria are chosen, care should be taken that the corresponding Factors too are as closely linked as possible to the Success Factors.

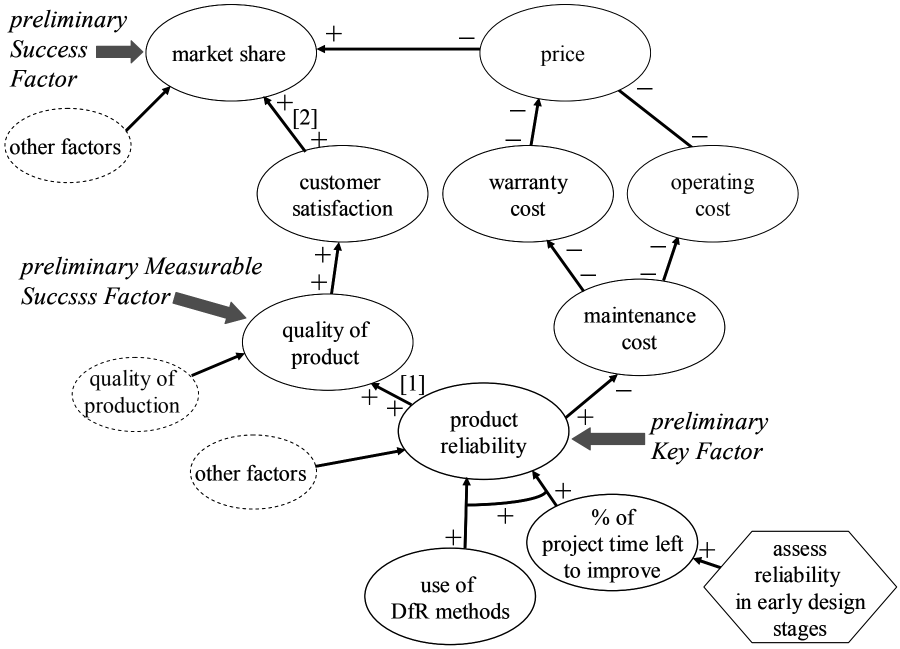


Figure 3.11 Initial Impact Model with preliminary Criteria

3.4.2 Research Questions and Hypotheses

We often observe that PhD students present their work describing what they are doing or planning to do, but fail to state the research questions and hypotheses behind their work. When research questions are formulated, they are often very vague, too encompassing to be answered within one PhD project, and concerned only with the support to be developed (*e.g.*, 'how to assess reliability'). To undertake research, the formulation of the main research questions and hypotheses early in the project is essential. During the course of the project, these will be refined and elaborated on to focus the various stages of the research process.

A *research question* is a question for which no answer exists yet. Research questions can be formulated in various ways, such as: What are the characteristics of a successful product? How often do designers..? How do designers do..? What are the ways in which..? How long does it take to..? When does…? Why is…? 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. In our example, some of the research questions would be: What causes the lack of product reliability? How does product reliability influence maintenance cost? How can we assess reliability in an early stage?

An *hypothesis* is a tentative answer to a research question in the form of a relationship between two or more concepts, or in our case, between two or more influencing factors, including the Success Factors. That is, an hypothesis is a claim or 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. In our example, an hypothesis behind the expected effect of the support is that 'If a lack of reliability is detected in an early design stage, sufficient project time is left to improve the product'. Every link can be formulated as an hypothesis, *e.g.*, 'An increase in product reliability will increase the quality of the product' (see Figure 3.11). Because this link is crucial in the Reference Model and based upon a reference from a different area of application, it would be important to investigate this hypothesis.

The main research questions and hypotheses can be derived from the research goal, the Initial Reference and Impact Models, and the related discussions. A detailed discussion on how to formulate research questions and hypotheses, so that they can be answered and verified, can be found in Section 4.5.2.

3.5 Selecting Type of Research

The next step is to identify the type of research suitable to answer the chosen research questions and verify the hypotheses. In Section 2.3, Figure 2.2, the seven main types of design research within the DRM framework were presented (reproduced here as Figure 3.12).

Research Clarification	Descriptive Study I	Prescriptive Study	Descriptive Study II
1. Review-based →	Comprehensive		
2. Review-based →	Comprehensive →	Initial	
3. Review-based →	Review-based →	Comprehensive →	Initial
4. Review-based →	Review-based →	Review-based → Initial/ Comprehensive	Comprehensive
5. Review-based →	Comprehensive →	Comprehensive →	Initial
6. Review-based →	Review-based →	Comprehensive →	Comprehensive
7. Review-based →	Comprehensive →	Comprehensive →	Comprehensive

Figure 3.12 Types of design research projects (iterations omitted)

As discussed in Section 2.3, a review-based study is based on the review of the literature on design or on design support only. A comprehensive study is a study in which the results are produced by the researcher, *i.e.*, an empirical study, the development of support, or the evaluation of support is undertaken by the researcher. A comprehensive study always includes a review-based study. An initial study closes the project and involves the first few steps of a particular stage to show the consequences of the results and prepare the results for use by others. Each research type is discussed in more detail in the following paragraphs.

Type 1. Comprehensive Study into Criteria

This type of project is undertaken when Success and Measurable Success Criteria are little understood, and therefore, a Comprehensive DS-I into understanding these criteria, their links and their relationships with the research problem is to be carried out. The outcome will be a better understanding of what constitutes success and which metrics can be used.

Type 2. Comprehensive Study of the Existing Situation

This type of study is undertaken when the criteria can be established, but a better understanding of the existing situation is necessary to identify the factors that are most relevant to address in order to improve this situation. A Comprehensive DS-I is necessary when the literature review reveals that understanding is:

- non-existent: the literature does not provide links between the factors of interest and the selected Success Factors;
- insufficient: the literature provides links but with insufficient detail; results are inconclusive or contradictory; evidence is based on a different context from the research; evidence is weak because of the small number of cases involved or the research methods applied;
- potentially incorrect: validity of the method(s) used is doubtful.

Once sufficient understanding is gained, an Initial PS is to be undertaken to indicate how this understanding can be used to improve design. This involves determining the factors that, when addressed, are most likely to have a large impact on success, and suggesting ways of addressing these factors.

Type 3. Development of Support

When the understanding of the existing situation obtained from the literature review and reasoning (Review-based DS-I) is sufficient to start the development of support a Comprehensive PS is undertaken if existing support is:

- non-existent: the literature, however, indicates or demonstrates the need to develop support to improve the existing situation;
- insufficient: the literature indicates or demonstrates that existing support is not used, does not work properly, only addresses part of the problem or is no longer effective or efficient in the context of new technologies, requirements and contexts.

The resulting Actual Support will be subject to an Initial DS-II for evaluation.

Type 4. Comprehensive Evaluation

In this case, support already exists. An evaluation of its application, however, is not available. A Comprehensive DS-II is undertaken to evaluate the support. The evaluation is based on a Review-based DS-I to understand the situation the support is intended to improve, and on a Review-based PS to understand the support and the expected effects (the desired situation). These reviews are necessary, because the support can fail due to incorrect assumptions or incorrect development. The evaluation can involve the comparison of multiple support. A Comprehensive DS-II may be necessary when current evaluations are

- non-existent: no formal evaluation of the application and effect on success of the support can be found in the literature;
- insufficient: earlier evaluations focused on Support or Application Evaluation, rather than Success Evaluation; the observed effects are unclear or contrary to expectations; earlier evaluation results were negative and the reasons are unknown;
- potentially incorrect: validity of the method(s) used is doubtful.

A Comprehensive DS-II is followed by suggestions for improvement (Initial PS), or further development (Comprehensive PS).

Type 5. Development of Support Based on a Comprehensive Study of the Existing Situation

A research project of this type is a combination of Types 2 and 3. The aim is to develop support, but the level of understanding of the existing situation is poor. Therefore, the research involves both the development of the understanding (Comprehensive DS-I) and, based on this, the development of support (Comprehensive PS). As with any comprehensive support development, this is followed by an Initial DS-II.

Type 6. Development of Support and Comprehensive Evaluation

A project of this type combines Types 3 and 4. The level of understanding of the existing situation obtained from the literature (Review-based DS-I) is sufficient to develop the support (Comprehensive PS), and the project resources allow formal evaluation of the support (Comprehensive DS-II). Depending on the results of the evaluation and the available resources, this is followed by a revisit of the PS or DS-I stage, either as an Initial study or a Comprehensive study.

Type 7. Complete Project

This is a project in which comprehensive studies are undertaken in each DRM stage. The RC stage will have shown that little has been done in the area of interest, yet indications are that the area has potential. As a result, research projects of this type involve; a comprehensive study of the existing situation (Comprehensive DS-I); development of support (Comprehensive PS); and a formal evaluation of this support (Comprehensive DS-II). This is followed by modifications to the support and understanding where necessary. In certain projects it may be required to start with a detailed investigation into criteria itself (Comprehensive DS-I as for Type 1). As carrying out all these stages in depth requires substantial time and resources, this type of research is more common for the work of a research group, unless a problem with a very specific scope is addressed.

Reliability Example

In our example, the researcher aims for Type 5. When the outcome of DS-I reveals a considerable lack of knowledge about reliability assessment, it might be necessary to change to Type 2.

3.6 Determining Areas of Relevance and Contribution

In our example, the initial literature review (see Section 3.3) focused specifically on reliability, and revealed details about its various dimensions. The aim was to use this understanding to develop the Initial Reference and Impact Models and a preliminary set of criteria, in order to determine the research topic and to select the type of research (Sections 3.3 to 3.5). In other words, the aim was to *identify* the research problem.

In order to *solve* the research problem, however, this initial review is not sufficient. The literature needs to be looked into in more detail, considering all potentially relevant areas, not only those related to the research topic or one's own discipline. It is important to consider a wide range of areas and disciplines. To draw a comparison with product development; no company will develop something without looking for what exists and for interesting ideas in other products. Other disciplines and areas might have undertaken interesting studies, might have developed interesting theories, methods, concepts, solutions, *etc*., that could be relevant, if looked at carefully in an analogical way. For example, for analysing the icons used in user interfaces and for developing support to develop intuitive user interfaces, Hurtienne used linguistic theories about metaphors (Hurtienne *et al.* 2008). Sometimes the research methods in other disciplines can be very interesting for design research. Breakthroughs in research often emerge at the intersection of areas or where knowledge, ideas and methods have been transferred from one area to another.

In our reliability example, the literature on the following topics can also be relevant: robust design, tolerancing, ageing, wear, information exchange (between service/maintenance and designers), life-cycle costs, maintenance, product liability design methodology, conceptual design, Design-for-Manufacturing, Design-for-Assembly, design thinking, human–computer interfaces, *etc*.

To avoid getting lost in this 'jungle' of the literature it is useful to ask the following questions:

- What are the areas that could be related to the topic in question?
- How directly relevant are these to the topic: which ones seem essential, which ones useful and which ones might be useful?
- In which of these areas is the researcher's contribution likely to be? This area or these areas should be compatible with the researcher's expertise, as well as with the goals of the project.

To help represent the answers to these questions, we developed the **Areas of Relevance and Contribution diagram**[8] (**ARC diagram**). This representation clarifies the foundation on which the research is to be based and the area(s) of contribution of the research. Students have found this diagram very useful for structuring their literature search, for structuring the literature chapter in their publications, for presenting, discussing and reflecting on the areas they consider relevant, and to clarify the area of their contribution.

For developing an ARC diagram, we suggest the following steps.

1. Draw an oval (or any other form) carrying the research title, goal, topic, or main research question.
    - Separate diagrams can be drawn to address additional (sub)goals, topics and research questions, or – as done here – one diagram can be drawn to cover all of the research.
    - Additional, more specialised diagrams can be drawn in each DRM stage to focus on the questions addressed in that particular stage.

2. Draw areas around this central oval, labelled with those disciplines that could be relevant for the research topic in providing possible theories, models, background information, existing methods, results of empirical studies, *etc.*. Figure 3.13 shows an example, based on the diagram developed by one of the PhD students attending our Summer School.
    - Analyse every word in the main oval as well as those in the Initial Reference and Impact Models, in the formulations of the research problem, and in the questions and hypotheses in order to identify relevant disciplines.

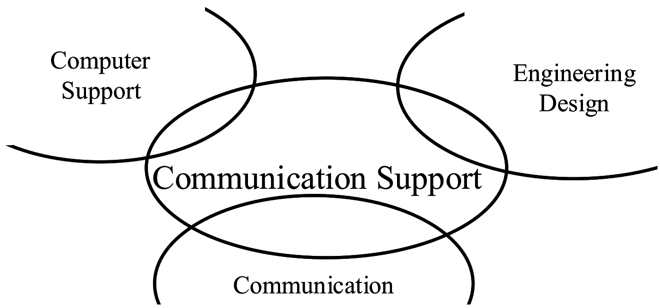


Figure 3.13 Example: Second step of setting up the ARC diagram for the project on 'Analysis and support for communication throughout the design process' (these and the following figures are adapted from the diagram of Thomas Flanagan, Summer School participant, unpublished, with permission).

[8] In earlier publications, we called this diagram the Theoretical Foundation and Contribution (TFC) model.

- Be as specific as possible if you are familiar with a discipline, *e.g.*, thermodynamics, rather than physics, or cognitive psychology, rather than psychology.

3. Identify the specific areas or topics within these disciplines that seem relevant, and put these in or around the discipline areas (see Figure 3.14).

   - Sub-areas can be represented using smaller ovals or circles connected to the related main area, but further away from the centre.
   - Rearrange the areas such that clusters of areas can be identified easily.
   - Try to be as informed as possible. In order to identify areas in a particular discipline other than one's own discipline, it is useful to look at: websites; handbooks, or lecturing materials of key institutes in the discipline (student editions with high edition numbers are likely to provide commonly accepted descriptions of main research areas and concepts); book series; refereed journals.
   - Consider a broad range, but be selective in the final choice: everything could potentially relate to everything.

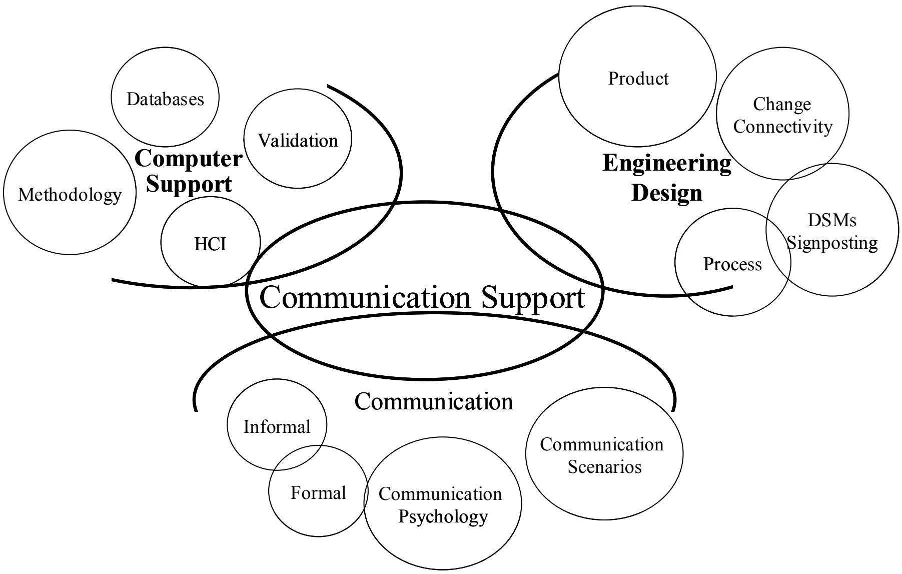


Figure 3.14 Example: Third step of setting up the ARC diagram for the project on 'Analysis and support for communication throughout the design process'

4. Indicate which of the areas seem *most relevant* to your work. Distinguish between essential areas and useful areas, *e.g.*, by colouring or hatching as in Figure 3.15.

   - As to which areas are most relevant may change during the course of the project. Similarly, new areas and disciplines may have to be added as understanding increases, and existing ones may have to be removed if they lose their relevance.

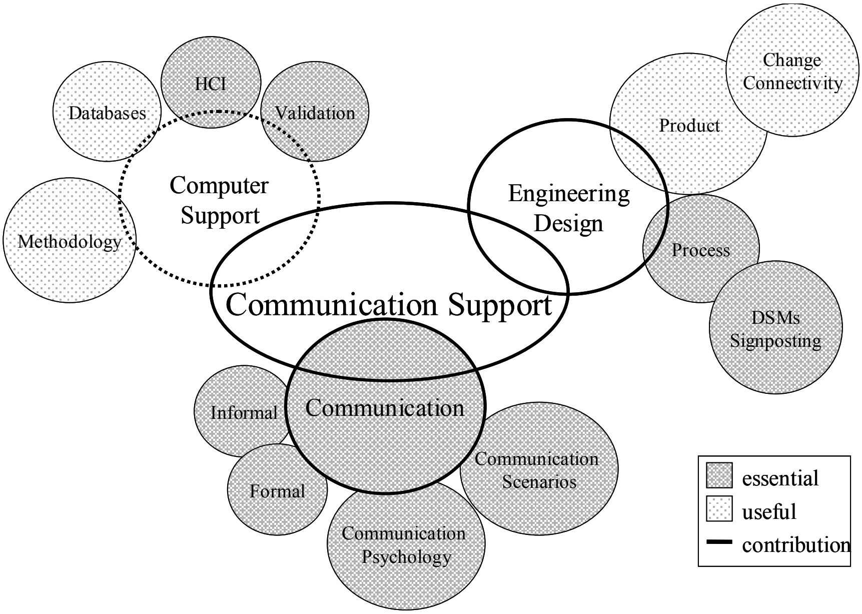


Figure 3.15 Example: Final ARC diagram for the project on 'Analysis and support for communication throughout the design process'

5. Indicate the area(s) to which the research project will *contribute*, *i.e.*, the area(s) in which the research is expected to make the biggest changes. Highlight these areas by, for example, thickening or colouring their borders, as in Figure 3.15.

   - Theories, models, findings and methods from various areas will be relevant, but it is possible to contribute to only few of them. For example, in a project on capturing rationale using a computer-aided design (CAD) system, it might be essential to look into the area of databases, and to use this as a basis for developing the support. However, the contribution will be in the area of engineering design, not in computer science; a new database might have been developed, but probably not a new database concept.

Other graphical representations can be used than the one presented here, such as MindMap™.

3.7 Formulating Overall Research Plan

The final step of the RC stage is the formulation of the Overall Research Plan for the project.

3.7.1 Overall Research Plan

An Overall Research Plan should include the following:

- research focus and goals;
- research problems, main research questions and hypotheses;
- relevant areas to be consulted;
- approach (type of research, main stages and methods);
- expected (area of) contribution and expected deliverables;
- time schedule.

The deliverables are the intended outcomes from the various stages of the research type chosen, listed in Section 2.6.

It is important to note that an initial plan is better than no plan at all; it provides a direction for research, a yardstick for measuring progress, and a sense of achievement to carry on beyond this stage.

The time available for a research project will be constrained by the possible duration of the project and the number of people involved. Since a detailed plan requires knowledge of the specific research questions to be answered, the plan cannot be made very concrete at this stage: the questions to be answered in a particular stage depend on the outcome of the previous stage. However, the results of the RC stage as proposed thus far in this chapter, provide a reasonable indication of the scope of the project, its main stages, and the type of research methods required to address the research problem.

The following chapters provide more information about methods available for carrying out each stage. It is useful to read these chapters before drawing up the research plan. The types of research method give an indication of the required resources and help develop a fairly realistic overall time schedule for realising the research goals. The plan should be monitored, modified and refined on a continuous basis as understanding increases during a research project and unforeseen circumstances and outcomes can occur.

A possible way of representing the aims or questions and hypotheses of the research project against the stages was developed by one of the PhD students who attended our summer school. He aimed to answer the following research questions (Eriksson 2007):

- What noises and enablers are there in product development decision making that effect project performance, and how? (Q1)
- How does decision maturity effect project performance? (Q2)
- How can the decision-making process be supported to continuously increase project performance? (Q3)
- What additional noises and enablers are there when product development projects are distributed and how do they affect project performance? (Q4)

- How can a distributed decision-making process be supported to continuously increase project performance? (Q5)

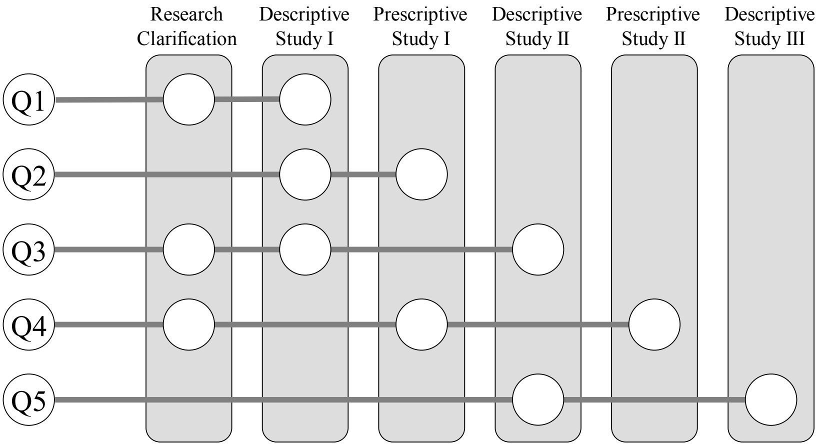


Figure 3.16 Aims against stages, after Eriksson (2007)

3.7.2 Visualisation Exercise

A particular problem we noticed is the ability to communicate the ideas about the deliverables. Even in an early stage, most researchers will have an image of what they want to achieve. However, this often remains implicit, making it difficult to judge the viewpoint and assumptions underlying the identified problem and research plan. The so-called Visualisation Exercise, proposed by Mogens M. Andreasen, provides a very good technique to make ideas and views explicit by visualising the concepts or outcomes mentioned in the research plan, using models, schemes, scenarios, and other graphical means, and to reflect on and discuss the result. In this early stage, it is useful to consider variants of these visualisations to avoid fixation on an initial idea. This exercise should be used in all stages of the research project. Note that the Impact Model shows the impact of the outcome, but not the intended outcome (the support) itself. The visualisation exercise is intended to show the intended outcome.

An *example* illustrates this exercise. A student planned to develop a methodology specifically for designing sports equipment. However, it was not clear from his description how sports equipment and its design is different from other products and design processes and hence requires a special methodology. He was therefore asked the following two questions in order to visualise his assumptions and views:

- Mention a typical sports equipment that has built-in medical, ergonomic, anthropometric, *etc*., knowledge, as this is what you said differentiated sports equipment from other products.

- Try to model how these peculiarities have influenced project strategy, team manning, process (special plans or activities), criteria, organisation, *etc*., in ways that are different from traditional ways.

The visualisation did not show the differences with other design processes. It became clear that there was not sufficient understanding about the typical types of knowledge required for developing sports equipment, and where and how these types of knowledge and the domains involved influence the process. The conclusion that a new methodology was required, seemed premature, and the research focus should be on gaining this understanding.

Other examples of visualisation requests that we have given to students are:

- Show an example of the content of your model of what you call ‘product assortment dispositions’, and show in a scenario how this understanding is used: by whom, for what, in what situation.
- Illustrate a family of products, their commonality and their variations. Illustrate how you can reason ‘cross-family’ concerning manufacturability and life-cycle costs. What are actually life-cycle costs? What insight about the product life do you need to have in front of you?
- Show an example of data merging from different domains and how this would tell about what you call ‘maturity and assembly capability’. If possible, show what more is known if higher maturity is achieved.
- Your guideline on patient safety will contain some type of system model that shall instruct, motivate and orientate the guideline user. This model may become the essential part of your research. Show us such a guideline!
- The ‘form development process’ of automobiles may be seen as a combination of an industrial design process focusing upon appearance (controlled by quality of form, reliability of form, *etc.*) and other processes, mainly technical. Make an activity model showing these parallel processes, to show what goes on in the industrial design processes and the other processes, and to show their goals or criteria. Show where in this model your framework shall operate.

As these examples show, the visualisation questions focus on the terminology and concepts used. The true content of these often remain implicit. Having ‘more knowledge’ about maturity, *e.g.*, does not show what sort of knowledge this entails, or at which level of detail. Similarly, a ‘guideline’ may be very generic or very specific. In this stage it is not possible to exactly know the outcomes, but there has to be a vision about the sort of outcome that is expected, knowing very well that this vision might not be correct. Making it explicit at least allows discussion about the vision.

3.7.3 Reflection on RC

Before the RC stage is completed, the following checklist may be useful to reflect on the deliverables of this stage:

- Why do you ask this research question?

- Why do you believe this is a relevant research question?
- Why do you believe that you have or can obtain the competences to answer the research questions and solve, or contribute to solving, the identified problem?
- Where do you believe that you can be original, *i.e.*, your results can bring a contribution to practice as well as to knowledge?
- Why do you believe your Overall Research Plan leads to a result?
- Why do you believe your work is scientific?

Oliver (1991) provides some heuristics of how to recognise an important contribution to science. One is that the contribution will significantly change the ways of thinking or working of others in the area. A researcher can recognise whether their contribution is important, amongst others, by checking if their results not only evaluate favourably to their own data but conforms well to other well-known information, or even more, if it can relate data previously unfamiliar to the researcher. The more diverse and numerous the compatible but previously unknown data is, the greater the chances are of the discovery being a major one.

3.8 General Guidelines on Doing Research

Many authors stress that researchers should have independent judgement and that they should be optimistic. They should nurture concentration – a sustained orientation of all faculties toward a single object of study (Ramon y Cajal 1999). A researcher needs to have devotion to truth and a passion for reputation for being able to discover the truth (Oliver 1991; Ramon y Cajal 1999). As research can often be a long and lonely activity fraught with failures; having enthusiasm for the work and an ability to "enjoy the struggle, not the spoils" is thus essential (Oliver 1991). Patience and observational abilities are often critical in scientific inquiry (Ramon y Cajal 1999).

Acts of creation, which includes research, are opportunistic in nature (Dasgupta 1994). It is important to be flexible (all within limits, of course) and opportunistic, that is, to have the courage to pursue promising, unexpected avenues opened up during research that may lead to exciting new solutions, even though they may not necessarily fit into the existing research goal and plans (Dasgupta 1994). The creative agent is not only knowledge rich, but is able to wander freely about the knowledge space and retrieve whatever seems to relate to the goal at hand (Dasgupta 1994).

Serendipity – accidental discovery – is commonplace in science, but only to those "whose minds are prepared for it" (Pasteur in Root-Bernstein (1989)), *i.e.*, those who show curiosity and perception. Serendipity can be encouraged in the following ways.

- Searching, assimilating and using a wide range of knowledge: creation, at least in the sciences, is knowledge intensive (Dasgupta 1994) and as we discussed it is important to look into, not just the publications of direct interest for the work at hand, but also others.

- Making and recording expected and unexpected observations (Lenox 1985).
- By maintaining flexibility in thinking and interpretation (Lenox 1985). Adams (1993) enlists various kinds of mental blocks that typically prevent us from thinking flexibly, and approaches of how these could be avoided.

Notwithstanding the importance of serendipity: large insights are composed of a possibly intricate but describable network of small steps (Dasgupta 1994). Doing research means working hard and meticulously, rather than waiting for the big moment of inspiration to arrive.

While it may be immensely beneficial to "learn from the masters" (Truesdell 1984), it is important not to be over-impressed by the work of predecessors (Oliver 1991; Ramon y Cajal 1999); this does not mean being disrespectful to other researchers, but being not too reverent to their work. It is important never to believe an hypothesis, law or principle completely, especially because this can lead to indoctrination, and blindfold one from the truth (Oliver 1991). One should beware of pursuing sophistication for its own sake, as this is both a distraction and a waste (Oliver 1991). This is not to say sophistication should not be pursued, but that it should be pursued only when required by the overall purpose of the enquiry.

It is particularly important to have a strong inclination toward originality (Ramon y Cajal 1999); alternative explanations should always be entertained, and evaluated against the yardstick of observation. Often it is useful to think like a child, and to force oneself to see things in a different light (Oliver 1991; Root-Bernstein 1989). Innovators and discoverers often reason by analogy (Oliver 1991). Especially when probing into the unknown, this can be very powerful. An hypothesis need not have its origin in facts or observations, although it eventually has to be validated by these. Speculation and subjective thinking for generating ideas should not be discouraged by over-critical annihilation of initial, bright ideas. "It is often more difficult to identify what is right in an idea than what is wrong in it", and "one must see the important features of an imperfect idea rather than totally discard it" (Bligh 1990).

There are a number of caveats that are often advised to be avoided. One is the excessive use of jargon (Oliver 1991), which substantially hinders communication, especially in interdisciplinary areas like design research. Furthermore, "There is no limit to what you can accomplish if someone else gets the credit" (Oliver 1991): passing credit freely to whomever and wherever it is due, is both polite and essential in areas like scientific enquiry where knowledge is built by successive addition. Pretence should be avoided at all costs as it is dishonourable and sooner or later it will be detected if the matter is of sufficient interest. If multiple alternative explanations exist, 'Occum's razor' – the principle that the simplest explanation is the best – should only be used when all observations on the matter are considered (Oliver 1991). Returning from time to time to the basic principles and laws in a particular area, is one way of staying on course.

Tunnel vision, *i.e.*, remaining fixated on a single solution or explanation, should be avoided by considering a range of alternatives and evaluating these. As Pauling suggests (N.N. 1977) "Just have lots of ideas, and throw away the bad ones". The ability to speculate is particularly important: "Be as bold in the conception of the hypothesis as rigorous in their demonstration" (Darwin and Richter in Root-

Bernstein (1989)). Perhaps while suggesting answers, one should not be too reasonable. Freeman Dyson, quoted in Root-Bernstein (1989), claims that "for any speculation which does not at first glance look crazy, there is no hope". However, this craziness should not be haziness (Root-Bernstein 1989). Therefore, the wilder the idea, the better they must be anchored by the accepted methods (Monod 1969). As Fermi, quoted in Root-Bernstein (1989), suggests: "only those guesses should be followed which define the answer to a problem, even if non-specific".

3.9 Main Points

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

- This chapter provides an approach and methods to support the early stage of a design research project or programme: to identify and refine a research problem, and set up an overall plan for carrying out research for solving this problem.
- This stage has six, iterative steps: identifying the overall topic(s) of interest, clarifying the current understanding and expectations, clarifying criteria main questions and hypotheses, selecting type of research, determining areas of relevance and contribution, and formulating the Overall Research Plan .
- There are three central issues that constitute an overall topic of interest: issue of interest, activity in or stage of the design process, and area of application. The topic may come from researchers, sponsors, research community or a combination, based on the belief this topic has an effect on design practice, although concrete evidence may not exist.
- In design research, the goal is to identify and solve problems of interest that have a degree of generality and application across products and practices. For a topic to qualify as a research area, it should be academically and practically worthwhile, as well as realistic.
- It is important to gather available information about the topic through an exploratory the literature review, and to make the expectations, beliefs and underlying assumptions of each stakeholder explicit, in order to obtain a first shared picture of the existing and desired situations and the Success Criteria. These are documented in the Initial Reference and Impact Models.
- The Initial Reference and Impact Models will indicate the focus of the research, whether this has sufficient research potential, what type of research would be suitable, and the criteria against which the research outcomes should be judged.
- Different entry points are possible. Research could start with: a descriptive goal of understanding a situation, a prescriptive goal of support development, or a descriptive goal of support evaluation.
- Irrespective of the entry points, initial models of both the existing situation (the Initial Reference Model) and of the desired situation (the Initial Impact Model) including the preliminary Success Criteria are necessary to clarify understanding and expectations and select the type of research.

- The exploratory literature review will lead to extension and adaptation of the Initial Reference and Impact Models, which were mainly based on beliefs and expectations. It is particularly important to find supporting evidence for the central assumption that applying support will have the desired effect on practice. There should be a link between the factors of interest (Key Factors) and success (represented by the Success Factors).
- Success Criteria usually remain static throughout a project; however, Measurable Success Criteria often do not. As a research project progresses, Measurable Success Criteria become more precisely defined and may change.
- Determining criteria and topic is an iterative process. Topic and criteria may be redefined several times before they are clear and well connected.
- Research questions and hypotheses can be derived from the Initial Reference and Impact Models. A research question is a question for which no answer exists yet. An hypothesis is a tentative, refutable answer to a research question in the form of a relationship between two or more concepts – or factors, in our definition.
- To determine the type of research to be undertaken, *i.e.*, on which DRM stages the research should focus, depends on the current state of the research. If for a particular stage, results are available, a review of the state-of-the-art is sufficient. If not, a comprehensive study is required, where results are substantially the researcher's own findings.
- Based on combinations of these possibilities, seven types of research are identified covering individual research projects as well as research programmes: four of these types focus comprehensively on one DRM stage only. The rest focus comprehensively on two or more stages.
- In the research project it is necessary to consider all potentially relevant areas, not only those related to one's own topic or discipline. It is important to consider a wide range of areas and disciplines.
- To identify areas of relevance and of contribution, these questions can be asked: what areas relate to the topic, how relevant these are (essential or useful), and in which of these the researcher's contribution is most likely to be. The area of contribution should be compatible with the researcher's expertise and project goals.
- A so-called ARC diagram helps represent the answers to these questions and is a good basis for discussion and reflection.
- An Overall Research Plan should include following: research focus and goals; research problems, main research questions and hypotheses; relevant areas to be consulted; approach (type of research, main stages and methods); expected (area of) contribution and expected deliverables; and time schedule.
- Since a detailed research plan needs knowledge of the exact research questions, the initial Overall Research Plan created in the RC stage cannot be very concrete. However, it still provides a direction for research, a yardstick for measuring progress, and a sense of achievement to carry on beyond this stage.

- The RC stage provides a good indication of the scope of the project, its main stages, and the type of research methods needed to address the research problem. The plan should be monitored and refined continuously as understanding increases.

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.

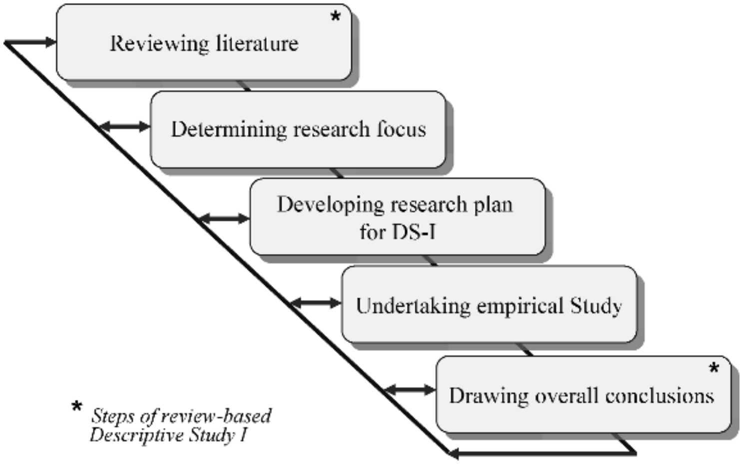


Figure 4.1 Main steps in a Comprehensive DS-I, stars (*) indicating the steps in a Review-based DS-I

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.

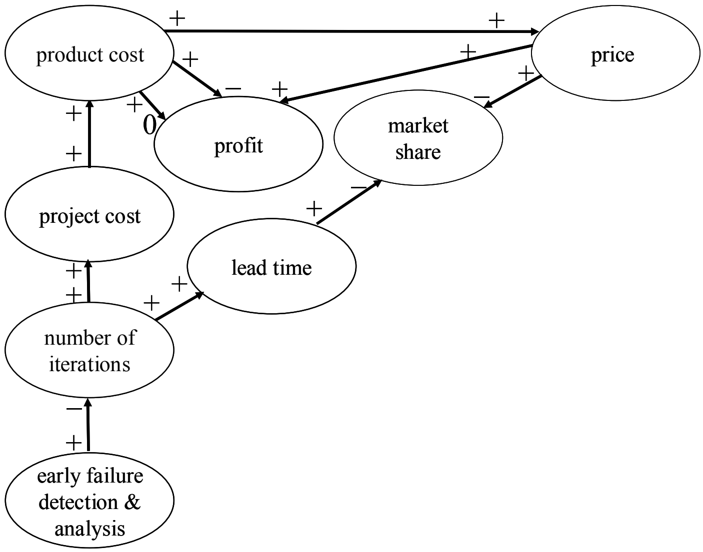


Figure 4.2 Partial Initial Reference Model related to the effects of failure detection and analysis in an early stage

- 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.

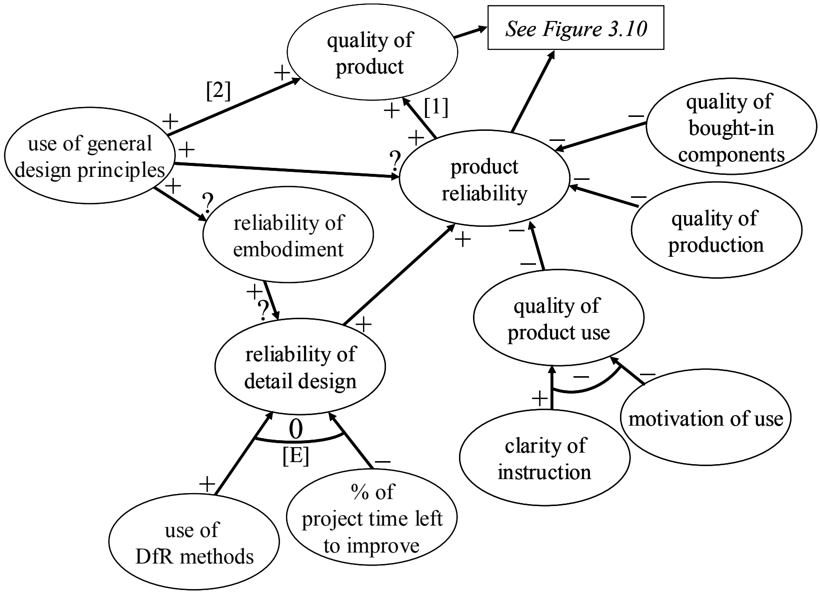


Figure 4.3 Partial Initial Reference Model based on some of the findings in the literature related to reliability

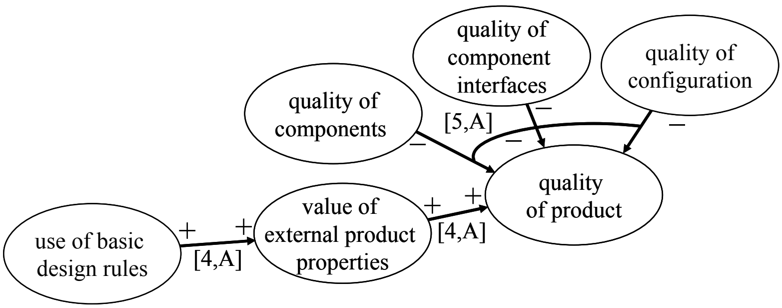


Figure 4.4 Partial Initial Reference Model based on some of the findings related to design rules

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.

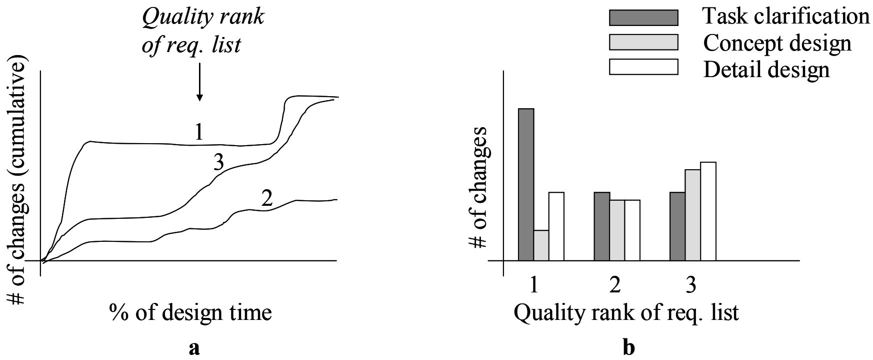


Figure 4.5 Different possible ways of presenting the results, affecting the research questions to be asked and research methods to be chosen

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.

	Interview users		Interview designers		Reliability assessment excercise		Analyse reliability documentation	
	Daily users, not buyers		Different companies Only experienced		Ask experts Use definitions Existing designs		Historical development and reliability	
Research Question 1	√√		√					
	R	P	R	P				
Research Question 2	√				√√		√	
	R	P			R	PP	RR	
Hypothesis 1	√√							
	R	PP						
Hypothesis 2			√				√√	
			R	P			RR	P

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:

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

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.

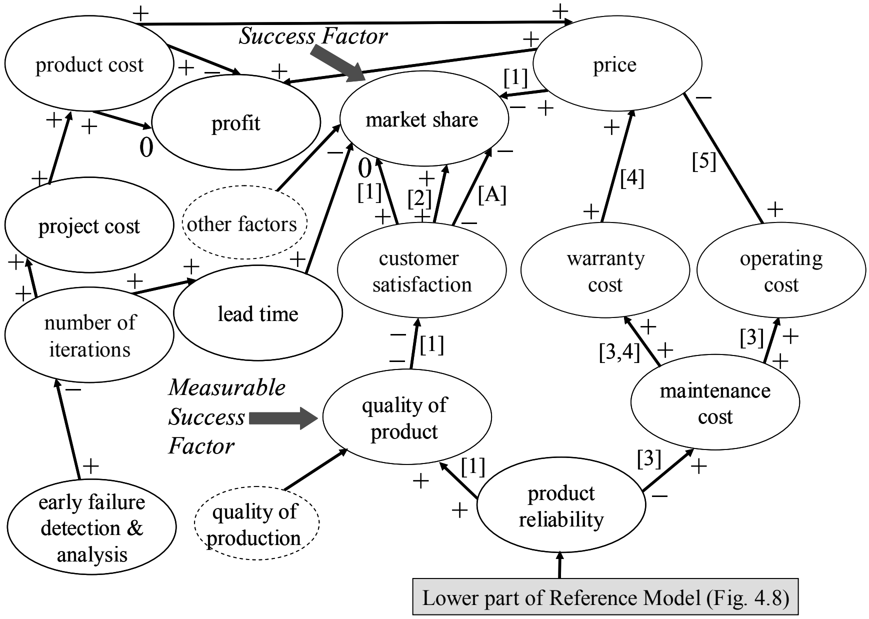


Figure 4.7 Upper part of the Reference Model resulting from the DS-I stage

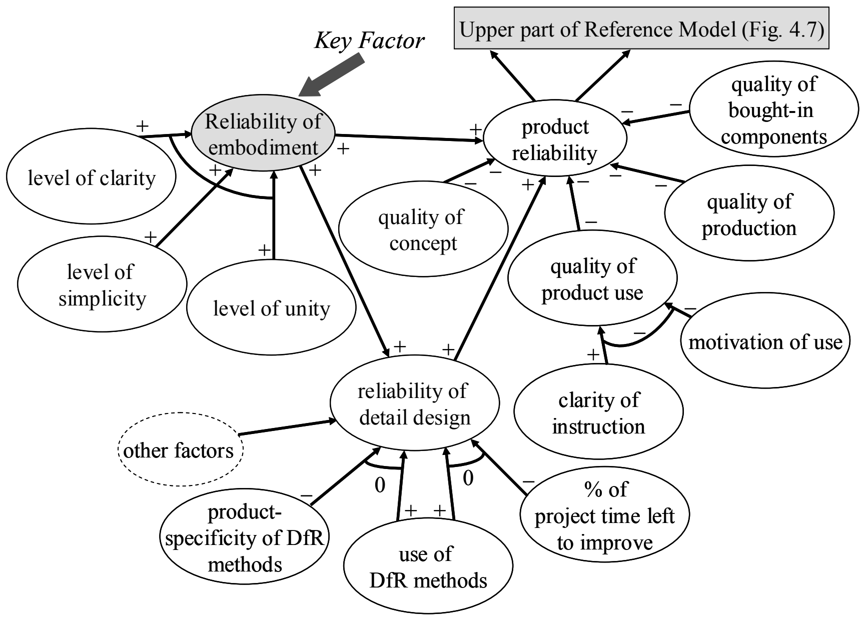


Figure 4.8 Lower part of the Reference Model resulting from DS-I

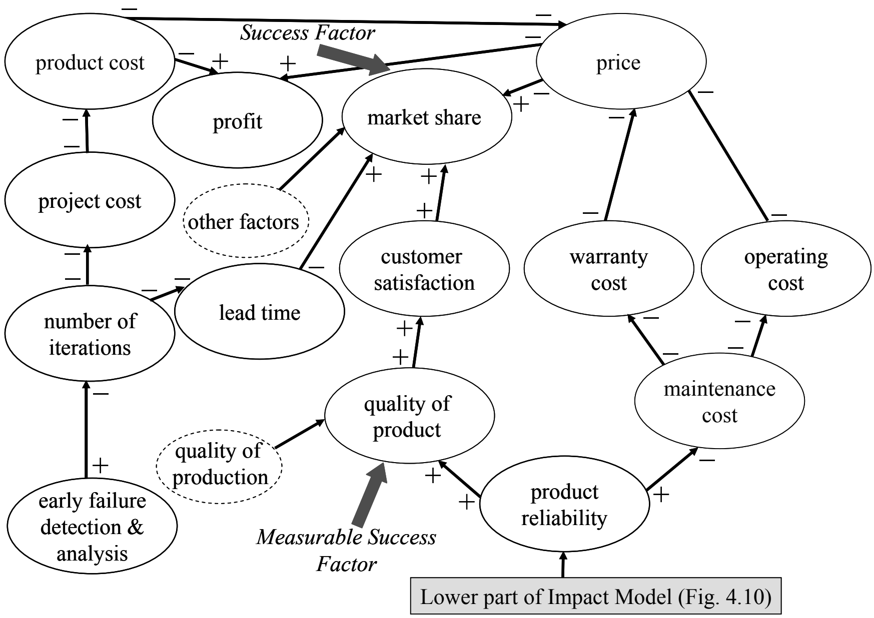


Figure 4.9 Upper part of the updated Initial Impact Model resulting from DS-I

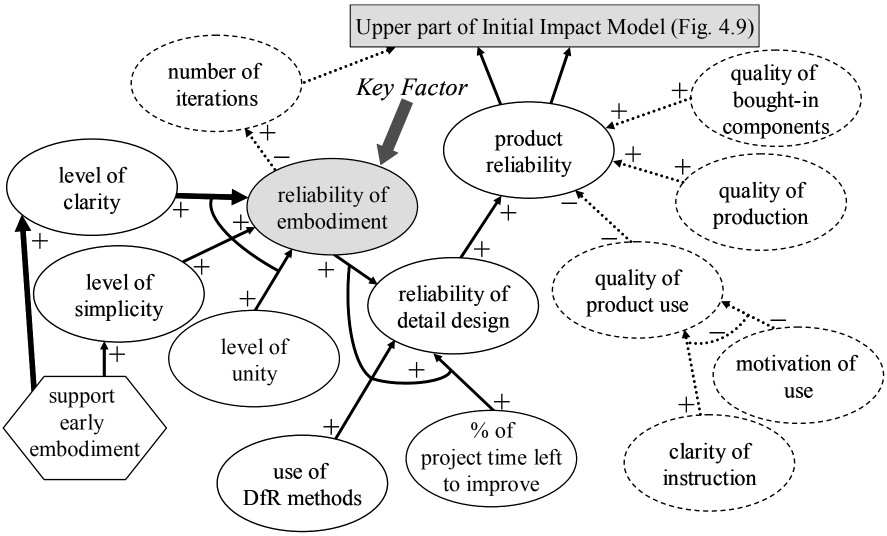


Figure 4.10 Lower part of the Initial Impact Model resulting from DS-I, focusing on improving the levels of clarity and simplicity, emphasising clarity as one of the main factors

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.

5

Prescriptive Study: Developing Design Support

The focus of this chapter is on the third stage of DRM: the PS stage. It discusses how, starting with the results from DS-I or DS-II, one can proceed to develop a design support (*i.e.*, knowledge, guidelines, checklist, methods, tools, *etc.*) in order to enhance, eliminate or reduce the influence of some of the critical factors found in DS-I or DS-II. Such a PS is a purposeful activity with the resulting support or its concept as the end product, and is, therefore, a design task in itself. The support is usually not a direct derivative of the findings from DS-I or DS-II. Creativity and imagination is required to develop effective and efficient design support. For this, a number of generic problem solving and development methods can be used.

The objectives of the PS stage are:

- to use the understanding obtained in DS-I or DS-II to determine the most suitable factors to be addressed in PS (the Key Factors) in order to improve the existing situation;
- to develop an Impact Model, based on the Reference Model and the Initial Impact Model, describing the desired, improved situation that is expected as a consequence of addressing the selected Key Factors using the support;
- to select the part of the Impact Model to address and to determine the related Success and Measurable Success Criteria;
- to develop support – the Intended Support – that addresses the Key Factors in a systematic way and to realise this to such a level of detail that an evaluation of its effects can take place against the Measurable Success Criteria;
- to evaluate the Actual Support with respect to its in-built functionality, consistency, *etc.*, – the Support Evaluation – in order to determine whether or not to proceed to DS-II to evaluate the effects of the support;
- to develop an Outline Evaluation Plan to be used as a starting point for the evaluation in DS-II.

The deliverables of the PS stage are:

- Documentation of the Intended Support consisting of:

  - Intended Support Description: what it is and how it works;
  - Intended Introduction Plan: how to introduce, install, customise, use and maintain the support as well as organisational, technical and infrastructural pre-requisites;
  - Intended Impact Model;

- Actual Support: workbook, checklist, software, *etc*.;
- documentation of the Actual Support:

  - Actual Support Description;
  - Actual Introduction Plan;
  - Actual Impact Model;

- Results of the Support Evaluation;
- Outline Evaluation Plan.

Note that an Initial PS (see Section 5.2) covers the first deliverable only, *i.e.*, the documentation of the Intended Support.

Section 5.1 provides examples of the most common types of design support. Section 5.2 introduces the three types of PS distinguished in DRM. The process we propose for a Systematic PS is described in Section 5.3. Each of the steps is discussed in Sections 5.4 to 5.8. Appendices B.1 and B.2 provide further details about available product and software development methodologies, and enlist a variety of specific methods and tools from these methodologies to help support developers to clarify tasks, formulate requirements, and generate and evaluate ideas. Appendix B.3 is a special section on guidelines for user-interface design and B.4 provides a checklist to summarise the support.

5.1 Types of Design Support

As noted in Chapter 2, design support includes all possible means, aids and measures that can be used to improve design. These are *prescriptions* – suggesting ways by which design tasks *should* be carried out – and include strategies, methodologies, procedures, methods, techniques, software tools, guidelines, knowledge bases, workbooks, *etc*. Below, we define and provide examples of some of the common types of support found in design research literature.

By a *design approach* or *methodology***,** we mean an overall framework for doing design. Common examples are the design methodologies proposed by Pahl and Beitz (Clausing 1994; Pahl and Beitz 2007) VDI 2221 (VDI 1993), and Total Quality Management (Clausing 1994). Newer, but less established work includes methodologies for specific types of products, such as mechatronics (Möhringer 2005).

By *design methods* we mean sequences of activities to be followed in order to improve particular stages of the design process (task clarification, conceptual design, detail design, *etc*.), and specific tasks within these stages (*e.g*., generation, evaluation, *etc*.). For an overview of design methods, see Adams (1993); Cross

(1994); French (1985); Jones (1970); Pahl and Beitz (2007); Roozenburg and Eekels (1995). For a list of these methods, see Appendix B.

Design guidelines are rules, principles and heuristics that are useful to follow in attaining some design objectives. Examples include the principles of order, embodiment, eurhythmy, symmetry, property and economy as applied to architecture by Vitruvius (circa 56 AD) (Vitruvius 1960), the conceptual design principles suggested by French (1985), principles of design embodiment for simplicity, clarity and safety proposed by Pahl and Beitz (2007), the many Design-for-X sets of guidelines, such as Design-for-Manufacturing, Design-for-Cost or Design-for-Environment guidelines, and the TRIZ principles proposed by Altschuller (1984).

By *design tools* we mean hardware and software for supporting design, based on some design approach, method or set of guidelines. The design tool supports the effective and efficient use of the approach, method or guideline. Sometimes, their use would not be possible without a computer tool. A large variety of computer tools is available, in particular for the embodiment and detail design stages. Design tools focus on certain design objectives, certain design stages or activities, or certain types of products. Some require intensive human–computer interaction; others execute design tasks almost without human intervention. Some tools can be made company or product specific through customisation, others are generic. Examples of design tools include CAD tools, Product Data Management tools, Finite Element tools, process planning tools, requirement capture tools, Life-Cycle-Assessment tools, and software for the design of specific products.

Design support can address any of the facets of design shown in Figure 1.1. This thus covers a spectrum as diverse as the supports mentioned above, as well as communication tools for distributed design, project management tools, procedures for introducing methods, suggestions for new organisational structures, and proposals for governmental strategies.

5.2 Types of PS

Three types of PS are distinguished in DRM, as shown in Figure 3.12 representing the different types of design research:

- Initial PS;
- Comprehensive PS;
- Review-based PS.

Any type of PS will start with Task Clarification, *i.e.*, a review of the research goals, the Reference Model, the (Initial) Impact Model, and of the literature on existing means with similar goals.

An **Initial PS** results in a description of the Intended Support. It therefore covers only the first two steps of the Systematic PS process discussed in the next section and shown in Figure 5.1: Task Clarification and Conceptualisation. As a consequence, the support cannot be formally evaluated: only a concept exists. An Initial PS is often carried out when time and resource constraints do not allow a Comprehensive PS but thoughts about possible or improved support based on the results of the previous stage are required to round off the research. As discussed in Section 3.5, we consider an Initial PS necessary to illustrate how the findings of DS-I could be used to improve design, even if the focus of the research is on gaining understanding and not on the development of support (Design Research Type 2 in Figure 3.12). For example, a study on how users experience interaction should not only result in a verified framework (theory) describing user experience and methods to assess this, but also in a set of recommendations based on this knowledge for the development of interactive systems (see, *e.g.*, (Mahlke 2008)). When a research project includes a comprehensive evaluation (DS-II) of a support (Design Research Types 4, 6 and 7 in Figure 3.12), an Initial PS is to be undertaken to show the consequences of the outcome of the evaluation on the support and on the Impact Model.

A **Comprehensive PS** results in a support that is realised to such an extent that its core functionality can be evaluated for its potential to fulfil the purpose for which it was developed. All the steps of the Systematic PS process discussed in the next section and shown in Figure 5.1 are covered: Task Clarification, Conceptualisation, Elaboration, Realisation and Support Evaluation. How comprehensive the PS stage has to be, *i.e.*, how complete the support has to be depends on what is required and what is possible within the constraints of time, resources and environment, for the intended evaluation (DS-II). This depends, amongst other things, on whether a designer or the researcher has to be able to use it, and whether it should be used in the real, intended environment or whether a test environment would be sufficient. The support developed does not have to be a fully realised system; a demonstrator or a prototype system is often sufficient to evaluate at least the core functionality.

A **Review-based PS** is necessary in projects that focus on the evaluation of existing support (DS-II) that has been developed without the researcher being involved (Design Research Type 4 in Figure 3.12). All steps in the Systematic PS process are carried out in order to reconstruct, or develop when not available, the input needed for DS-II. If the intended purpose, the underlying concept, the expected impact and available functionality of the support is not known, a proper evaluation is not possible. A Review-based PS is based on a review of the literature and of the documentation about the support and on discussions with users and developers, where possible. In other words, a Review-based PS is a Comprehensive PS, but carried out with the help of the existing literature and knowledge sources on already available support to be able to formally evaluate the support.

5.3 A Systematic PS Process

There are two areas that can provide general guidance to developing support: product development methodologies and software development methodologies. A summary of the various methodologies can be found in Appendix B. Interestingly, no methodology for developing methods and guidelines (heuristics) could be found.

The product and software development methodologies divide the process into a number of stages, in each of which the product or software under development becomes more detailed. In each stage the basic problem-solving cycle (see below) is applied, although not all methodologies make this explicit. This combination of gradual development and problem-solving cycle are seen as the key elements to managing complexity and enhancing success in product development. The process, it is always emphasised, is not linear: many iterations take place within and between the stages.

The most commonly used stages are:

- task clarification: specification of problem and requirements;
- conceptual design: concept development;
- embodiment design: giving overall shape to the concept;
- detailed design: finalising the details of the embodiment and prototyping;
- testing is sometimes included in these stages, sometimes mentioned as a separate stage.

The basic problem-solving cycle, which takes place in each of the stages, essentially has the following four steps:

- establish need or clarify the problem to be solved;
- generate potential solutions for fulfilling the need;
- evaluate solutions by comparing them with each other and against the problem or need;
- decide if a suitable solution is found; if not, return to the first or second step, depending on the results.

Based on the existing methodologies, we have developed a Systematic PS process with five main steps to aid support development. It is important to note that – as in any methodology – iterations between steps will and have to take place, as the development of support is a continuous process of generation and evaluation, frequenting between various levels of abstraction.

An important feature of the proposed process is the distinction we make between Intended Support and Actual Support. The **Intended Support** is a *description* of the *complete* support as envisaged by the researcher. The **Actual Support** is a *realisation* of the Intended Support that may cover only a part of functionality of the intended and may be implemented in a different way, but can still be used as a proof-of-concept for the purpose of evaluation. The focus of the Actual Support should be on the core contribution of the research project, *i.e.*, the core functionality of the Intended Support (see discussion in Section 5.7.1).

In order to be able to develop the Actual Support, it is necessary to start developing an Outline Evaluation Plan, as the intended evaluation will determine the comprehensiveness of the Actual Support. Hence, a second feature of the proposed Systematic PS process is the parallel development of the support and the Outline Evaluation Plan: the DS-II stage starts during the PS stage.

A third important feature of the proposed Systematic PS process is its emphasis on developing a strong concept and generating variants before detailing the concept. This implies that the type of implementation (paper-based, software, *etc.*) is not selected until the end of the conceptualisation, that is, until it is clear what type of support is most suitable for achieving the goal of evaluating the concept. When software is to be developed, this implies that the concept is developed on paper before actual programming is initiated. Although this seems obvious, considering the recommendations of design methodologies, too often we have seen research projects setting out to develop software to address a particular issue, without considering or investigating whether software is the best solution: maybe a workbook or checklist would be sufficient. Similarly, we have rarely seen researchers consciously generating variants (even in those cases in which the tool itself was intended to help generate variants to support designers). Developing support should follow the same best practice recommended for developing products and software. Many aspects of design methodologies can be applied for support development. The process we present here blends the characteristics of several of these methodologies.

Note that while we primarily focus on the process aspects for carrying out support development, gathering domain knowledge plays a crucial role in design research in general, and support development in particular. This was emphasised earlier in Chapter 3, when determining the scope of research: specifying the domain in which the results are applicable is necessary, and must be kept in mind all through the research process. The domain provides the content of the support and the context for use and application and thus has an influence on the focus of each DRM stage: where data is gathered; in what context the support is to be used; what functionality is to be included; which knowledge to incorporate; in which context to evaluate the support, *etc*. In Appendix B.2.5, some methods for knowledge acquisition are described.

The steps of the Systematic PS process are shown in Figure 5.1.

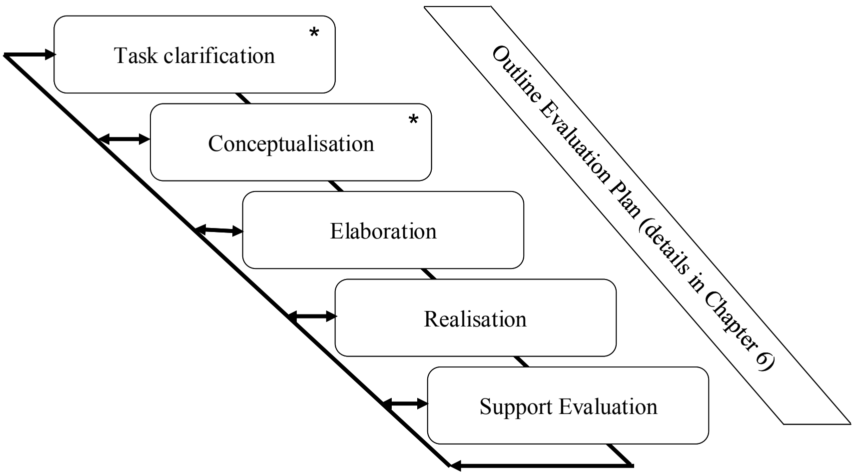


Figure 5.1 Main steps in the PSstage; stars (*) indicate steps of an Initial PS

1. Task clarification (details in Section 5.4):

   - The results from earlier stages and earlier projects, such as goals, findings from DS-I or DS-II, the Reference Model, the Initial Impact

Model and the associated Criteria are gathered. A full Impact Model might be available if the PS stage follows a DS-II stage.
- Using the Initial Impact Model as the preliminary focus, the literature on existing support with similar goals, *i.e.*, with similar expected impact, is reviewed in order to identify the extent to which current support fulfils the goals, and where scope exists for developing support.
- With this information, alternative Key Factors, Measurable Success Criteria and relevant links are identified from the Initial Impact Model to create Impact Model alternatives for the Intended Support. These are compared and evaluated to create the Impact Model representing the desired effects of the Intended Support (**Intended Impact Model**).
- The Intended Impact Model, the project goals and the results from DS-I or DS-II are used to formulate a list of requirements for the support, covering the whole life cycle.

2. Conceptualisation (see Section 5.5):
   - The functions and sub-functions of the Intended Support are identified.
   - Concepts of how to fulfil these functions are proposed, evaluated, compared and selected.
   - Based on the selected Intended Support concept, a concept plan for introducing the Support (**Intended Introduction Plan**) is developed.
   - The Intended Impact Model is updated to take into account the consequences of the chosen concept and the Intended Introduction Plan.
3. Elaboration (see Section 5.6): This and the next step are carried out only in Review-based and Comprehensive PS:
   - (in the case of tool development) Existing literature on technologies with which the Intended Support could be realised is reviewed.
   - The Intended Support is fully described.
   - The Intended Introduction Plan is elaborated.
   - The Intended Impact Model is finalised.
4. Realisation (see Section 5.7):
   - The core functionalities of the Intended Support are identified and an Outline Evaluation Plan is generated (see Chapter 6).
   - (in the case of tool development) The literature on development platforms and technologies suitable for the realisation of the *Actual* Support are reviewed (note that these are not necessarily the same as for the Intended Support).
   - The Actual Support is developed to such an extent as to enable the evaluation of the core functionalities with the available resources.
   - An **Actual Introduction Plan** is developed.
   - An **Actual Impact Model** is created to reflect the limitations and particular features of the Actual Support and the Actual Introduction Plan.

5. Support Evaluation (Section 5.8):
    - The Actual Support is evaluated for completeness, internal consistency, *etc*, and modified if necessary.
    - Based on the results of the Support Evaluation and the modifications to the Actual Support, the Outline Evaluation Plan is updated for use in DS-II (Chapter 6).

5.4 Task Clarification

In developing a support, the first stage, Task Clarification, is to establish the problem to be solved by the support, to clarify its requirements and to better define the desired situation. The support must be such that its requirements reflect the goal of the project, and help attain the changes envisaged by the difference between the current situation (as represented by the Reference Model) and the desired situation (as envisaged in the Impact Model). The changes proposed must be novel and significant in their impact on the Success Criteria.

The starting point is the results from earlier stages or earlier projects, such as the goals, findings from DS-I, the Reference and Initial Impact Models and the associated Criteria.

When the PS stage, whether Initial or Comprehensive (see Section 5.2), follows the evaluation of existing support (DS-II), the input to the Task Clarification also includes the documentation of the support, the findings from the evaluation and the Impact Model used earlier. In most cases, the required input will be available.

A Review-based PS, however, focuses on obtaining information about the *existing* support that is to be evaluated in DS-II (see Section 5.2). This is required because the support is not developed by the researcher him or herself and nothing more than the Actual Support and some documentation may be available. The task of the researcher is to gather as much data as possible from users and developers to reconstruct the missing input, in particular the Impact Model, to have sufficient understanding of the support to be able to develop an Evaluation Plan and prepare a DS-II that is in line with its original goals and its functionalities.

The description in this chapter will focus on a Comprehensive PS, as the process for an Initial or Review-based PS can be easily derived from this. In situations where the latter processes deviate strongly from that of a Comprehensive PS, this will be mentioned explicitly.

The focus of the literature review at this stage is primarily on existing support with similar goals, *i.e.*, aiming at a similar impact on the Measurable or Success Criteria or trying to address similar factors. The relevance of the literature is guided by the Initial Impact Model. The aim is to identify the extent to which current support fulfils parts of the Initial Impact Model, where scope exists for developing new support that can have a significant impact, as well as to find ideas, fragments of concepts, implementations and technologies that could be used for developing the support. Note that the primary focus of this literature review is to help clarify the problem and requirements for the support to be developed, and to better specify the desired situation.

Typically, a review of the literature at this stage investigates each relevant, existing support for its scope, functionalities, application area, underlying concept, assumptions behind the implementation, implementation technologies and evaluation results, if available. The "Support Outline: Summarising Scope and Assumptions" (Appendix B.4) can be useful here. The checklist is intended to help researchers document the scope and assumptions underlying the support they are developing, but can equally well be used to describe existing support.

The literature review is used to verify the Initial Impact Model. There will be interesting alternative sets of Key Factors that could be addressed. These alternative sets will be linked to the Success Criteria in different ways and might have a different impact. Furthermore, taking into account the context within which the support should be used, and the intent and drivers behind the support can also lead to alternative Key Factors and Criteria. The result is a set of alternative Impact Models. By exploring in each model the links between Key Factors and Success Criteria, looking for factors as close as possible to the Success Criteria and yet measurable within the project, potential Measurable Success Criteria for each alternative Impact Model can be identified.

The alternative Impact Models are compared and evaluated to select the most promising Impact Model for the Intended Support, the *Intended Impact Model*. The choice of the Key Factors, Measurable Success Criteria and links will be governed by their potential for scientific contribution (academically worthwhile) and the envisaged strengths of the impact of addressing the Key Factors on the Success Criteria (practically worthwhile).

The Intended Impact Model can differ from the Initial Impact Model as well as the Reference Model in various ways:

- additional nodes and links: *e.g.*, the support intended to train novice designers to exhibit expert behaviour is expected to encourage novices to exhibit a more inquisitive behaviour (a new node and related links) asking certain questions they otherwise would not have asked;
- removed nodes and links: *e.g.*, novices are expected to no longer make certain mistakes they made when they did not use the support;
- modified links, *i.e.*, links with changed signs: *e.g.*, novices who use the support less often ask typical 'novice' questions.

Most of the links in the Impact Model are assumptions. Many of the links are modifications from existing links in the Reference Model or are introduced as an anticipated consequence of the introduction of the support to be developed. Even the links that are understood well in the existing situation might not be the same in the new desired situation because of unexpected side-effects of the introduction of the support. Each link in an Impact Model should be provided with argumentation about why the effect is expected, based on the literature, assumptions and reasoning.

The Intended Impact Model describes the effects of the Intended Support, the requirements are not known yet. To start with, a *problem statement* is formulated, describing the core problem addressed by the support. Then, *a list of requirements* to be satisfied by the support is formulated, taking into account the documentation and information gathered thus far. The list should cover the entire life of the support, including its implementation, testing, installation, introduction, use and maintenance phases, and keeping in mind organisational, technical and other contextual pre-requisites. Many of these requirements cannot be formulated yet or formulated precisely; the list will become more complete and detailed as the PS stage progresses. At this stage, it is important not to presume a particular type of support: the problem statement and the requirements list should be as solution-neutral as possible. Appendix B1.1 lists several existing requirements identification and evaluation techniques.

At this stage it is useful to start the process of documenting the scope and assumptions of the Intended Support using the already mentioned Checklist for Summarising Scope and Assumptions.

The outcome of the Task Clarification is an Intended Impact Model, a problem statement and a list of requirements for the Intended Support.

Reliability Example

To clarify the development task in our reliability example, the Reference and Initial Impact Models resulting from DS-I, shown in Figures 4.7 to 4.10, are analysed. Some of the main conclusions are:

- Increased ‘market share’ has been taken as the Success Criterion, because the company involved had identified market share as their primary concern. The suggested Key Factor has been ‘reliability of embodiment’.
- Primary findings from DS-I (see Figure 4.8) indicate that previous failures in improving reliability at the detailed design stage could be due to poor reliability of embodiments, and that the use of general rules for embodiment (clarity, simplicity and unity) adequately demonstrated their impact on improving product reliability. In particular, the ‘level of clarity’ and the ‘level of simplicity’ were found to have a significant link with ‘product reliability’ as long as the ‘level of unity’ (assessed in terms of strength of components of a product) remained adequate. The Initial Impact Model at the end of DS-I therefore suggests support to determine ‘level of clarity’ and ‘level of simplicity’ only, with ‘level of clarity’ as the most influential factor.
- From other available evidence, it was seen that ‘early failure detection and analysis’ (see Figure 4.9) has additional impact on ‘market share’ *via* ‘lead time’. The Initial Impact Model from DS-I therefore suggested to focus on improving reliability of embodiments, which will involve earlier analyses of failures than in detailed design.
- That the three factors ‘levels of clarity’ *etc.*, can be assessed early in embodiment, was also the result of investigations in DS-I. Improvement of ‘reliability of embodiment’ by means of providing ‘support for early embodiment’ was therefore chosen as the target for the support to be developed in PS.

In order to identify existing support, if any, for aiding improvement of clarity and simplicity, a literature review is undertaken, but no such support found other than very generic guidelines and examples. Sufficient support exists for determining the 'level of unity'. However, the link between the level of unity and the levels of clarity and simplicity is not addressed at all, although the levels of clarity, unity and simplicity together influence the reliability of the embodiment; the level of unity must remain adequate for clarity and simplicity levels to become effective. It was therefore decided to revise the earlier suggestion and to address reliability of early embodiment by means of *all three factors*: level of clarity, of simplicity *and* of unity. The main reason for not choosing these as Key Factors is that the levels of clarity, unity and simplicity *together* influence the reliability of the embodiment: addressing each one separately will not have the same effect.

The focus of the PS-stage is thus on supporting the assessment of all three levels as early as possible during the embodiment stage, the combination of these into one measure of reliability of embodiment, and the improvement of the reliability of the embodiment. This results in the updated Intended Impact Model shown in Figure 5.2. The factors and links shown with dashed lines are outside the scope of the PS as discussed in Section 4.8.2. Note that the choice of any other set of Key Factors to be influenced by the support, *e.g.*, 'use of DfR methods', or 'quality of product use', would have led to a different, alternative Intended Impact Model.

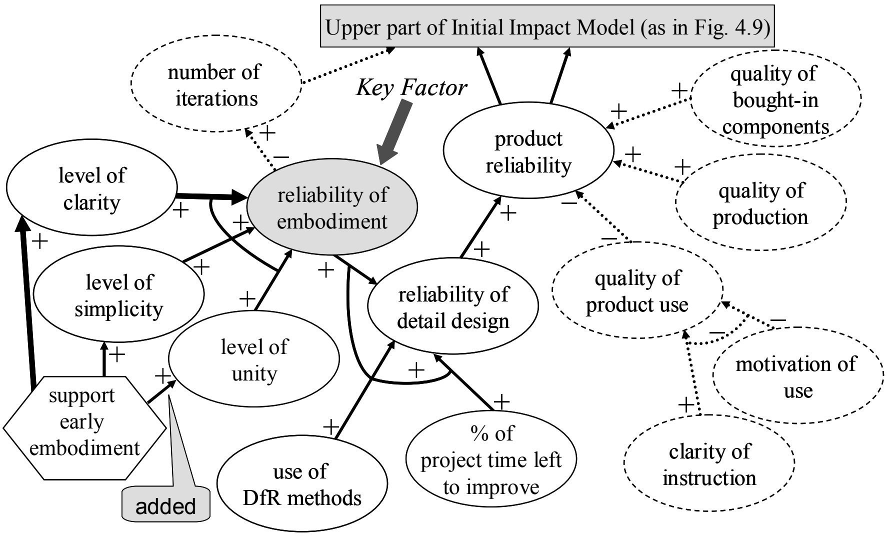


Figure 5.2 Lower part of the Intended Impact Model after Task Clarification in PS

Based on all above information, a list of requirements for the support is formulated. A partial list is given in Table 5.1, which is based on the checklist of Roozenburg and Eekels (1995). As it is not certain whether it is possible to suggest modifications on the basis of the assessed levels of clarity, simplicity and unity, the related requirement is listed as a wish, rather than a demand.

Table 5.1 Partial requirements list for the support related to the Intended Impact Model in Figure 5.2 (*D* = demand, *W* = wish)

Problem statement:	
Develop a support to help experienced mechanical designers improve reliability of early design embodiments	
Performance:	
D	The support should help assess the levels of clarity, simplicity and unity.
D	The support should help assess the reliability of the embodiment.
W	The support should suggest modifications of the embodiment to improve these levels.
D	The support should be able to support the design of mechanical systems
W	The support should be able to support the design of electromechanical systems
D	The support should be able to use the information available in engineering drawings or CAD models as input.
W	The assessment should be fast enough to be used as a regular activity within the design process
Ergonomics:	
D	The support should be usable by individual experienced mechanical designers
D	The support should be easy to introduce
D	The support should be easy to learn
D	The support should be easy to use
W	The support should be easy to maintain
D	The support should be easy to install
Cost:	
D	The support should cost less than XX to buy
W	The support should cost less than YY to maintain
Introduction:	
D	It should be possible to use the support in conjunction with existing support available in mechanical design offices
W	No additional hardware or software will be necessary
Life:	
W	The support, with maintenance, should have indefinite life
Disposal:	
W	If the support involves hardware and consumables, these should be limited and environmentally benign.

5.5 Conceptualisation

5.5.1 Determining Main Functions

The first task in Conceptualisation is to identify and decide which functions the support needs to have in order to affect the Key Factors in the intended way. As the example below will show, this process will lead to further elaboration of the Intended Impact Model. A function may address a particular factor or link, or a composite of factors, and new functions may lead to new factors and links.

Reliability Example

In the reliability example, the problem statement is to develop a support to help experienced mechanical designers improve reliability of early design embodiments.

The *first* potential function is the assessment of the levels of clarity, simplicity and unity of early embodiments (henceforth abbreviated as CSU). Assessing levels of CSU does not directly influence the levels of CSU itself, but the designers' knowledge of these levels. This leads to the introduction of new influencing factors in the Intended Impact Model; the factors referring to the knowledge of the designers are added. The increased level of knowledge does not improve the levels of CSU either, unless modifications are made and these are of good quality. Hence, 'quality of improvement' is added as a factor.

Knowing the levels of CSU can have an impact on the 'quality of modification' by helping to determine which modifications could be relevant. However, this is not sufficient. The designer needs to have sufficient knowledge to come up with ways of modifying the embodiments. To support this, it is decided to add a *second* function to the system: supporting the modification of the embodiment based on the knowledge of the CSU levels. Improving the quality of the modifications should improve the levels of CSU, which in turn should influence the reliability of these embodiments. Furthermore, designers need to have knowledge of the trade-offs between the levels of CSU and their combined effect on reliability. This gives rise to a potential *third* function: determining the reliability of the support based on the assessed levels of CSU. Depending on which of the functions is included in the support, alternative concepts can be generated:

- to support the assessment of C, S or U only, or a combination of two of those (which was considered in the Initial Impact Model and rejected);
- to support assessment of all three levels, shown by the smallest hexagon 'early assessment of C,S,U' in Figure 5.3, which should improve the level of knowledge of the designers;
- to support assessment and modification, shown as the medium-sized hexagon in Figure 5.3, which should not only improve the level of knowledge of the designers, but also the 'quality of modification';
- to support assessment of CSU, modification of embodiment and assessment of reliability, shown as the largest sized hexagon in Figure 5.3, which should have the most direct influence on the chosen Key Factor: reliability of embodiment.

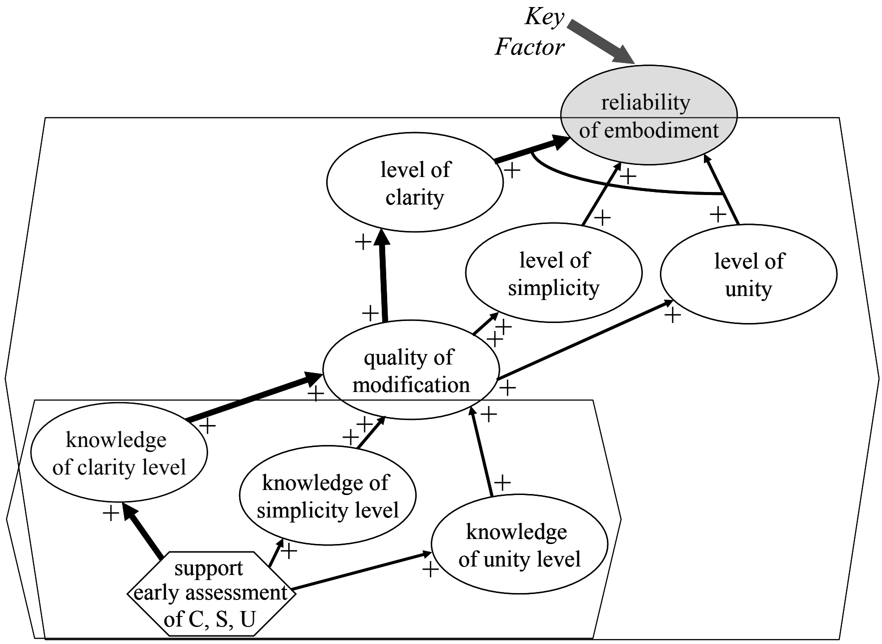


Figure 5.3 Three alternative, superimposed, Intended Impact Models (partial models) representing three different Intended Support concepts covering alternative sets of functionalities, indicated by hexagon shapes

In the example, the researcher decides to only support assessment of the levels of clarity, unity and simplicity, leaving modification and assessment of reliability to be carried out by the designer. The resulting Intended Impact Model is shown in Figure 5.4 (the rest of the model remains as shown in Figure 5.2).

Choice of Functions

A nice abstract view of how to define the realistic and relevant functions of the Intended Support is given in Figure 5.5, which was developed for diagnostic research (Verschuren 1997). The large circle in the figure indicates *possible* future scenarios. The smaller circle indicates the *likely* futures, *i.e.*, scenarios that will happen without the introduction of the support. The oval represents the scenarios of the *desirable* future. The support needs to focus on the scenarios that are possible and desirable but not likely to happen (the dark area in the figure). Scenarios that are desired but likely to happen need not be addressed with the support, as they will happen anyway. Taking our reliability example, the researcher decides that it is *desirable* to have a measure of reliability of embodiments. Preliminary studies suggest that knowing the levels of CSU are an important pre-requisite to determine reliability, but there is no existing support. The researcher decides that it is currently *not possible* to determine reliability from the levels of CSU without the help of the designer, but that it should be *possible* for the support to assess the levels of CSU. It is *unlikely* that these levels can be determined without support.

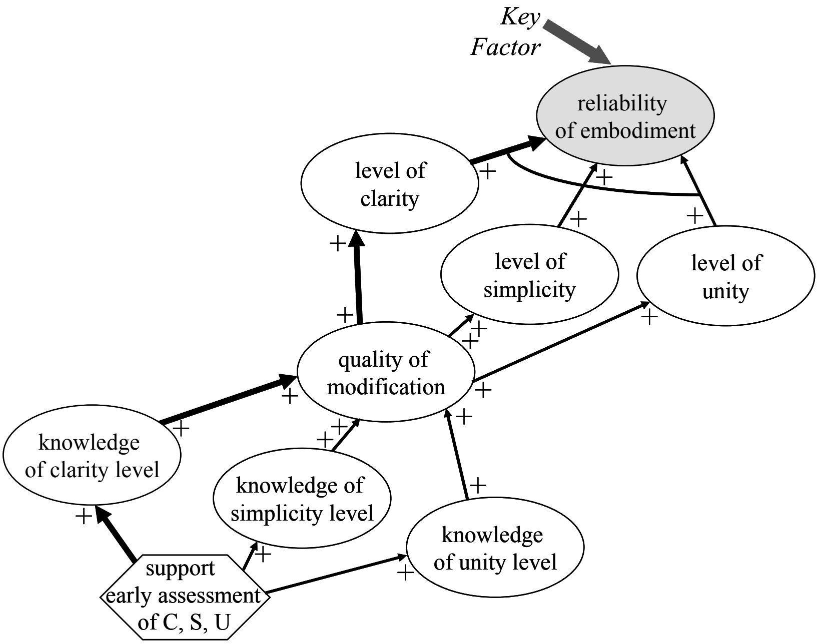


Figure 5.4 Partial Intended Impact Model reflecting the chosen focus of the Intended Support in the reliability example

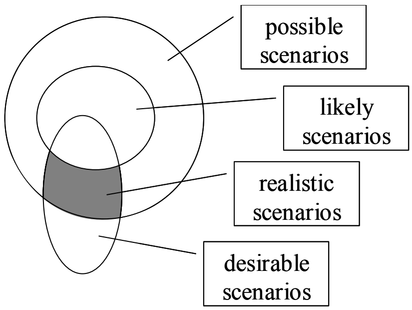


Figure 5.5 Schematic view of the area in which the Intended Support should functions, after Verschuren (1997)

5.5.2 Generating and Selecting Support Concepts

After the main functions have been selected, alternative concepts of the support are generated. Through evaluation, combination and refinement, a viable and promising concept is selected for further detailing and implementation. A review of the literature, focusing specifically on the functions of the support is useful at this stage, so as to identify means or ideas from the literature that could be used to fulfil the functions.

To generate new concepts it is important to be receptive, open and inquisitive, and let the environment trigger one's brain cells. There is much knowledge available that can be relevant, that just needs to get linked. It is also important to be critical about one's own concepts and to try to deliberately search for faults, potential shortcomings, uncertainties, *etc*., in order to generate alternative, better concepts.

It is often found that conceptualisation in product design is achieved by *co-evolution*: early, high-level requirements guide the generation of potential solutions, their evaluation leads to the generation of solution-specific requirements, which are then addressed by modifying the solution to add further detail, and so forth (Nidamarthi *et al.* 1997). In our experience, co-evolution holds true for support development as well: generation or adaptation of a particular concept for fulfilling a certain function will often lead to specific, detailed requirements related to that concept. In Appendix B pointers to useful methods can be found.

As pointed out earlier, making assumptions explicit is crucial in research and essential for developing and evaluating support. This implies that the Impact Model should be updated in every stage. The already-mentioned Checklist of Scope and Assumptions (see Appendix B.4) helps in identifying sources of possible assumptions.

Reliability Example

In the reliability example, in order to develop concepts for assessing the levels of CSU, it is necessary to define measures for these levels that can be applied with the data available in the early embodiment stage. As discussed in Chapter 4, clarity, simplicity and unity relate to components, interfaces and their configuration. Based on the literature, unity was defined as the mechanical strength of the components, for which tools are available. Simplicity was defined as the number of components and interfaces, which is relatively easy to determine. Because the literature did not provide a clear definition of clarity, other than that it depends on the clarity of the interfaces, the clarity measurement in DS-I was based on expert opinion. For obvious reasons, this is not an option for support. It is necessary to find a relevant definition of the level of clarity of an interface that can be determined in early embodiment and to find the relationship between the clarity measures of the individual interfaces and the clarity of the embodiment. To that end, information from existing work on clarity has to be combined with the original contribution from the researcher to fill in the voids in current understanding. The focus on components and interfaces lead the researcher to propose the concept of a *component-interface diagram* that can be derived from early engineering drawings or CAD models, assuming that a description of components, interfaces and configuration at a functional or conceptual level description is sufficient to make the measurements and thus assessments.

The PS stage thus focused on three tasks: (1) developing a component-interface diagram from the information available at the early embodiment design stage, (2) developing measures for simplicity, unity and in particular clarity, (3) developing a method for assessing the levels of CSU.

Synthesis Example

At this point we would like to introduce another (simplified) example about a researcher who becomes interested in improving the process of generating product ideas (synthesis) after hearing and reading some literature about the importance of idea generation. This example is inspired by the research presented in Appendix C.3 of this book. We will henceforth refer to this as the *synthesis example*. The synthesis example project focuses on PS and DS-II (research Type 6 in Figure 3.12), whereas the reliability example project had a focus on DS-I and PS (research Type 5 in Figure 3.12). The results of the RC Stage and the Review-based DS-I stage can be summarised as follows.

The researcher reviews the existing literature in depth to understand the current situation and finds several interesting empirical studies that support her idea to aid idea generation. One study showed that most designers do not consider more than a few ideas during their design process. Another study on the design processes of companies with products of poor quality showed that what these companies had in common was also that they did not consider more than a few ideas in each process. Several other studies revealed that an increase in product quality improves customer satisfaction as well as the amount of profit. At the end of the Review-based DS-I the researcher puts the findings together in a Reference Model, a simplified version of which is shown in Figure 5.6. She decides to take the 'number of ideas considered during design' as the Key Factor influencing the quality of the product. High profit and customer satisfaction are taken as Success Criteria. Given the timeframe of the research project, the researcher decides to focus on increased product quality as Measurable Success Criterion. Other factors, such as cost of production, known to affect the amount of profit, are considered outside the scope of the research project. Because of its importance, cost of production is added to the Reference Model, but graphically marked as out of scope.

Based on this Reference Model, the researcher concludes that considering a *large* number of ideas may lead to products of *high* quality, although she also realises that this may require the ideas to be quite different. That is, she assumes that the values of the factors shown in the Reference Model alongside the links can be reversed, provided that a wide range of ideas is considered. Obviously, this is still only an assumption: no studies were found showing that companies that develop products of *high* quality *do* consider a large number of ideas and of a wider range during their design processes.

The researcher envisages the desired situation as one in which a support is available to help designers consider a wider range of ideas in design. This should help achieve better-quality products, which in turn should improve customer satisfaction and the amount of profit. She represents the assumptions and her line of argumentation in an Initial Impact Model (see Figure 5.7).

The researcher proceeds with a PS. After setting up a requirements list for the envisaged support, she considers various alternative concepts for supporting designers to consider a wider range of ideas. One alternative concept for providing a wide range of possible ideas is to develop a catalogue of existing product ideas that can be explored by designers (in the hope they consider these).

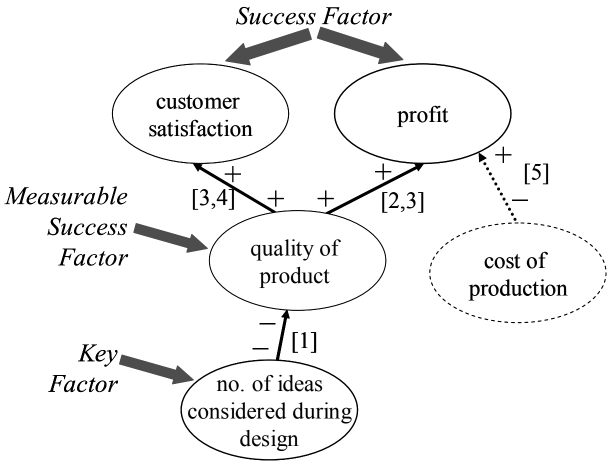


Figure 5.6 Reference Model for the synthesis example (simplified)

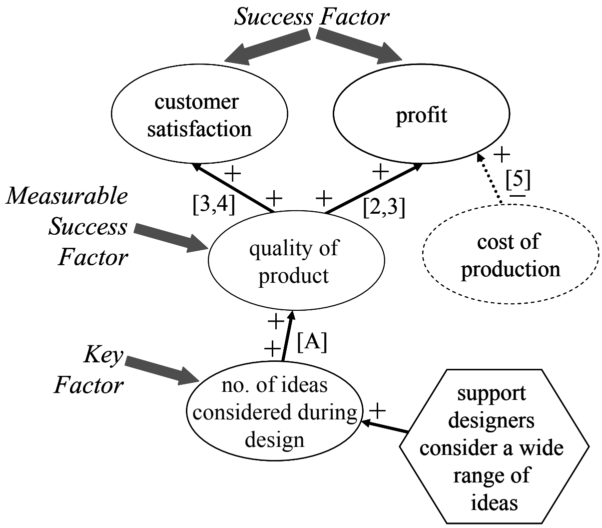


Figure 5.7 Initial Impact Model for the synthesis example (simplified)

Another concept is to provide designers with a wide range of possible ideas that are developed automatically by synthesising exhaustive combinations from a set of idea-building blocks. This concept was chosen as potentially more successful in generating a wide range of ideas, and shown as alternative A in the Impact Model in Figure 5.8. Elaborating on alternative A while taking into account the worry that designers may not consider the ideas that are generated, a third alternative was generated (Concept B in Figure 5.8). This concept not only provides a wide range of ideas but also encourages these ideas to be considered by supporting their exploration. Both concepts lead to several research questions, such as: What is a suitable set of building blocks? What are methods for exhaustively combining these? What form should exploration take? How should the ideas be presented so that the user has an overview?

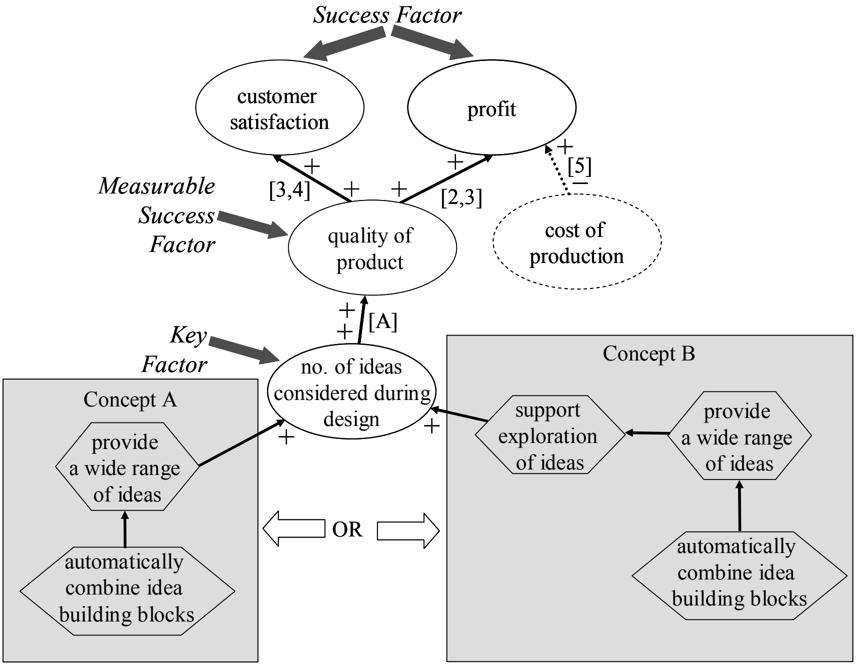


Figure 5.8 Impact Model with alternative support concepts for the synthesis example

The answers to these questions will have consequences for the support, resulting in additional functions. This enforces further refinement of the Impact Model. Take for example the above question about how the ideas should be presented. This was raised because the fulfilment of the function 'provide wide range of ideas' was considered necessary but not sufficient to encourage designers to consider a larger number of ideas. The range should be presented such that it allows an overview, thus introducing an additional function: 'provide overview of ideas'. Both functions might have particular effects that are not yet represented in the Impact Model. Such effects thus lead to additional factors and links, namely those that are influenced by these functions. The first function, for example, affects a factor that can be called 'quality of overview'; the second function affects the factor 'range of ideas'. These factors, *together*, are responsible for influencing the 'number of ideas considered' in the original Intended Impact Model and should be added to the model (not illustrated here).

5.5.3 Introduction Plan

While conceptualising the Intended Support it is useful to start considering the life-cycle phases of the support, *i.e.*, the processes of *introduction, installation,* *customisation, use and maintenance*, and to document these together with any *organisational, technical and infrastructural pre-requisites* in what we call an Intended Introduction Plan. This involves exploring questions such as: How is the support supposed to be introduced, installed and used? Is customisation required and how is this done? Who is involved in each life-cycle phase and in what role? Do the contexts in which the life-cycle stages take place put specific constraints on the support that must be accounted for, such as the type of users, the circumstances and other available support?

The visualisation of the various processes in the life-cycle of the support, *e.g.*, in scenarios or flow charts, is a useful way to identify additional functions and features of the support, necessary to realise these processes. The following questions can be asked for each step in each process:

- Who is involved in this step and in which role? For instance, in the reliability example, who creates the component-interface diagrams (henceforth called CI diagrams) necessary for assessing clarity, simplicity and unity? What interfaces are commonly used by these users?
- How is the step executed (note that this may reveal additional steps that were originally not anticipated)? For instance, if the CI diagram is to be created manually from a CAD model or an engineering drawing, a procedure to verify the correctness of the CI diagram against the CAD model or drawing may be needed. If a support needs customisation, a separate interface may be required to support the various steps in the customisation process.
- How difficult is the execution of the step in the intended context, what is involved? For instance, a CI diagram for a large system might require an enormous effort to generate and might be difficult to use. Hierarchies of diagrams may be required.
- How error-prone is the step? For instance, manual creation of the CI diagram may be highly error-prone, potentially influencing the quality of the outcome as crucial components or interfaces may be left out by mistake, or translated incorrectly into the diagram.

One difficulty in support development is that the effectiveness and efficiency of a support depends on the characteristics of the users and the support, as well as of the nature of their interaction. The higher the degree of freedom for the user as to how the support can be used and the more the support allows different interpretations, the more difficult it will be to ensure that the support will be effective and efficient. Therefore, it is important to identify possible alternative uses and interpretations of the support during the life-cycle phases, while answering the questions listed in the previous paragraph. These considerations can be used to:

- make explicit how the support should and should not be introduced, installed, customised, used and maintained;
- improve the support so that alternative uses and interpretations that are detrimental to the impact of the support are minimised;

- ensure that the evaluation of the support (in DS-II) can be planned and executed such that the observed impact is primarily due to the support and not to other circumstances.

The Intended Support and the Introduction Plan together should ensure that the life-cycle phases can take place as intended and reduce the chances that alternative uses and interpretations have a large negative impact on the effectiveness and efficiency of the support.

Note that the Intended Impact Model assumes that customisation, installation, introduction, use and maintenance are carried out as planned, since no factors are introduced to represent deviations from the envisaged plan. Such deviations can affect the impact of the support and thus result in a situation that differs from the desired situation represented by the Intended Impact Model. For this reason, the Introduction Plan is an important input for the evaluation of the support in DS-II.

The concept for the Intended Support, the Intended Introduction Plan and the Intended Impact Model are developed together. As for any other step, it is important to keep note of the rationale behind the decisions that were taken, such as the problems envisaged, proposals considered, the arguments behind the decision, *etc*.

Reliability Example

In the reliability example, the analysis of the life-cycle processes resulted, among others, in the following. The main users are experienced mechanical designers from the company involved, who would use the support individually. The most effective introduction of the support in the particular context is considered to be a workshop led by the researcher, and potentially by a representative from the company trained by the researcher. The workshop should include an explanation of the support and allow the participants to solve example cases. The alternative to develop a paper- or computer-based tutorial for self-learning was considered to require too much effort. The available systems within the company require organisational and technical procedures for the introduction and installation of the support as well as specific functionalities within the support.

The analysis of the processes through scenarios and using the questions listed above provide:

- a more detailed picture of the processes involved and of the necessary features and functions of the support; *e.g.*, the possible elements in the CI diagram based on the available information in the CAD models or drawings of early embodiments, and the possibility to reuse CI diagram elements and sub-systems;
- an evaluation of the weak points of the support: *e.g.*, the manual creation of a CI diagram task is potentially error-prone and tedious, and every modification of the embodiment requires a modification of the CI diagram. The accuracy of the CI diagram was therefore considered an important issue to be addressed.

The greater clarity of the functions, the concept and introduction of the Intended Support allows a further improvement of the Intended Impact Model. In our example, the above and other results of the analysis are used to update the Intended Impact Model as shown in Figure 5.9.

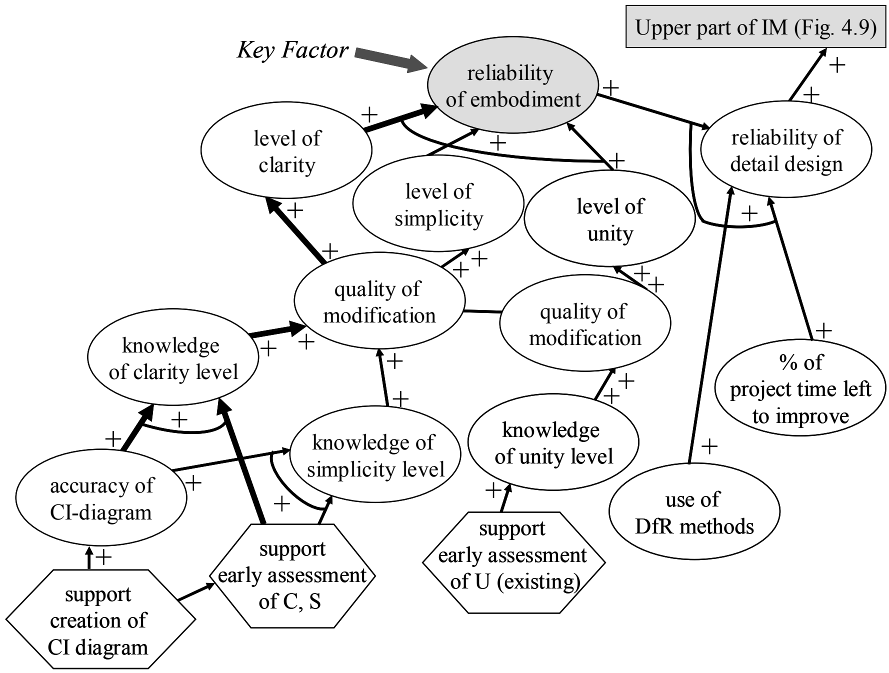


Figure 5.9 Lower part of the Intended Impact Model for the reliability example after the Conceptualisation step in PS

User Issues

Particularly relevant in this stage of support development are user issues. In many dissertations, support concepts are described without any clear indication of who the users are: are they designers, administrators, information maintenance personnel, a combination of these, *etc*? It is often also not clear what is done by the user(s) and what is provided or done by the support, or who takes the initiative for a particular interaction. In this section we focus on how to determine the type and amount of interaction required. This is necessary to clarify the kind of support intended, which in turn is essential for developing the right concept. Further details about user-interface design can be found in Appendix B.3.

In the case of tool development different types of interaction are possible. The types of interaction in the interaction diagram in Figure 5.10 are based on who initiates the interaction process and who transfers data or information.

- The user initiates interaction as well as transfers data, *i.e.*, the user *provides* data.
- The user initiates the interaction while the support provides the data in response, *i.e.*, the user *requests* data of the support.

- The support is the initiator while the user transfers data in response; *i.e.*, the user *replies* to a request.
- The support initiates interaction as well as transfers data; *i.e.*, the user *receives* data from the support based on a decision taken by the tool that this interaction is necessary.

Who initiates interaction / Who transfers data	Human (user)	Support
Human (user)	**Provide**	**Reply**
Support	**Request**	**Receive**

Figure 5.10 Interaction diagram from a user's point of view

Various variants of this diagram exist, both from a user point of view as well as from a system point of view. The labelling of the types of interaction in Figure 5.10 is based on a user point of view, with an increasingly active role of the support. Depending on the proportion of initiative taken by the support, the support can be:

- *manual*, where the user initiates interaction and also transfers data (comprising 'provide' interactions);
- *passive*, where the user is the initiator and the interaction is a combination of 'provide' and 'request' types of interactions;
- *interactive*, which would involve many types of interactions. Depending on the amount of initiative taken by the support, this can be more tool-initiated or more user-initiated;
- *automated* where interactions are mainly of the 'receive' type, with some 'requests' and 'replies' in particular at the beginning and end of the interaction.

These types of interaction should not be taken as rigid divisions. In any one support, several types of interaction can take place. However, it is important to indicate the dominant interaction characterising the type of Intended Support.

Synthesis Example

In Concept A of the synthesis example, the aim is to provide a range of ideas, but not to support their exploration. This reduces the complexity of interaction considerably compared to Concept B in which both range and exploration must be supported. Concept B requires a more intensive user interaction, and strategies for supporting this function have to be developed. It was decided to first develop concept A, as this alternative is simpler and is part of Concept B.

5.6 Elaboration

For each support function, the concept needs to be elaborated by identifying the necessary user interaction and suitable means (or combination of means). We recommend that wherever possible existing means should be used for individual functions, in particular in the case of tool development and for the functions that do not constitute the core contribution of the project. The reason is that the contribution of design research is unlikely to be in the detailed technologies used, but most likely on new functions and concepts for the support. Using existing means not only makes it easier to detail and subsequently realise the support, but also easier to assess the likelihood that the function and thus the support can realise the desired impact.

The Elaboration of the Intended Support starts with a review of the literature to identify ideas and available means to embody the functions of the support (in the case of tool development, there is more in Section 5.7.4). This information is then used to elaborate the concept as far as possible, ideally until the point at which it has been described to such a level that it can be realised. At the same time, the Intended Introduction Plan is elaborated and the Intended Impact Model updated where necessary.

The evolving description of the Intended Support, the Intended Introduction Plan and the Intended Impact Model are assessed by asking questions such as:

- Is the description plausible?
- Does it relate to the need?
- Does the literature suggest opposing evidence to any of the claims?
- Is the description such that the claims can be evaluated?

Reliability Example

In the reliability example, the researcher decides that the Intended Support should automatically create a CI diagram and provide a partially automated assessment of the levels of clarity, simplicity and unity. The reasons are that for all embodiments a CAD model is available, and that the manual creation of a CI diagram is a tedious and error-prone process. A requirement is that the Intended Support is to be used as part of the designers' current process.

For the assessment of the level of unity, a suitable computer tool was found, which should be integrated into the Intended Support. For the assessment of clarity and simplicity levels nothing is available. A method has to be developed such that the Intended Support can determine the CSU levels semi-automatically using the CI diagram it has created, the results from the unity assessment tool, and input from the designer. The support should prompt the designers for the necessary inputs, such as the level of clarity for each interface, using a pre-specified format. The researcher also considered whether the support should point to relevant guidelines for improving the CSU levels in a given context, but finally decided to leave this out, because the guidelines were found to have been difficult to use. More work on the guidelines is necessary before the above functionality can be included.

The consequences of these decisions are taken into account in the Intended Introduction Plan and the Intended Impact Model. For instance, the decision to automate the creation of the CI diagram is expected to significantly reduce the time needed to evaluate embodiments, and therefore should increase the '% time left to improve', as seen in the updated Intended Impact Model in Figure 5.11.

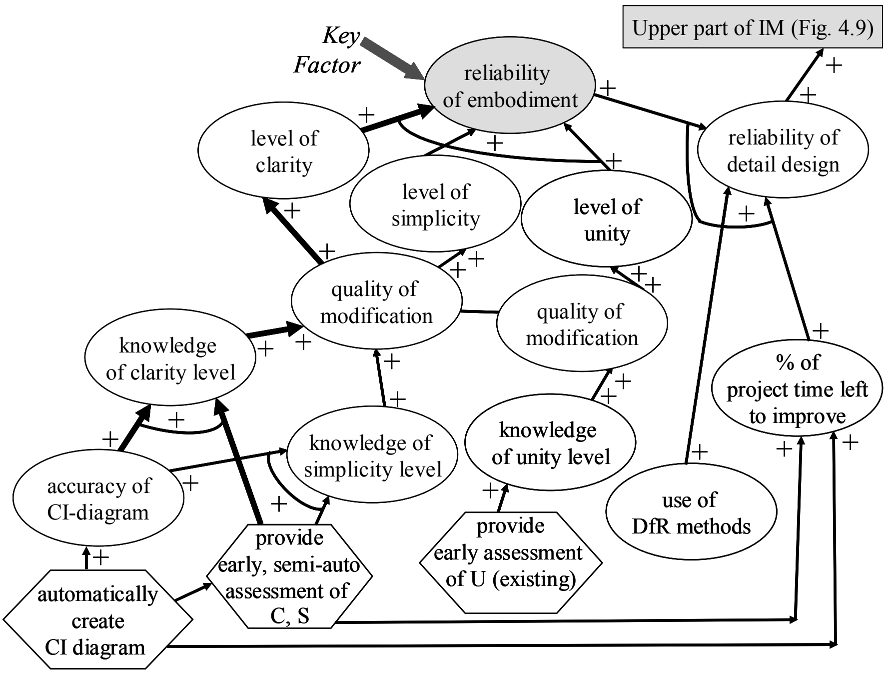


Figure 5.11 Lower part of the Intended Impact Model of the reliability example at the end of the Elaboration step in PS

Knowledge Issues

In many research projects, knowledge issues will be particularly dominant at this stage of development, although they will also play a role in the task clarification and conceptualisation steps.

In support development, research questions are often associated with a specific design issue, activity, stage and application. For instance, Stephenson (see Appendix C.8) asks "How can we improve evaluation (activity) of reliability (issue) at the embodiment design stage (stage) for earth-moving equipment (application)?" Chakrabarti (Appendix C.3) asks "How can we support designers in generating (activity) a wider range of mechanical designs (application) at the conceptual stage (design stage) to fulfil the intended functionality (issue)?" The intended use of the support for a particular application makes it imperative to gather domain knowledge. The issues, activities and stages provide the relevant context and thus focus.

Domain knowledge will be necessary as part of some support for its functioning (*e.g.*, a set of guidelines for design for Ergonomics), while for other support knowledge will be provided by the user (*e.g.*, a morphological chart). Where domain knowledge plays a significant role, systematic gathering and structuring of such knowledge can be a significant portion of PS unless it has already been collected as part of DS-I.

In the *reliability example*, for instance, the researcher needed to gather technical details and failure data of specific sub-systems of existing equipment, as well as the guidelines and procedures designers apply in the early stages of embodiment and the information they have at their disposal, in order to be able to develop the assessment procedures and the CI diagram concept. Most of the data had already been collected in DS-I.

In situations where the required domain knowledge is not available or not complete, this knowledge must be generated. This requires an understanding of the domain. In the *synthesis example*, for instance, the Intended Support depends on the availability of a sufficiently rich, minimal, generic set of building blocks that can be easily combined into variants early in the development process. Being able to identify fundamental building blocks is a crucial aspect of this work and one of its main contributions: the building blocks suggested in the literature did not fulfil the requirements. The researcher, therefore, analyses the designs available in the domain for basic patterns (similarities) at the functional level. The result is a model and a modelling approach, as well as a set of generic building blocks. These are to be used in the Intended Support to generate new design concepts to be explored by the designer. This implies that the contribution of this research project is not only a support but also new (domain) knowledge about how to model technical systems at an abstract level, so as to support their creation.

The *synthesis example* illustrates a case in which the elaboration of the Intended Support requires a revisit of the DS-I stage, driven by specific questions dictated by the needs of the PS stage. This second DS-I focused on obtaining an understanding about similarities and fundamental functional differences of technical systems with a focus on mechanical designs. The results were models of building blocks and rules for their combination. The modelling language used was a creation of the researcher, as nothing suitable was available. The development of the building block models and combination rules could have been part of PS, had the aim of PS been to develop a tool for modelling building blocks. In the example this is not the case: the building blocks are given and a tool for their use was to be developed.

Various techniques for knowledge acquisition are available, especially from the discipline of knowledge engineering, see Appendix B.2.5. In addition, many data-collection methods from DS-I may be relevant (see Chapter 4 and Appendix A.4).

Deliverables

The Elaboration stage results in the completed **Documentation of Intended Support**. This includes the Intended Impact Model, the Intended Introduction Plan, and the **Intended Support Description**. The Intended Support Description describes the support in terms of the need or problems addressed, the goals and objectives of the support, its elements, how it works, the underlying concepts, theory, assumptions and rationale, and how it is to be realised.

At this stage of research it is useful to ask oneself the following reflective questions:

- Why do I believe this support leads to a result?
- What is my contribution to this support?
- Why do I believe this contribution to be academically worthwhile?
- Why do I believe the support to be practically worthwhile, or to contribute to a practical goal?
- Why do I believe that I have the competences or can obtain the competences to realise this support (if applicable).

5.7 Realisation

The Intended Support is a description of the complete support as envisaged by the researcher. The purpose of the Actual Support is to evaluate the *core* contribution, *i.e.*, it is a proof-of-concept of the main ideas and novel elements of the support, not of all functionalities the Intended Support is planned to have. Furthermore, as mentioned in Section 5.3, the resources available in a research project are often insufficient to realise the entire range of intended functionality or the detail necessary for introduction, use and evaluation in the target context. What is actually realised – the Actual Support – can therefore be, and often is, more restricted than the Intended Support.

5.7.1 Core Contributions, Support Functionalities and Outline Evaluation Plan

The Realisation step starts with determining the core contributions and essential functionalities of the Intended Support. Then, an outline for a plan is generated on how to evaluate these contributions and functionalities – the **Outline Evaluation Plan**. This plan is then evaluated for its feasibility, taking into account constraints such as available time, competences required, availability of participants, *etc*.

The Actual Support should be realised to such an extent that it can be evaluated. As discussed in Chapter 2, we distinguish three types of evaluation. Support Evaluation involves the continuous assessment of the support during its development focusing on in-built functionality, consistency, *etc*. Application Evaluation focuses on usability and applicability, *i.e.*, the ability of the system to address the Key Factors as intended. Success Evaluation focuses on the usefulness of the support, *i.e.*, its ability to realise the expected impact and fulfil the Measurable Success Criteria.

Support Evaluation is the type of evaluation that takes place in the PS stage (see Section 5.8). Support Evaluation should ensure that the support is developed such that it can be used in DS-II for Application and Success Evaluations. The functionalities and features of the Actual Support and the Application and Success evaluations will thus strongly influence each other. This implies that the researcher needs to think already about the content of Application and Success Evaluation in the PS stage, by developing an Outline Evaluation Plan in parallel to the support (as illustrated for tool development in Figure 5.14).

The scope of the Actual Support, *i.e.*, the chosen functionalities, is a balance between what needs to and can be evaluated, and what functionalities the Actual Support must and can have in order to make this possible. The essential functionalities are determined on the one hand by the functionalities that represent the core contribution of the research project – as these are to be evaluated – and, on the other hand, the necessary and possible evaluation methods.

To find the core contributions and essential functionalities that should, ideally, be realised in the Actual Support, the following questions can help:

- Which features and functionalities of the Intended Support are your core contributions?
- What is the focus of the evaluation, *i.e.*, which factors and links in the Impact Model are the most useful to evaluate, *e.g.*, those that connect most strongly to success? Given the project constraints, which ones can be evaluated?
- Which features and functionalities are essential given the scope of the evaluation?
- Which (additional) features and functionalities are essential to measure the impact?

The evaluation may require additional functionalities: for instance, if designers are going to evaluate the support, a good user interface is needed, even though the user interface may not be a core contribution of the researcher. In addition, the Actual Support may include functionality that will not be part of the Intended Support, but added, *e.g.*, for evaluation purposes or because of the chosen medium. A typical example is a data log function to evaluate the use of a support tool. This is illustrated in Figure 5.12.

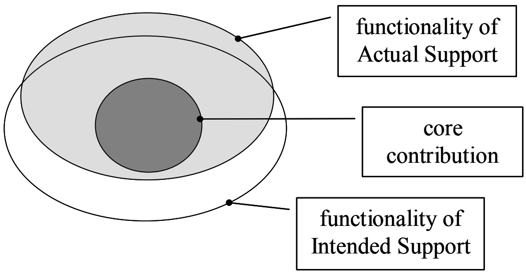


Figure 5.12 Core contribution in relation to Intended and Actual Support

The choice of functionalities to be realised also depends on the context in which the Actual Support is to be evaluated, in particular the type of users and their role. The Actual Support will be quite different depending on whether designers are involved as users in the evaluation, whether designers are used as a control group with the researcher or computer as ‘user’, or whether designers are not involved at all. Similarly the type of task to be solved during the evaluation will affect the necessary functionalities. As we will discuss in Chapter 6, the context for the evaluation of the Actual Support does not necessarily have to be the same as that of the Intended Support for the evaluation to provide useful information about the effects of the support. It all depends on the research questions and hypotheses to be addressed by the evaluation. Although the available time for the research does play a role in determining the evaluation focus, it should not *dominate* it. Not involving designers as users just to reduce the effort in developing an appropriate user interface is not acceptable if it compromises the evaluation questions.

Our recommendation is to limit the scope of the Actual Support to what is absolutely essential and core to the project, at least for the first evaluation, and to clearly formulate the aim of the evaluation and the related research questions and hypotheses. It is better to thoroughly evaluate some aspects, rather than to try to evaluate everything, potentially ending up without any strong evidence.

Creativity is required to find an effective and efficient combination of evaluation approach and Actual Support functionalities and features. The iterative process of co-evolving the Outline Evaluation Plan and the Actual Support ends with a list of core contributions and essential functionalities that can and must be evaluated within the available boundary conditions, and an Outline Evaluation Plan that indicates how this evaluation will have to be carried out.

Details of how to set up an Evaluation Plan is described in Section 6.6. We recommend, however, reading the whole of Chapter 6 on evaluation before determining the functionalities and features of the Actual Support.

Synthesis Example

In the *synthesis example,* the main emphasis of the Intended Support based on Concept A (see Figure 5.8) is to ensure that designers are able to describe easily a design problem for which the support is then to provide a wide range of ideas. The core contribution and hence the core functions are to help designers describe a design problem easily, to provide a wide range of ideas, and to help designers access these ideas without difficulty. The researcher assumes at this stage that if she is able to do this with help of the support, this might suffice as a first confirmation of these core functions, even though this does not mean that designers would necessarily be able to do so in a real project. Based on this assumption, she decides not to focus much effort on developing an elaborate user interface, but instead on a support that provides a means to help describe the problem and to offer a wide range of ideas in response. Note that by doing so, she also decides about the type of evaluation, namely the use of the support by researchers rather than designers, and on the main evaluation questions. These questions are whether users can easily describe a design problem, whether the support can provide a wide range of ideas, and whether the users can access these ideas without difficulty.

5.7.2 Developing Actual Support

Once the functions of the Actual Support have been identified, a literature review is undertaken to find examples, platforms and technologies that can help the realisation of these functions.

It is useful to evaluate each potential means for realising a particular function by reflecting on:

- the potential benefits, which for existing or adapted solutions could be found in the literature;
- the potential drawbacks, such as potential side-effects, implementation difficulties, problems of integrating this solution with other parts of the support or related support. Again, for existing or adapted solutions the literature may provide information.

As discussed earlier, the Actual Support is a 'proof-of-concept' that may differ from the Intended Support in that:

- a different implementation is used, such as interactive rather than automated;
- a different medium is used, such as paper instead of software, an expert rather than a knowledge base;
- not all functionality is available, such as multi-user access, a maintenance function, help functions;
- domain coverage is not complete, such as a solution catalogue containing only certain types of solutions, a procedure only applicable to designs with few components, or a database that is only partially filled;
- performance is not optimised, such as low speed of processing, inefficient storage, or low robustness.

Since each of these differences affect what can be said about the applicability, usability and usefulness of the support and hence affect the evaluation, the choices regarding the realisation of the Actual Support must be done carefully, and guided by the evaluation questions and hypotheses. It is of little use to develop a support that cannot be evaluated in a way that allows useful conclusions to be drawn. The larger and the more fundamental the differences between the Intended and the Actual Support are, the more difficult it will be to extrapolate the results from the evaluation to the intended situation.

The differences between the Actual and the Intended Support will have consequences for the Introduction Plan and the Impact Model developed thus far. This will be discussed in Section 5.7.3.

Synthesis Example

In the synthesis example, the researcher selected exhaustive synthesis as a way of providing a wide range of ideas. She envisaged the Intended Support as one having an exhaustive database of idea-building blocks and an algorithm that can exhaustively combine these. However, developing an exhaustive database of building blocks seems unnecessary to demonstrate that this support concept would be capable of providing a wide range of ideas. For a proof-of-concept, the researcher decides that it is sufficient to develop a database of building blocks for a certain domain of devices and to develop an evaluation task related to this domain. She chooses the domain of mechanical transmissions. She decides to carry out two tests in the Support Evaluation. In the first test some colleagues, all PhD students with an engineering degree, are asked to generate ideas using pen and paper for a particular transmission design problem. These ideas are then analysed by the reasercher to determine whether these can all be described using the building blocks and their representations. If that is the case, then at least all the participants' ideas can be potentially generated by the intended algorithm. In the second test, the researcher uses the system to generate ideas for the same problem as used in the first test. If it can be shown that the algorithm also generates ideas over and above those generated in the first test, this would demonstrate that the support concept is able to provide a wider range of ideas than the participants, at least for the considered cases. The two tests are repeated for a number of problems in the chosen domain. If these are successful, this would imply that the support is likely to have the desired impact for the chosen domain. The researcher thus decides that the Actual Support should have a representation and a database of building blocks that is sufficiently general, and an algorithm to generate ideas, but does not require a database of building blocks as exhaustive as envisaged in the Intended Support.

Reliability Example

In the reliability example, the researcher decides that the Actual Support can take the form of a paper-based guide on how to draw up a CI diagram and how to assess clarity and simplicity. The reasons have partly to do with the lack of time available to realise a software program (the focus of the research was on DS-I). More importantly, however, the researcher argues that a computer-based realisation is not necessary for testing the *core* contribution, that is, to evaluate whether it is possible to draw up a CI diagram, whether the levels of clarity and simplicity can be assessed using the CI diagram, whether this has an effect on the 'reliability of embodiments' and whether this in turn has an impact on 'high product quality' – the Measurable Success Criterion chosen. The assessment of unity does not belong to the core contributions of the researcher: support already exists. To remain focused on the core contribution, the researcher decides to base the evaluation on embodiments with components that have sufficient strength (*i.e.*, an adequate level of unity). The consequences of the paper-based version on the speed of the assessment process and the level of complexity of the examples that can be evaluated are considered less relevant at this stage: if the principle does not work, speed and complexity do not matter. The functions of the Actual Support are therefore reduced from 'automatically create CI diagram' and 'provide early, semi-automatic assessment of CS' (see Figure 5.11) to 'support manual creation of CI diagram' and 'support manual assessment of CS'.

5.7.3 Actual Introduction Plan and Actual Impact Model

Once the Actual Support is developed, an Introduction Plan for this support needs to be drawn up, the Actual Introduction Plan, describing how the Actual Support is to be introduced, installed, customised, used and maintained within the context (organisation, infrastructure, users, *etc.*) in which it is to be evaluated. This plan should reflect the differences between the Actual Support and the Intended Support.

Compared to the Intended Introduction Plan, the Actual Introduction Plan may differ in:

- the type of introduction, such as an oral instruction rather than a manual;
- the level and type of installation;
- the level of customisation or the possibility to customise;
- the users, the tasks to be solved, and the environment;
- the need for maintenance.

The Intended Impact Model also has to be adapted to be compatible with the Actual Support and Introduction Plan. The result is an Actual Impact Model to be used for the evaluation of the Actual Support. Additional factors and links may have to be introduced to reflect (and evaluate) the possible side-effects of the Actual Support and some of the factors and links in the Intended Impact Model might not be addressed by the Actual Support. The support functions can also be different, as the reliability example above showed. This may influence the choice of Measurable Success Criteria, which, however, would affect the strength of conclusions that can be drawn from the evaluation. The Key Factor(s) and Success Criteria should still remain unaffected. It is the Actual Impact Model that will determine the Evaluation Plan in DS-II.

Reliability Example

The Actual Impact Model of the reliability example is shown in Figure 5.13. This model is different from the Intended Impact Model in Figure 5.11 in the following:

- Rather than computationally supporting the creation of CI diagrams and the assessment of clarity and simplicity, support is provided to do these manually.
- The assessment of unity is not part of the Actual Support, as this is already covered by existing support. The related factors and links are not evaluated (drawn with dotted lines). Since modifications made by the designer to improve clarity and simplicity can potentially impact the level of unity, a new link is added between 'quality of modification' in clarity and simplicity and the 'level of unity'. The '?' sign alongside this link indicates that it is not clear what the effects will be.
- The influence of the support on '% of time left to improve' is not part of the evaluation, since the Actual Support uses a manual process, which is not expected to influence this factor.

Deliverables

The Realisation stage results in the Actual Support as well as the completed **Documentation of the Actual Support.** This includes the **Actual Support Description**, the Actual Impact Model, and the Actual Introduction Plan. The Actual Support Description describes the support in terms of the need or problem addressed, the goals and objectives of the support, its elements, how it works, the underlying concepts, theory, assumptions and rationale, and how it is realised.

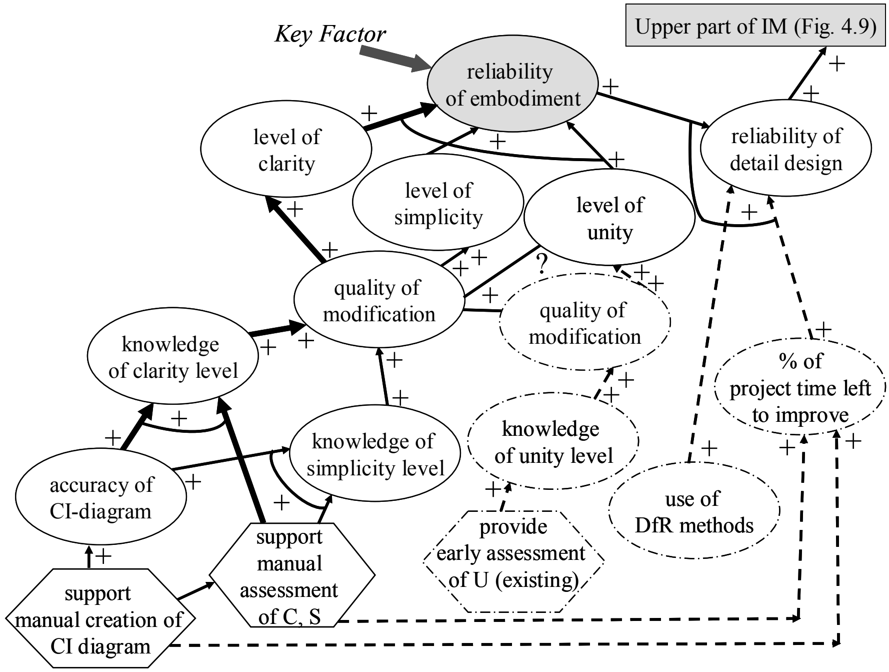


Figure 5.13 Lower part of the Actual Impact Model of the reliability example

5.7.4 Tool Development

In many instances, the Actual Support takes the form of a computer tool. Tool development can take up a substantial portion of the researcher's time. The software development methodologies in Appendix B.2 should be particularly useful to ensure that this time is used effectively. These methodologies have a number of well-defined stages that help the tool developer to:

- identify the interaction of the tool with the environment;
- clarify the interaction between its sub-systems; and
- gradually develop a clear picture of what form these sub-systems should take.

The main points are summarised in this section.

One useful feature of these methodologies is that they help develop the software such that the built-in functionality of the tool can be easily tested (Support Evaluation). However, these methodologies are intended mainly for large, commercial software systems where teams of software developers work together. Support development effort in research is often much smaller in scale, and therefore the researcher should be careful not to get lost in the details of these methodologies. We find them useful to follow in so far as they are helpful in developing the specification of the software at various stages of detail and in ensuring that these specifications are logically linked to one another.

In the context of knowledge based systems development, the Common KADS methodology (Schreiber *et al.* 2000) formulates a set of 'How' questions to guide the realisation phase, which can be useful for all tool development (see also Appendix B.2.5). These questions help clarify the technical system specification in terms of architecture, implementation platform, software modules, representational constructs and computational mechanisms required to realise the tool.

Raphael *et al.* (1999) propose in their Computer Aided Engineering (CAE) tool development methodology, that the main three issues to be addressed to realise a tool are:

- choosing representation(s);
- choosing methods (*i.e.*, reasoning procedures);
- defining visualisation and distribution needs (*i.e.*, user-interface needs).

We suggest adding to the latter the needs for introduction, installation, customisation, and maintenance.

We would like to highlight the CaeDRe methodology, developed by Bracewell *et al.* (Bracewell and Shea 2001; Bracewell *et al.* 2001; Langdon *et al.* 2001) as it supports the development of *design* tools in a *research group setting* and is based on our DRM philosophy. CaeDRe aims to provide a systematic process for producing evaluation-ready prototype systems targeted at improving design processes. The reasons behind the development of CaeDRe were that design researchers "need the necessary tool set and support to rapidly prototype research systems without being unnecessarily hindered by implementation details and fast changes in computer technology and standards", and develop these research systems (which we call design support tools) such that they can be evaluated for their "capabilities and merit for fundamental research output beyond initial benchmark tests". This should lead to the development of research systems that are "both theoretically capable and suited to achieving these capabilities within varying design processes", and "enable the development of useable computational design tools early on in research projects", taking into account that design researchers often do not have the breadth of experience necessary for software development and are constrained by the limited time available for implementation.

CaeDRe unifies three complementary approaches: the Cae tool development methodology developed at EPFL (Raphael *et al.* 1999), the product platform concept used in industry, and the DRM described in this book. Figure 5.14 shows the methodology, clearly illustrating the link with DRM[15] in the four stages, the need to use methods from the Social Sciences, as described in Chapter 4, and the need to start working on the Outline Evaluation Plan during the development of the design tool. CaeDRe adds methods from engineering software development and refines the PS stage specifically for tool development.

[15] Note that the terminology of the stages of DRM has evolved since the publication of this article.

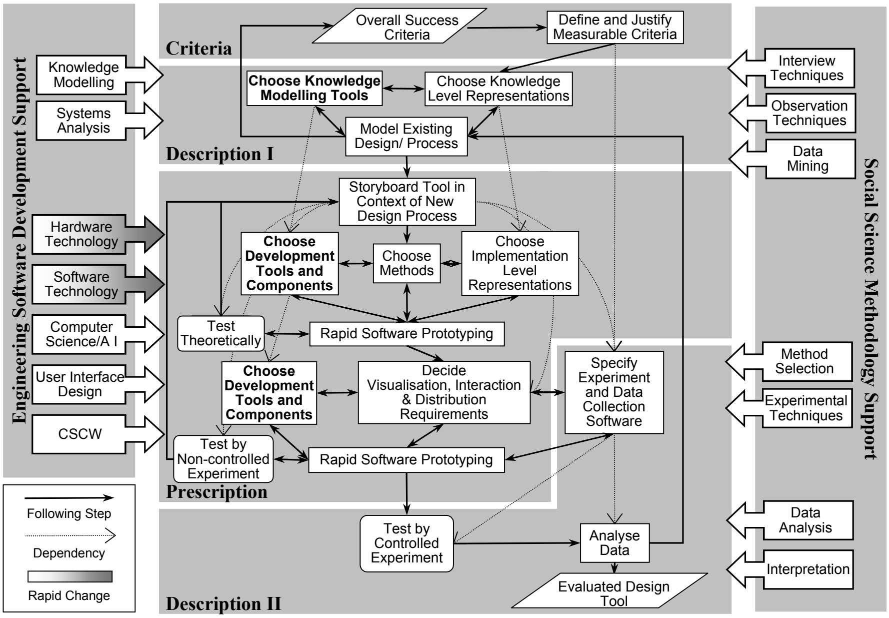


Figure 5.14 Tool Development Model from Bracewell *et al.* (2001)

The CaeDRe tool development process has five activities (see Appendix B.2.1 for details):

- Task definition.
- Choice of representations.
- Choice of methods.
- Definition of visualisation, interaction and distribution strategies.
- Theoretical and experimental validation.

Bracewell *et al.* (2001) add that requirements for testing and evaluation of the tool are a critical issue that needs to be taken into account while defining the visualisation and distribution needs, *e.g.*, the requirements for data collection.

Before starting programming, it is often helpful to write the structure and expected behaviour of the computer program using a formal model. Several alternative ways of modelling are possible within the two major software development paradigms: function-oriented and object-oriented (see Appendix B.2.2). This not only prepares the programming, but allows the use of logic to pre-check the program.

Programming can be supported by CASE (computer-aided software engineering) tools, workbenches or environments (see Appendix B.2.6). Choosing the right software platform can sometimes be difficult. It may be helpful to go through the examples or tutorials provided by commercial platforms, to see if the application is suitable for the job at hand. Procuring a software platform can be expensive, although more and more freeware is available on the Internet. In practice, the choice of a software platform will be limited by constraints such as the available hardware and software in the research group or the environment in which the Actual Support is to be evaluated.

During programming, it is natural and commonplace to make mistakes. In order to expedite error-free realisation of the tool, it is helpful to follow the popular *design-test-debug* cycle, where

- a provisional version of the program is implemented first;
- the program is checked for its required functionality;
- if the result is not satisfactory, errors are identified and the program modified;
- in the case of a more serious error in the algorithm itself, the concept is re-evaluated.

A software development methodology to speed up the design-test-debug cycle is called *prototyping* (Smith 1991) This methodology is especially useful in cases where it is hard to clarify what functionality may be useful or even necessary, or to evaluate a concept before it is elaborated, *e.g.*, a concept for a user interface. This entails the development of a quick computer implementation of some initial ideas, have it evaluated by some potential users and modify it based on the feedback received. In terms of DRM, this implies quick iterations between the PS and DS-II stages. The added advantage is that computer implementation will force the researcher to discipline thoughts.

We end this section with a note of caution about user-interface design: often researchers spend unduly long hours programming and modifying user interfaces, even though they are not part of the core contributions. It is essential to stay focused and implement only what is absolutely necessary in order to evaluate the program's functionality and impact. Because of the effect the interface can have on the use of the support and thus on its impact, it is important to do a pilot study, so as to identify and correct harmful side-effects of the user interface that can interfere with the desired functionality of the support. The pilot study should focus on obtaining specific feedback, rather than a general opinion, to allow effective modification of the user interface. For more details on user-interface design, see Appendix B.3.

5.8 Support Evaluation

As discussed earlier, the type of evaluation that takes place in the PS stage is Support Evaluation. Support Evaluation involves verification, that is, checking that the support fulfils the requirements. Support Evaluation can be useful:

- during Task Clarification and Conceptualisation: Typically these steps will generate descriptions of the support at various levels of detail. Support evaluation involves checking for consistency (that each part at one level of detail is addressed by some part at the other) and completeness (that each function intended to be addressed by the support is indeed addressed). It also involves logical checking of the detailed functionality at the lowest level. Together, these ensure that the detailed functionality has a strong chance of realising the intended impact.

- during Elaboration and Realisation: The evaluation should check that the smallest elements of the support can function as intended, that they are consistent with one another, and then check progressively if the whole support integrating these elements, can work as intended (see Appendix B.2 for more details of software testing).

It might be useful to carry out some of the Application Evaluation (see Chapter 6) before finalising the PS stage. This increases the chances of the support being applicable and reduces the number of iterations required. The absolute minimum Application Evaluation involves the researcher as the user using a provisional version of the support to evaluate its usability and applicability. Rather than taking an imaginary task or situation, the researcher could take an existing case or data set to run through the support. This will increase the likelihood of finding problems. The familiarity of the researcher with the support, however, remains an obvious disadvantage. A more effective approach is to ask other people, preferably those with similar backgrounds to potential users, to use the support. Students and colleagues are two common options. For more details see Chapter 6.

The results of the Support Evaluation have to be reflected not only in the Actual Support, but also in the Actual Introduction Plan, the Actual Impact Model and the Outline Evaluation Plan.

Synthesis Example

In the synthesis example, the two Support Evaluations are undertaken as planned (see Section 5.7.2). The first Support Evaluation is successful: the ideas generated by the colleagues using pen and paper can all be described using the set of building blocks and their representations. The set seems to be sufficiently complete for the chosen domain, and the ideas can potentially be generated by the support. The second Support Evaluation is also successful: the tool indeed generated a wide range of ideas, even more than those generated by the participants. Based on these results, the researcher decides to continue with the evaluation of the actual use of the tool, *i.e.*, to undertake an Application Evaluation as outlined in the Outline Evaluation Plan, and proceeds to the DS-II stage.

Reliability Example

In the reliability example, the researcher does a Support Evaluation to assess the completeness of the instructions for creating CI diagrams and assessing the levels of clarity and simplicity. For this purpose, he runs through an example case, using the instructions. Whether designers can understand and use the instructions, whether the ‘accuracy of CI diagram’ and the ‘knowledge of CS levels’ of the designers improve, and whether the Key Factor – reliability of embodiments – changes as desired, is not the topic for Support Evaluation. The research plans to investigate these using an Application Evaluation in DS-II, based on the Outline Evaluation Plan developed in PS.

5.9 Main Points

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

- This chapter focuses on how to develop a design support to enhance, eliminate or reduce the influence of some of the critical factors found in DS-I or DS-II.
- Design support includes all possible means, aids and measures that can be used to improve design. These are prescriptions, suggesting ways by which design tasks should be carried out.
- Support development is usually not a direct derivative of the findings from DS-I or DS-II, but involves a highly creative and imaginative design process. Design methodologies can be used in this process. When a computer tool is developed, software development methodologies can be useful.
- User aspects and interaction issues are important in support development. User involvement at all stages is beneficial and at the latter stages often essential.
- When a support is developed for a specific domain, domain knowledge plays a major role, both in support development and testing. Methods are available for acquisition of this knowledge.
- The Systematic PS process detailed in this chapter blends two general features of product and software design methodologies: gradual detailing through steps and problem solving at each step. Its steps are: Task Clarification, Conceptualisation, Elaboration, Realisation, and Support Evaluation.
- We emphasise the need to develop a strong concept based on the generation of variants, before embarking on the elaboration.
- Three types of PSare distinguished in DRM: Initial, Comprehensive and Review-based.
- An Initial PS follows a Comprehensive DS-I or DS-II and is carried out if a Comprehensive PS is not possible. Only the Task Clarification and Conceptualisation steps of the Systematic PS process are executed, to illustrate how the results of DS-I or DS-II could be used to improve design. The support cannot be formally evaluated.
- A Comprehensive PS includes all the steps of the Systematic PS Process. The outcome is a support realised to such an extent that its core functionality can be evaluated for its potential to fulfil the purpose for which it was developed. How comprehensive the PS stage is, depends on the intended evaluation in DS-II.
- A Review-based PS is necessary in projects that focus on the evaluation of existing support (DS-II) that has been developed without the researcher being involved. All the steps in the Systematic PS process are carried out to reconstruct or develop the documentation needed for starting a Comprehensive DS-II.
- A distinction is made between Intended Support and Actual Support. The Intended Support is a description of the complete support as envisaged by

the researcher. Project constraints often prevent all functions of the Intended Support from being realised. The support that is realised is the Actual Support, which may be incomplete in several respects, but can be used for the purpose of evaluation – as a proof-of-concept. The focus is on the core contribution of the research project, *i.e.*, the core functionality of the Intended Support.

- The Task Clarification step helps clarify the requirements for the support. These must reflect the goal of the project, and help attain novel and implementable changes to be brought about by the support.
- Based on the Reference Model and Initial Impact Model, alternative Impact Models – based on alternative Key Factors, influencing factors, links or Measurable Success Criteria – should be generated, explored and evaluated before selecting the most suitable: the Intended Impact Model. The majority of links in the Impact Model are taken as assumptions.
- Based on the Intended Impact Model, project goals and the Reference Model, a list of requirements for the Intended Support is formulated. The requirements pertain to the whole life cycle of the support.
- In the Conceptualisation step, the functions of the support are identified, alternative concepts generated, a concept for the Intended Support selected, its Introduction Plan conceptualised, and the Intended Impact Model updated.
- The Intended Introduction Plan documents the intended processes of introduction, installation, customisation, use and maintenance.
- The Elaboration step starts with a literature review to identify ideas and available means to embody the functions of the support. The step results in the completed description of the Intended Support, the Intended Introduction plan and the Intended Impact Model.
- The Intended Support Description describes the support in terms of the need or problem addressed, the goals and objectives of the support, its elements, how it works, the underlying concept, theory, assumptions and rationale, and how it is to be realised.
- The Realisation step determines the core contributions and the essential functionality of the Intended Support to be evaluated, generates an Outline Evaluation Plan, consults the literature on existing ideas, means and technologies, develops the Actual Support, and develops the Actual Introduction Plan and Actual Impact Model by adapting the Intended Introduction Plan and Intended Impact Model to match the Actual Support.
- The Actual Introduction Plan describes how the Actual Support is to be customised, installed, introduced, used and – if applicable – maintained within the context in which it is to be evaluated.
- The Actual Support is the only support developed in PS and may differ from the Intended Support in: implementation, medium, functionality; coverage; performance. Additional functionalities or features may be required for the evaluation.
- In tool development, computer implementation is a substantial portion of the research project. It is helpful to follow the design-test-debug cycle.

- It is important to remain focused, and realise only what is absolutely necessary in order to evaluate the program's core functionality and intended impact.
- In PS, only Support Evaluation is carried out to evaluate completeness, internal consistency, *etc*. Users may be involved. Support Evaluation can take place throughout the PS stage.
- During the various steps of support development, assumptions are made. It is important to record these, since these can provide alternative explanations for the evaluation results.
- The deliverables of the PS stage are: Documentation of the Intended Support, Actual Support, Documentation of the Actual Support, Support Evaluation results, and Outline Evaluation Plan for DS-II . An Initial PS will only result in the Documentation of the Intended Support.

6

Descriptive Study II: Evaluating Design Support

The focus of this chapter is on the fourth stage of DRM: the DS-II stage. It discusses how empirical studies can be used to evaluate the application and impact of the design support that has been developed in the PS stage (see Chapter 5). Because the functionalities and means of the realisation of the Actual Support and the Evaluation Plan are closely linked, an Outline Evaluation Plan should already be formulated while developing the support, *i.e.*, the DS-II stage should start during the PS stage.

Many details of planning and undertaking an empirical study have been discussed in Chapter 4 and the reader is expected to be familiar with that chapter. This chapter focuses on: the types of evaluation we distinguish; the differences between DS-I and DS-II; existing evaluation approaches; and the specific issues related to the use of an empirical study for evaluation purposes.

As described in Section 2.6.4 the objectives of DS-II are:

- to identify whether the Actual Support can be used for the task for which it is intended and has the expected effect on the Key Factors (Application Evaluation);
- to identify whether the support indeed contributes to success (Success Evaluation), *i.e.*, whether the expected impact, as represented in the Impact Model, has been realised;
- to identify necessary improvements to the concept, elaboration, realisation, introduction and context of the support;
- to evaluate the assumptions behind the current situation represented in the Reference Model, and the desired situation represented in the Impact Model.

The deliverables of the DS-II stage are:

- results of Application Evaluation;
- results of Success Evaluation;
- implications and suggestions for improvement of:
    - the Actual Support;

- the Intended Support, its concept, elaboration and underlying assumptions;
- the Actual and Intended Introduction Plan including introduction, installation, customisation, use and maintenance issues;
- the Actual and Intended Impact Model;
- the Reference Model;
- the criteria used.

Section 6.1 discusses the importance of evaluation, the different types of evaluation in DRM and some evaluation approaches from other areas that are relevant for design research. The types of DS-II are described in Section 6.2, the proposed Systematic DS-II process is introduced in Section 6.3 and its steps discussed in detail in Sections 6.4 to 6.8.

6.1 Evaluation

6.1.1 Importance of Evaluation

When developing products, it is good practice to evaluate results throughout the process, in particular to determine whether the product (or its description) can be released to the next stage. In the same way design support should be evaluated at various stages in its development process irrespective of the type of support and the extent to which the support has been developed. The Institute of Electrical and Electronics Engineers (IEEE) defines what they call validation as the "confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled" (IEEE 1998). For a further discussion on validation of methods and in particular of design methods, see Frey and Dym (2006); Seepersad *et al.* (2006).

Evaluating design support is essential because its effects can only be assumed while developing the support. This is due to the fact that

- the support is a creation, involving various assumptions introduced during the translation or extrapolation of what was found in reality or in a theory, and during the development of the initial idea;
- the introduction of the support creates a new situation that did not exist before and about which only assumptions exist: many unexpected effects may occur;
- the context in which and for which the support is being created is dynamic.

Developing support is a creative process. From an investigation into the needs and problems – the current situation – and available theories, ideas and technologies, a new situation is envisaged. Support is developed that should change the current situation into the desired situation. This is a design step, the results of which can not necessarily be derived directly from an analysis of the current situation.

The introduction of support in the current situation is a change in its own right that has an effect: a new situation is created. The ‘old’ Reference Model is no longer valid. The introduction of the support will and is indeed intended to influence the factors in the Reference Model in order to realise the desired situation. Some effects, however, might have been unforeseen. These side-effects might be negative as well as positive. Side-effects that have a negative impact on the outcome are, *e.g.*, the time required to apply a support or the method used to introduce the support. Examples of positive side-effects are a temporary increase of motivation due to the introduction of something new and increased attention for the problem at hand. Side-effects can obscure the effects of the support: the evaluation could suggest a more positive effect than the support actually has, or even suggest a positive effect, where the actual effect is nil or negative. Because of the many facets involved in design (see Figure 1.1) positive as well and negative side-effects are likely to occur during the evaluation and when the support is actually used.

The context in which the development process takes place changes, irrespective of the introduction of design support: people learn, markets change, organisations evolve, new technologies emerge, new knowledge becomes available and new regulations are put in place. As a consequence, the original needs of the designers may no longer exist, the support’s concept might be overhauled, the acceptance of the support may change and with it the ease of introduction. For example, some years ago Internet technology to implement design support would have been regarded with scepticism and would have required more resources for training and for building an infrastructure. This is no longer the case.

Evaluation of design support in itself is difficult because:

- the level of implementation of the Actual Support usually does not cover all functionalities of the Intended Support (as described in Chapter 5);
- the effects of more heuristic support, such as guidelines, methodologies or approaches can be difficult to assess;
- it can take time until the effects occur if proper application requires learning and de-learning processes, and changes of mind set and working habits;
- design processes can be long compared to a research project so that the actual success-related effects might not emerge until many years after introduction;
- the expected outcome, *i.e.*, the desired situation, depends not only on whether the support is functioning (as evaluated in PS), but also on the validity of the description of the current situation and its problems, the quality of the description of the envisaged situation, the conceptualisation of the support, the realisation of the concept, its introduction into the current situation (which may require training, specific resources, organisational changes and customisation), its users and use, its application and maintenance.

These issues emphasise the need to focus not only on the outcome, but also on the process of applying the support in order to be able to interpret the evaluation results – and thus determine the true strengths and weaknesses of the support – and to suggest improvements.

6.1.2 Types of Evaluation in DRM

In DRM, evaluation of design support takes place in the PS and DS-II stages. The evaluation in the PS stage was labelled *Support Evaluation* (Section 5.8). In DS-II two types of evaluation are distinguished: *Application Evaluation* and S*uccess Evaluation*, with Success Evaluation being the most comprehensive. Support Evaluation is a pre-requisite for Application Evaluation and can support the explanation of results. Similarly, Application Evaluation is a pre-requisite for Success Evaluation. Figure 6.1 illustrates the focus of the three types of evaluation, using the Impact Model shown in Figure 2.8.

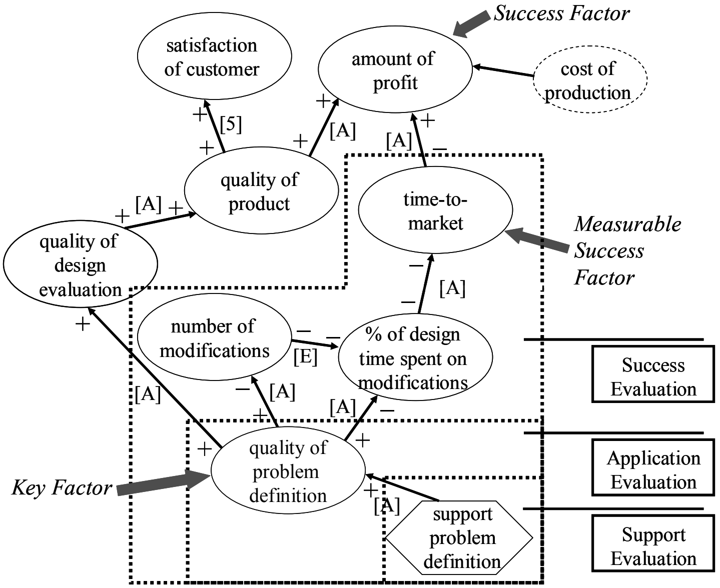


Figure 6.1 The focus of the three types of evaluation using the Initial Impact Model in Figure 2.8

As described in Section 5.8, **Support Evaluation** involves continuous testing during the development of the design support to ensure that the Actual Support is developed to such an extent that it can be evaluated in DS-II. The Actual Support differs from the intended support (see Section 5.7.2). This affects the evaluation methods that can be used in DS-II and the conclusions that can be drawn with respect to the Intended Support.

The first evaluation in DS-II is the **Application Evaluation**, which aims at assessing the applicability and usability of the support against the desired values of the Key Factors. The following questions are addressed:

- Can the support be used?
- Does the support indeed address those factors it is supposed to address directly (the Key Factors)?
- Are these Key Factors affected as expected?

In the terms used by Rossi (1999) this type of evaluation assesses the *proximal outcomes*, that is, the outcomes that should be affected directly. In our *reliability example*, Application Evaluation aims at assessing whether and to what extent the developed reliability method can indeed be used to determine the reliability of an embodiment, and whether the use of the method results in a measure of reliability that is correct, *e.g.*, against values obtained from actual testing of final products. As shown in the Actual Impact Model in Figure 5.13, this involves assessing the resulting 'accuracy of CI diagrams', and the improvement in the knowledge of clarity and simplicity levels of the designer, the resulting 'quality of modification', and how this influences the Key Factor 'reliability of embodiment'.

Success Evaluation aims at assessing the usefulness of the support, *i.e.*, how successful the support is in achieving the formulated aims. The following questions are addressed:

- Does the application of the support have the desired overall effect as defined by the Measurable Success Criteria?
- Does the application of the support result in the desired situation as represented in the Actual Impact Model, *i.e.*, have all factors involved (from Key Factors to the Measurable Success Factors) achieved their desired values?
- Does the overall effect suggest that the Success Criteria can be fulfilled?

As mentioned in Section 2.6.4, evaluating success is far more difficult than evaluating applicability and usability and the findings are not easy to generalise. Success can only be truly measured in the intended situation, *i.e.*, in practice, and in many instances only in the long term. The concept of Measurable Success Criteria was introduced as a proxy to the Success Criteria in order to address this issue and to enable the assessment of the usefulness of the support to a reasonable extent within the constraints of the project.

In the terms used by Rossi (1999) this type of evaluation focuses on the *distal outcomes*: those that are not directly addressed by the support and are expected to be ultimately affected. In our *reliability example*, Success Evaluation would aim at determining whether the application of the reliability method leads to increased product quality (Measurable Success Criterion), which is something that was considered assessable during the research project. Assuming the correctness of the Impact Model, increased product quality should lead to increased customer satisfaction, which is expected to contribute to market share (Success Criterion). These links, however, are not evaluated.

Usually the Key Factors addressed by the support do not directly influence the Measurable Success Criteria: several other factors are involved. In our *reliability example*, enabling the user to assess reliability in the embodiment stage (desired value of the Key Factor) is expected to increase the reliability of the detail design. This is then expected to increase the product's reliability that in turn is expected to increase the product's quality (the Measurable Success Criterion). Success Evaluation thus involves testing the existence and strength of the various links in the Actual Impact Model starting from the Key Factor (shown with solid lines in Figures 4.9 and 5.13), as well as any other side-effects envisaged, through various intermediate factors, leading up to the Measurable Success Factor – Product Quality.

The evaluation of all factors from Key Factor(s) to those related to the Measurable Success Factor(s) is necessary. If the expected impact is not realised, the network and the link(s) where the network 'breaks down' provide an indication of where the problem may lie and will thus inform improvement. Note that improvement may be required of other aspects than the support, such as the way in which the support was introduced, customised, or the evaluation was set up (see Section 6.5.1).

Application and Success Evaluations may be based on the same empirical study. In many situations, however, different studies are required because the research methods needed to address the different factors and links that are involved cannot be combined in one study. Success Evaluation may also require multiple studies to evaluate the different parts of the Actual Impact Model in order to address the Measurable Success Criteria.

6.1.3 Synthesis Example

At this point we summarise the results on the *synthesis example* introduced in Chapter 5. The results of the RC and the Review-based DS-I stage are illustrated in the Reference Model reproduced in Figure 6.2 (as discussed in 'synthesis example' in Section 5.5.2 and Figure 5.6).

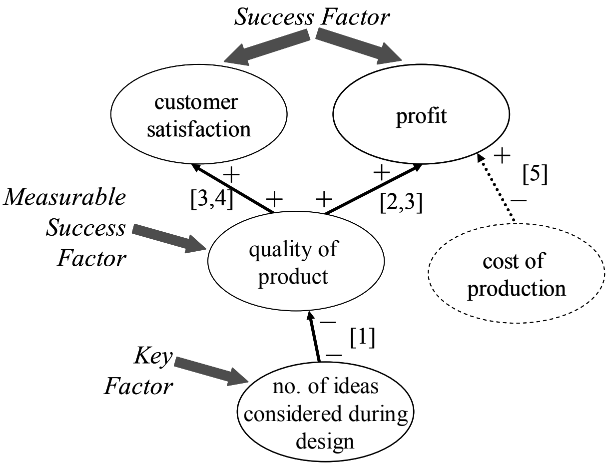


Figure 6.2 Reference Model for the synthesis example (simplified) (see Figure 5.6)

The researcher envisages the desired situation as one in which a support is available to help designers consider a wider range of ideas in design. This should help achieve better-quality products, which in turn should improve customer satisfaction and amount of profit. The Impact Model with the alternative supports considered is reproduced in Figure 6.3, see Section 5.5.2 for more detail.

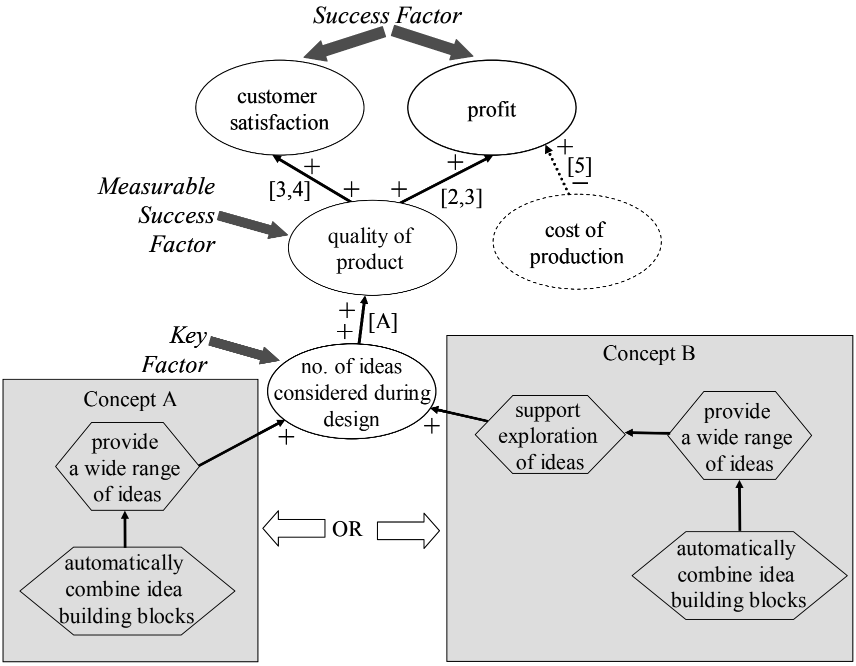


Figure 6.3 Initial Impact Model with alternative support concepts for the synthesis example (see Figure 5.8)

Concept A was developed, as this alternative is simpler and is part of alternative B. To generate a wide range of ideas, Concept A was developed into a software tool consisting, amongst others, of a database of basic building blocks, and an algorithm that generates an exhaustive set of solutions by combining these building blocks. As described in Section 5.8 the *Support Evaluations* were successful.

The researcher decides to start with an Application Evaluation: if the support cannot be applied, there is obviously no merit in evaluating its success, that is, in undertaking a Success Evaluation. The *Application Evaluation* takes place in a laboratory setting involving a comparative study with two groups of designers, those using the Actual Support (the experimental group) and those using pen and paper (the control group). The questions are whether the designers can use the support and whether the number of ideas *considered* by the designers using the support is higher than by those not using the support (desired value of the Key Factor). It was found that designers could use the tool without any problems, but that the number of ideas generated by the tool was too high for them to consider, let alone evaluate. Given the time constraints under which designers have to work, this problem would only be aggravated in a practical setting. The outcome of the Application Evaluation was thus negative. The use of the support did not have the expected effect because of a side-effect: widening the range of ideas presented also increases the time required to go through these ideas. As a consequence, the range of ideas actually considered only covered a small, rather randomly chosen subset of the solution space. The researcher had not explicitly included the factor ‘time required' in the evaluation, as it was not part of the Impact Model. It had not been considered during DS-I and hence not been included in the Reference and Initial Impact Models. Furthermore, the time required and any other side-effects related to the process of using the support had not been considered during support development and thus not included in the Actual Impact Model. As a consequence the required time was not introduced as a factor of interest in the evaluation. Had the effect been smaller, it might have gone unnoticed. This shows that potential side-effects have to be considered at all stages of the research project. The evaluation also points to the fact that the assumption 'the wider the better' that was the basis for the Support Evaluation, has its limits depending on the support provided. It seems that designers *also* need support to *explore* the many ideas in order to be able to consider these.

The researcher decides to revisit the PS stage to improve the support. Possible modifications she considers are the reduction of the number of solutions generated by the system - which runs counter to the overall aim – and supporting exploration. She chooses the latter, as this functionality was already envisaged in Concept B, although this was only to help the designer and not specifically for reducing the time required. She revises Concept B to include a clustering algorithm and a browsing feature (see updated Impact Model in Figure 6.4). The clustering allows the same wide range of ideas to be presented as before, but in a way that is easier to manage. Of each cluster only one representative example is shown. If a cluster seems interesting, the browsing feature can be used to explore the other ideas in the cluster. After a further Support Evaluation to ensure that the support is functioning, a renewed Application Evaluation takes place. This evaluation is successful: designers are now able to consider the whole solution space and within an acceptable amount of time. The range of solutions is no longer experienced as too wide.

Based on this outcome, the researcher undertakes a *Success Evaluation* to determine whether considering this wide range of ideas increases the chances of coming up with a higher quality product, which is the Measurable Success Criterion of the project. The Application Evaluations, though useful as a pre-requisite to Success Evaluation, only show that designers consider a wider range of ideas and can use the support. The Success Evaluation is the actual 'proof-of-the-pudding'.

In order to assess whether 'product quality' has increased, an operational definition is required. Of all components of product quality, the researcher selects 'product novelty' and 'envisaged usefulness' as important contributors to product quality and influenced by the Key Factor 'number of ideas considered during design'. These components are operationalised such that they can be used to assess fulfilment of the Measurable Success Criteria. The Success Evaluation involves a comparative study with designers working individually on two design problems A and B. One half of the designers work on problem A without the support and then work on problem B with the support. The other half does the reverse: working on B without the support and then work on A with the support. The results are positive: when the support is used, the number of ideas considered as well as the quality of the ideas selected is higher.

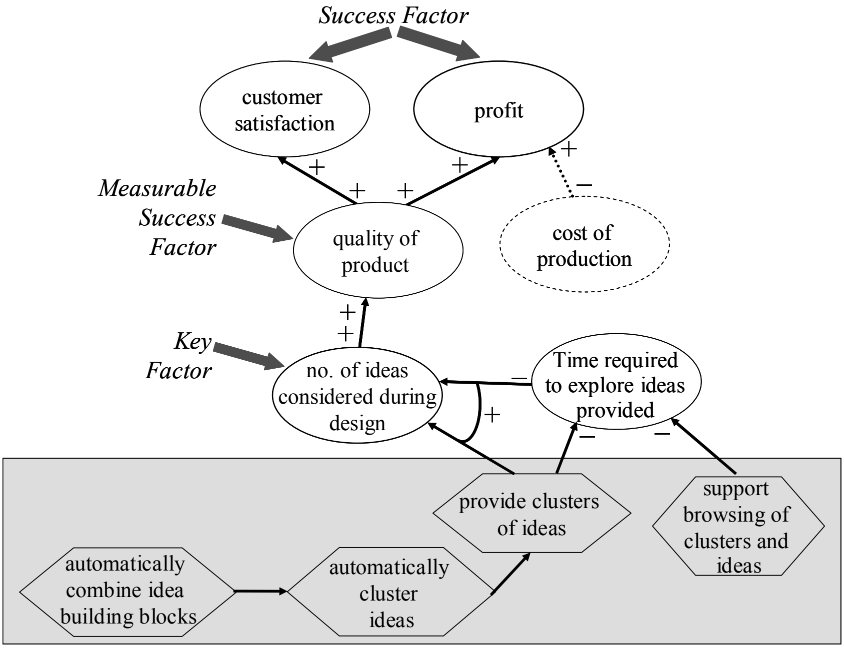


Figure 6.4 Revised Impact Model based on concept B in Figure 6.3

As the example shows, the Application and Success Evaluations inform about the next steps to be taken based on a comparison of the findings of the evaluations with the Impact Model (the desired situation). The results may be satisfactory and lead to the finalisation of the research project, the detailing of further functionality of the Intended Support, or even the commercialisation of the support.

If the evaluation is unsatisfactory, the results would first lead back to the PS stage to modify the developed support. However, as the example shows, the results can also increase our understanding of designing with the support. It may be necessary to modify the Impact Model, which affects the *type* of support required. Moreover, the evaluation may also increase our general understanding of design and reveal issues as well as factors and links that were not considered during the DS-I stage and thus not represented in the Reference Model upon which the Impact Model is based. In such a case, the DS-I stage might have to be revisited to further investigate the existing situation and to modify the Reference Model.

6.1.4 DS-I Versus DS-II

DS-I and DS-II both involve empirical studies to investigate particular phenomena but with different aims. DS-I aims to understand an existing situation to provide suggestions for support to improve this situation. DS-II aims at understanding a situation in which a support is introduced, in order to assess its ability to improve the existing situation and to provide suggestions for improvement of the support. The difference in aims leads to the following differences in approach.

DS-I focuses on understanding the current situation and therefore, generally, does not involve intervention[16] by the researcher: what is studied has not been deliberately influenced or introduced by the researcher. DS-II is always based on intervention: design support is introduced by the researcher or by others and the aim is to evaluate this support to suggest improvements.

DS-I may be based on hypotheses derived from the literature, but will often be focus on answering research questions. In DS-II hypotheses will be available, namely those describing the intended effects of the support to be evaluated, based on the aims and functions of the support and on the Impact Model developed in PS.

In DS-II causal relationships have to be identified to be able to determine the effect of the support. This implies that DS-II is likely to involve explanatory research and comparative studies as these provide much stronger evidence of causal relationships. In DS-I, exploratory or descriptive research can be a good starting point, although eventually a link between findings and success has to be established to determine which factors, when they are addressed, are mostly likely to improve the existing situation.

6.1.5 Existing Evaluation Approaches

Researchers in various disciplines develop, introduce and evaluate solutions for the problems they have identified through their investigations. These can be as varied a software tools, technologies, design methods, new curricula, medical treatments, change management programmes, diagnostic procedures, social programmes, standards, *etc*. Various ways of evaluating (often also called validating) the introduced solutions are used, some more established than others. The approach used depends to a large extent on what is to be evaluated and whether this is introduced specifically for the evaluation. The evaluation of a curriculum, *e.g.*, will require a different approach to that for the evaluation of an e-learning demonstrator for use by a single user. The evaluation of an established curriculum will also be different from that of a new curriculum introduced by the researcher/evaluator. When reading the literature on evaluation, it is necessary to understand the arguments and assumptions used by the authors and to choose the right approach to address one's own evaluation aims. The main approaches upon which we based this chapter are summarised below.

Evaluation Research

Literature on what is called 'evaluation research' usually concerns research into the effects of social, community, public health, or curriculum programmes. Several journals and organisations exist with websites where examples of evaluation

---

[16] It can be argued that most types of investigation will interfere with the normal situation, in particular when a change of environment is introduced such as a laboratory study. In the context of DRM, we do not speak of an intervention, when the factors *of interest* are not expected to be influenced by the chosen setting; when these factors are influenced, but the effects are known; or when, in a comparative study, all cases are affected in the same way.

procedures can be found (such as www.europeanevaluation.org and www.evaluation.co.uk). Although the types of programmes are different from those with which we are concerned, the procedures are relevant to design research where comprehensive or less tangible means of support are to be evaluated, such as new ways of working. It is the evaluation of these types of design support that is particularly difficult. For example, approaches such as concurrent engineering or knowledge management that involve and affect large parts of an organisation and are essentially heuristic in nature are much more difficult to evaluate than a software tool for individual use that automates certain design activities. For the latter, software engineering literature provides several detailed approaches. For the former, social research approaches are suitable, as they focus on evaluating organised social behaviour. Several facets of design are instances of such behaviour. This section focuses on these approaches.

Evaluation research distinguishes two main approaches. The *classical* evaluation involves an experimental, hypothetical-deductive approach. Hypotheses are formulated, variables determined, tests developed and outcomes measured to accept or reject the hypotheses. The collected data tend to be quantitative. Authors such as Patton (2002) and Blalock (1999) point out that these are performance measures focusing on outcomes, not on why and how. They favour *naturalistic enquiry* for evaluation to better understand the situation and context as it occurs. Naturalistic enquiry is central to fieldwork and aims to understand realities and details by being close to the people and situations being studied (Patton 1987). The approach is inductive in that variables are not pre-determined. Rather than using hypotheses, questions are used and patterns emerge from the observations. The approach is clearly data-driven. Qualitative as well as quantitative data can be collected, but a preference exists for qualitative data.

Evaluation research also distinguishes between *formative* evaluation to provide information for programme improvement and *summative* evaluation to determine whether a programme is effective and should be continued. Both can be objectives of evaluation in design research.

In design research two situations can occur. The evaluation can focus on existing support or on newly developed support. An example of the former is a project on the actual use and effects of the systematic design approach introduced several years earlier in a company. This is the sort of project Patton is referring to and for which naturalistic enquiry seems very appropriate. More common in design research, however, is the evaluation of support that is introduced by the researcher or his/her group and only developed up to a certain extent. It is the change in the existing situation that we wish to measure. Although the premises of a naturalistic enquiry do not hold, we agree that it is essential to know whether, but also why and how support is successful or not successful, if improvement of the support is to be effective, efficient and satisfactory. Qualitative methods are necessary, as these allow one "to see possibly more than standard tests and measures may tell" (Patton 1990). "Quality is what separates and falls between those points on a standardised scale". In evaluating design support it is one thing to know that someone considers a new design method satisfactory by ticking the appropriate box in a questionnaire. It is another thing to know *why* this person is satisfied with the method and with which particular features.

Despite the advantages of empathetic understanding to better understand how those involved experience the support, their objectivity and the trustworthiness of human reports are often debated. In particular for the evaluation of methods and approaches introduced by the researcher, empathetic understanding can be problematic because of the closeness of the researcher to what is being evaluated. This can bias the researcher as well as the participants involved, who might try to be 'nice'. Whether using an emphatic approach or not, every method must be applied correctly, as "closeness does not make bias and loss of perspective inevitable; distance is no guarantee of objectivity" (Patton 1990).

Depending on the type of the Actual Support, the factors to be evaluated, the research questions to be addressed, the hypotheses to be verified, and the theoretical understanding that exists, choices have to be made about the evaluation approach: qualitative, quantitative or a combination, experimental or naturalistic, formative or summative, data- or theory-driven. That which we want to know should guide the Evaluation Plan, not the paradigm (after Tashakkori and Teddlie (1998)). Patton (1990) argues that a qualitative-naturalistic-formative approach is especially appropriate for programs that are developing, innovative or changing, where the focus is on program improvement, facilitating more effective implementation, and exploring a variety of effects on participants. Blalock (1999) provides a useful overview of the various research designs, their primary purpose, the kind of issues for which each design is most appropriate, and the most suitable research methods.

Several authors emphasise that evaluation should not only focus on an assessment of the outcome, such as the reliability of the embodiment, as the impact on this factor might be due to a variety of factors other than the support itself. The systematic approach for evaluating social programmes[17] proposed by Rossi *et al.* (1999) takes this into account. We found this approach to be very similar to our own approach for evaluating support. They too emphasise that there is no one style or paradigm[18] and allow the integration of various methods.

Furthermore, Rossi *et al.* define what has to be available – apart from the programme to be developed – before a proper evaluation can take place. This overlaps with our deliverables of PS (for a comparison see Section 6.4 and Figure 6.6). According to Rossi *et al.* *programme theory* is the set of assumptions that relate the programme to the benefits it is supposed to realise and the strategy and tactics used to achieve the goals and objectives. A programme theory consists of a process theory and an impact theory. The *process theory* consists of the organisational plan describing the required resources in the organisation, and the service utilisation plan describing how the target population receives the intended amount of the intervention. The *impact theory* describes the chain of causes and effects that show how the intended intervention brings about the benefits.

---

[17] In Rossi *et al.*'s original work the spelling is 'program'. The spelling has been changed for consistency with the rest of the book, which uses UK English spelling.

[18] Their short discussions about scientific versus pragmatic evaluation postures (p.29) and the diversity in evaluation outlooks and approaches (p.32) are worth reading.

Their approach is divided into five types of what they call conceptual and methodological frameworks corresponding to the types of common evaluation questions:

- Needs assessment: answers questions about the social conditions that a programme is intended to address and the need for the programme.
- Assessment of programme theory: answers questions about programme conceptualisation and design, such as whether the stated goal and objectives do relate to the conditions that the programme is intended to improve, or whether the assumptions underlying the programme represent a credible approach.
- Assessment of programme process (or process evaluation): answers questions about programme operations, implementation, and service delivery.
- Impact assessment: impact evaluation or outcome evaluation: answers questions about programme outcomes and impact.
- Efficiency assessment: answers questions about programme cost and cost effectiveness.

Program evaluation, according to Rossi *et al.*, involves assessment of one or more of the above. In our approach, we consider all these assessments essential and encourage their inclusion at least in the planning of the evaluation, to ensure that their impact on the outcome can be controlled or at least understood, even though the focus of the evaluation, *i.e.*, the actual assessment, may be only on some of the assessments mentioned above.

Action Research

An approach that is fairly similar to ours and provides very useful methods is Action Research. An example is the Soft Systems Methodology of Checkland (1999) described in Section 2.7. In contrast to the other approaches, evaluation as well as development is part of the Action Research approach. Action Research provides a set of research methods that aim at alternating action (development, implementation and introduction) with critical reflection (evaluation). Its evaluation aspect is formative. Action Research is usually qualitative and participative, and can be seen as a form of rapid prototyping.

In terms of our methodology, Action Research corresponds to a cyclic process involving the PS and DS-II stages after an initial DS-I stage. We agree that iterations between PS and DS-II always take place, but whether a series of small iterations or a few large iterations is more suitable depends on the type of support developed and on our understanding of the current and desired situations. The main difference between Action Research and DRM is that Action Research focuses on obtaining conclusions about specific support (often a programme or an approach) in a specific situation. The aim is to gradually improve the support for use in that situation until a full, optimised implementation is achieved. DRM in supporting design research aims its evaluation stage at obtaining generic statements about partial implementations. This difference in aim (individualisation or generalisation) is important when consulting the literature to formulate an Evaluation Plan, because it does affect the choice of approach. Patton, *e.g.*, recommends qualitative case methods when individualisation rather than generalisation of the outcome is the aim.

Some of the differences between the approaches of Checkland and Rossi *et al.* are of interest for the choice of evaluation approach, although the borderlines between the two approaches are not always clear. Checkland focuses on the evaluation of situations that are problematic. His approach is about change, about understanding the problem and finding solutions or alleviations. Rossi *et al.* focus on the evaluation of specific programmes that have been introduced. The situation might not be problematic; on the contrary, the evaluation may show how good the situation is. Both approaches include an assessment of the current situation: Checkland through a comparison with an imagined, future, better situation within the company, but not within other sites or groups; Rossi *et al.* through a value-judgement that could include a cross-site or cross-group comparison. Both Checkland and Rossi *et al.* do field research. Checkland's approach, however, is essentially ethnomethodological and data driven, whereas Rossi *et al.* follow a more quasi-experimental, hypothesis-led approach. Checkland actively involves problem-owners in the evaluation and researchers in the change process (a participative approach), whereas Rossi *et al.* take a more distant position, not intending to change the situation. Checkland's approach is inherently cyclic; that of Rossi *et al.* essentially linear (although iterations will obviously take place) as their approach does not include the improvement of the support and the subsequent round of evaluation. As soon as Checkland's approach has produced and implemented a change and this is then to be evaluated (note that Checkland does not explicitly mention evaluation), the aims of the two approaches become very similar.

Software Evaluation

Software development is another area where evaluations, in the form of, *e.g.*, α- or β-testing, are very common and reasonably well described, although the approaches tend to refer to the development and evaluation of commercial systems, rather than proofs-of-concept (see also Appendix B.2). Approaches to evaluate user interfaces can be found in Human Computer Interaction (HCI) literature (see Appendix B.3). The focus of these approaches is mainly on usability, which is divided into effectiveness, efficiency and satisfaction.

Design Research

The difference between the types of evaluation used in disciplines such as social science and computer science, and those used in design research, is that the former usually evaluate support that is realised to its full extent and introduced into the field for real use, or at least intended to be so. As discussed in the previous chapter, in most design research projects only demonstrators or concepts are available. The project and thus the evaluation is aimed at obtaining a proof-of-concept. The literature does not address how to select those elements of the support that, when realised, are sufficient to evaluate the concept. The fact that a selection usually has to be made is the reason why DRM distinguishes between Actual Support and

Intended Support and why it emphasises the development of an Outline Evaluation Plan in parallel with the realisation of the support.

6.2 Types of DS-II

There are two types of DS-II: an Initial and a Comprehensive DS-II (see Section 2.3, Figure 2.2). Both cover all steps in the DS-II process. The Outline Evaluation Plan developed in PS only covers the initial steps and is the basis for an Initial as well as for a Comprehensive DS-II.

6.2.1 Initial DS-II

When a PS has been undertaken, a full evaluation might not be possible for two reasons. First, the support is developed and realised, but the project duration does not allow a full evaluation. Second, the support has not been developed and realised to the extent that the support can be applied by future users as intended, but the researchers want to be able to evaluate at least some of its applicability, usability and if possible, usefulness. As we argued in Section 2.3, at least an initial evaluation is required to round off a research project, that focused on the development of support to be able to draw any conclusions about the relation between the support and the aims of the research project. This type of research project corresponds to the third type of project shown in Figure 2.2. Minimally required are:

- an indication of the applicability, usability and usefulness of the support;
- an indication of the issues, factors and links that need detailed evaluation;
- a suggestion for a proper Evaluation Plan.

This is best done through an Initial DS-II focusing on Application Evaluation. The link to the Measurable Success Criteria is argued by using the literature. An Initial DS-II contains all steps of a Comprehensive DS-II, but to a lesser extent or in lesser detail.

6.2.2 Comprehensive DS-II

For a proper evaluation that can be used to assess the effects of the proposed support and inform further development, a Comprehensive DS-II is required. DS-II may be part of a larger research project or follow up another research project in which design support has been developed and an extensive evaluation is required. DS-II may also be used when serious doubts exist about the success of existing design support. Examples of the latter are research into reported difficulties in: implementing the concept of concurrent engineering; using QFD in its entirety; composing design teams using methods for selecting team members based on personal characteristics, or the effectiveness of brainstorming.

A significant difference between the two situations is that when the evaluation is part of a larger project, detailed information about the support is available and the researcher or the research group is involved in the actual introduction. This is less likely when the evaluation involves existing support. The danger is that "where outcomes are evaluated without knowledge of implementation, the results seldom provide a direction for action because the decision maker lacks information about what produced observed outcomes" (Patton 1987). For this reason, the DS-II process we propose emphasises the need for reviewing the existing documentation on the support and to ensure that the documentation is as complete as possible before undertaking the evaluation.

The description of the DS-II process in the rest of this chapter focuses on a Comprehensive DS-II. To which level of detail the steps have be executed in an Initial DS-II depends on the aim of the research project.

6.3 Systematic DS-II Process

The systematic approach we propose for the planning and execution of a DS-II is shown in Figure 6.5. Iterations between steps will occur, and even more so between the activities within a step as many of these are done in parallel. The first two steps (reviewing existing documentation and determining evaluation focus), as well as the results of each individual evaluation study will determine the number of studies required and thus the number of times the various steps in the process are undertaken. Continuous documentation of steps, results and arguments is necessary as this will constitute an important basis for the publication of the evaluation results.

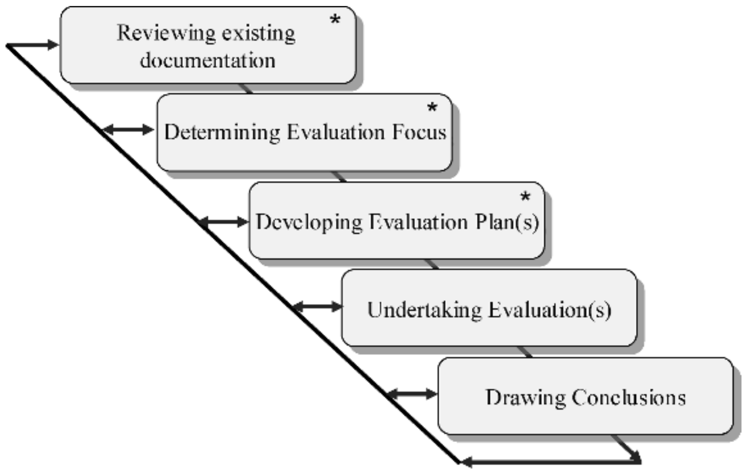


Figure 6.5 The main steps in the DS-II stage, stars (*) indicating steps that start during PS to develop an Initial Evaluation Plan

The steps in our systematic approach are the following.

1. Reviewing existing documentation. This involves collecting and analysing existing documentation, results from DS-I and PS, results from any

evaluations that have already taken place, and the Introduction and Outline Evaluation Plan(s) if available. Missing information has to be tracked and missing documents generated.

2. Determining evaluation focus. This involves reassessing the Outline Evaluation Plan, selecting the factors and links to be addressed, (re)formulating the research questions and hypotheses, and selecting the type(s) of evaluation.
3. Developing Evaluation Plan(s).[19] This involves developing measurements and setups, selecting and developing research methods and their combination in one or more studies, adjusting the Actual Support and the Actual Support Documentation to match the Evaluation Plan, developing the necessary material, undertaking one or more pilot studies, and adjusting the Evaluation Plan, the research methods, the setup and the materials used.
4. Undertaking evaluation(s): this involves collecting, processing, analysing and interpretating the data, verifying the results, drawing conclusions, comparing the results with the Actual Impact Model, updating this model, deciding on further empirical studies within the project if not already planned, and documenting deviations from the Actual Introduction Plan and Evaluation Plan.
5. Drawing overall conclusions: this involves combining the results of the different evaluation studies, identifying the net effects, the effects of the processes and the effects of the research, completing the Actual Impact Model, reflecting on the Evaluation Plan, determining the consequences for the Actual and Intended Support, for the Actual and Intended Introduction Plan, for the Intended Impact Model, for the support concept and underlying assumptions, and for the Reference Model, as well as determining the next stage.

As discussed in the previous chapter, the choice of functionality realised in the Actual Support and the components of the Evaluation Plan are closely linked. The first three steps (without developing the material and doing the pilot study) should therefore already start during the PS-stage to develop an Outline Evaluation Plan, which is then elaborated in the DS-II stage into an Evaluation Plan. The steps will be described in more detail in the remainder of this chapter.

6.4 Reviewing Existing Documentation

To ensure an efficient, effective and meaningful evaluation more is required than the availability of the Actual Support and and Outline Evaluation Plan. Additional information is necessary to understand the support and choose the right focus and methods for the evaluation. This is the reason why a variety of deliverables are required from the PS stage as input for the evaluation. These are listed in Figure

[19] We opted for the term ‘evaluation plan’ rather than the commonly used term ‘evaluation design’ to avoid confusion with our domain ‘design’.

6.6, which also shows the terminology used by Rossi *et al.* discussed in Section 6.1.5.

Evaluation input

- Actual Support
  (Rossi: Programme)
- Actual Support Documentation
  (Rossi: Programme Theory)
  - Actual Introduction Plan
    (Rossi: Process Theory)
  - Actual Impact Model
    (Rossi: Impact Theory)
  - Actual Support Description
- Intended Support Documentation
  - Intended Introduction Plan
  - Intended Impact Model
  - Intended Support Description
- Reference Model

Figure 6.6 The input required for planning an evaluation, including a comparison with the input proposed by Rossi *et al.* (1999)

Apart from the Actual Support itself, Support Documentation has to be available. Support Documentation consists of an Introduction Plan describing the introduction, customisation, use, maintenance and use context (including the user) required for the support to realise its impact, as well as the Impact Model. The Support Description (not in Rossi *et al.*) describes the support in terms of the need or problems addressed, the goals and objectives of the support, its elements, how it works, the underlying concept, theory, assumptions and rationale; and how it is realised (software, workbook, *etc.*).

As discussed in Section 5.7, the Actual Support is likely to differ from the Intended Support affecting not only its functionality, but also the Introduction Plan, the Impact Model and the Support Description. New versions for each of these are created in the PS stage for the Actual Support. Although the evaluation involves the Actual Support, we also wish to be able to draw conclusions about the Intended Support. For this reason we require *two* sets of documentation for the evaluation: the Actual Support Documentation and the Intended Support Documentation.

The differences between these two sets reveal the differences between the Actual and the Intended Support. This will determine which statements about the applicability, usability and usefulness of the Actual Support might apply to the Intended Support and thus determine which of the factors and links that *can* be evaluated are *useful* to evaluate. Take for example the case where paper rather than software is used to realise the Actual Support. The Intended Support aimed at supporting designers storing what they document in a new structure in the Intranet.

Functionalities such as ‘cut’ and ‘paste’ cannot be evaluated when using a paper realisation; hence nothing can be said about their influence. The ‘user interface’ can be assessed, but the outcome probably does not say much, if anything at all, about the intended type of interface. The core functionality, however, might still be assessed: do users understand the categories used in the new structure, can they assign their documents to these categories, are the categories unambiguous, are any categories superfluous or missing, *etc*.?

Apart from the documentation of the support, the Reference Model is needed for planning the evaluation, as this model describes the situation that gave rise to the development of the support and is likely to be the situation in which the support is to be introduced. If an Outline Evaluation Plan has already been developed, *e.g*., in a previous project, this is also collected.

The careful preparation of the documents during the PS stage helps to understand the Actual Support, the assumptions underlying the support and the way it is supposed to be used (What, Why and How). This is essential for the selection of the factors and links to be evaluated, the formulation of the criteria to assess the outcome, the formulation of research questions and hypotheses, the identification of the outcomes that are reasonable to expect, and the identification of the limitations caused by the differences between the Actual and Intended Support.

In particular when the support has not been developed by the evaluators, the documents provide invaluable information. It is therefore important to determine whether any information or documents are missing from the existing documentation. If this is the case, attempts have to be made to complete the set to avoid an evaluation that is useless because it was based on the wrong premises. This might require the reconstruction of essential information such as Impact and Reference Models, or details of the Introduction Plan. If possible, the original developers should be involved in this process.

6.5 Determining Evaluation Focus

In order to determine the focus of the evaluation, it is necessary to consider the various aspects that can influence the outcome of the evaluation, other than the Actual Support. These will affect the type of evaluation(s), the factors and links that are to be addressed, as well as the set of research questions and hypotheses.

6.5.1 Aspects to be Considered

The Actual Impact Model describes the expected impact of the Actual Support *provided* that it has been developed and used as intended. Despite all best intentions, this might not be the case:

- the support might not address the real needs of the users or the needs at the time of use, but instead address obsolete or perceived needs;
- the theories and assumptions behind the concept that link this concept to the expected benefits might be incorrect or unsuitable;
- the concept might not be well thought out;

- the support might not have been realised correctly or using means (training, workbook, software, *etc.*) that cannot reveal the expected impact;
- the support might not have been introduced correctly or using unsuitable means, *e.g.*, insufficient training or unclear introduction;
- the changes to the working environment caused by the introduction itself might not have been considered;
- the support is used in a context that is dynamic and where politics have a large influence;
- the competencies of the users are not appropriate;
- the users' preferences, beliefs, interests and motivations are not in line with the working and goals of the support;
- the available resources such as people, time and equipment are insufficient;
- the available environment (organisational and technical infrastructure, help service) is unsuitable.

The evaluation should therefore not focus only on measuring the factors and links in the Actual Impact Model and assume that the measured effects are due to the support only. As the above list shows, several aspects have to be considered in the evaluation plan to ensure that the net effect can be assessed, that is, that the results can be attributed exclusively to the support rather than to other influences or to chance. These aspects are the following:

- need of the users;
- conceptualisation of the support and the underlying assumptions;
- the Actual Support;
- introduction;
- impact: desired and undesired, indirect and direct, immediate and long term;
- efficiency;
- user competencies, preferences, beliefs, interests and motivations as well as behaviour during use;
- organisational, technical and other contextual pre-requisites.

Considering these aspects does not imply that all have to be assessed as part of the evaluation. Depending on the features of the support, the purpose and focus of the evaluation and the context, some of the aspects may not play an important role. Nevertheless, thought should go into the possible effects of the support on these aspects and the effect of these aspects on what the support is intended to achieve. Many aspects can be dealt with by a careful selection and preparation of the evaluation. This may require additional actions, such as developing a training exercise or creating the appropriate infrastructure. A typical example is the introduction of a new support. The quality of the introduction does influence the use of the support and thus the outcome of an evaluation. An appropriate introduction to the support might have to be developed for use in the evaluation and at the same time, questions about the introduction might be included in the evaluation itself.

Other aspects that have to be considered are those related to the evaluation methods that are used. These have been discussed in Chapter 4 and Appendix A.

6.5.2 Determine Focus

The available documentation and the list of aspects to be considered provide a boundary around the space of possible DS-II studies. The aim of the evaluation and the available time and resources will provide the final focus.

We suggest starting with the aim of the evaluation: what is expected to be achieved by the evaluation in order for it to provide a contribution, that is, to be academically and possibly practically worthwhile. It is important to focus on the *core*, *essential* and *unique* features of the Actual Support and to evaluate these thoroughly, rather than to try to evaluate everything, potentially ending up without any strong evidence.

The evaluation may have several aims, which may require multiple studies. With this in mind, the Actual Impact Model and Actual Introduction Plan are analysed to identify the set of factors and links that should ideally be addressed. This set is assessed against the features of the Actual Support and Actual Introduction Plan as well as the Intended Support and Intended Introduction Plan to verify that addressing the particular factor or link is worthwhile. This set is furthermore assessed against the available time and resources to verify that the factors and links can be addressed at all. Indications about possible or necessary evaluation methods that emerge should be documented.

The resulting set of factors and links should include at least one complete chain between the Support and the Measurable Success Criteria, although some paths in the Actual Impact Model might not be included, as the Actual Impact Model of our reliability example in Figure 5.13 shows. Note that if the researchers were involved in the development of the support, this comparison should have taken place during the PS stage as part of the co-development of the support and the Outline Evaluation Plan (see Section 5.7.1).

Once the initial set of factors and links have been selected, the aspects (see the above list) that may potentially influence these are determined, paying particular attention to the effects of the features of the Actual Support and the Actual Introduction Plan. Finally, the main research questions and hypotheses are formulated that cover the factors, links and aspects. This process is iterative: formulating research questions and hypotheses may give rise to the need to include other factors, links and aspects. At this point in time, it will not be possible to determine the relevance of all aspects as their potential effects might relate to the chosen evaluation method(s). The exact focus of the evaluation will therefore only be clear once the Evaluation Plan has been developed. This may require the adjustment of the set of factors, links, aspects, research questions and hypotheses while remaining consistent with the chosen Measurable Success Criteria.

6.6 Developing Evaluation Plan(s)

Planning the evaluation of design support is a complex, challenging task that requires creativity and involves careful preparation in order to obtain meaningful results. The various factors and links to be addressed, the aspects to consider and the chosen criteria result in a variety of research questions and hypotheses that are likely to require a combination of approaches and methods, either in parallel or in sequence. Hypotheses can exist about the effects of the support, and questions can exist about the effects of certain aspects. Quantitative methods might have to be used to measure certain effects and qualitative methods to understand the reasons why. The challenges are "to match the research procedures to the evaluation questions and circumstances as well as possible, and to apply them at the highest possible standard feasible in those circumstances" (Rossi *et al.* 1999). In doing so "evaluators must often innovate and improvise as they attempt to find ways to gather credible, defensible evidence".

If several methods and approaches are required, it might be necessary to develop several detailed Evaluation Plans. We suggest starting with the development of one Outline Evaluation Plan, preferably during the PS stage. The Outline Evaluation Plan suggests the number of studies, their focus and how they are linked. The latter is important for the operational definitions used to assess the factors, links and aspects (see below). If the results of two studies have to be linked, it is necessary that the same operational definitions are used. In our synthesis example, the ideas produced by the system were defined in terms of the building blocks upon which they were based. The ideas produced by the designers were represented by the researcher in terms of the same building blocks. Only then was it possible to compare both number of ideas and type of ideas.

For each study in the Outline Evaluation Plan, a detailed Evaluation Plan is developed. Depending on the situation, it might be useful not to fully detail all the Evaluation Plans at the start of the DS-II stage, but to subsequently detail each plan to take into account the results of the previous evaluations. The outcomes might be negative, making further evaluation futile until the support has changed. The outcomes might also be unexpected, suggesting other evaluations than were originally planned.

The order in which the evaluations are to take place can be relevant in the light of potential negative and unexpected outcomes. We suggest starting with Application Evaluations before undertaking the usually more extensive Success Evaluations.

Selecting a particular approach is taking a particular view. When evaluating a support one has developed oneself, one is very vulnerable to taking a view that suits the support. If not careful, these views and their related preferences and beliefs as well as those expected from the peer group or examiners can affect the way the evaluation is set up and executed and the data analysed. A careful planning of the evaluation can avoid some of the bias or at least reduce its impact.

6.6.1 Develop Measurements

To prepare the evaluation, *each factor and link* in the Actual Impact Model up and including the Measurable Success Factor(s), as well as any other factors used in the research questions and hypotheses are operationalised by the following:

- a definition of the terms used: for example 'reliability of embodiment' and 'level of clarity' have to be defined;

- the information needed to determine the existence and value of a factor or link, *e.g.*, a formula and specific product data to calculate reliability, or the opinion of an expert on reliability;
- the criteria for accepting a factor or link as existing and successful, *e.g.*, the reliability value required to determine whether the reliability is 'high', or the required improvement ('higher') of the reliability value when using the support, *i.e.*, relative to when not using the support;
- additional information needed to judge and interpret the results, *e.g.*, values for comparison, such as the reliability value resulting from not using the support, or values taken from other benchmarks, such as standards.

While specifying the factors and links it is important to consider any aspects that might affect these factors or links or their measurement in a way that can affect the evaluation results – that is, give rise to alternative explanations. Some of these aspects have already been identified while determining the research focus (Section 6.5.1). For each aspect it has to be determined whether and how their effects can be controlled or taken into account. Each aspect to be considered in the evaluation then has to be specified in the same way as the factors and links.

In order to draw conclusions regarding further development, the issue of success has to be addressed, given the available resources. This implies that the Measurable Success Criteria might have to be given a much narrower definition than commonly used. For example, the quality of the product might have to be narrowed down to the result of an assessment of the embodiment drawing as to whether the depicted product is expected to fulfil the requirements list. The narrowing down of the definition potentially reduces the strength of the conclusions that can be drawn. For this reason it might be necessary to introduce new factors that can provide additional data from which to draw conclusions.

An example is given in Blessing (1994) where an additional criterion was selected at the evaluation stage. The reason was that the Actual Support was not expected to give results for the two Measurable Success Factors – product quality and design time – that would allow strong conclusions to be drawn, because the definitions of both had to be narrowed down to allow measurement within the constraints of the research project. This additional Measurable Success Factor was the 'level of confidence' designers had in the solution they created. This was defined as "the feeling they had, that they could not have come up with a better solution" and measured by means of a questionnaire. Based on the literature, it was assumed that confidence expresses that designers consider a substantial part of the solution space and pay attention to evaluation, both of which are characteristics of successful design processes. In other words, success was now defined as a combination of product quality, design time, and level of confidence. As expected, the evaluation did not show a significant difference in design quality or design time, but the observed difference in level of confidence favoured the proposed support.

The Evaluation Plan should also include the assessment of whether the support was introduced as foreseen in the Actual Introduction Plan, and whether the introduction could have had an influence on the observed impact of the support. This will require additional aspects to be measured.

Apart from the effectiveness of the support, its efficiency and the satisfaction of the users may be important to be considered, even if they are not explicit factors in the Impact Model. A support may have the required effect, but this might be of little use, if the use of the support substantially increases the time required or is not accepted by the users. In design research it is difficult to measure the efficiency of the proposed support because the realisation is usually incomplete. Efficiency of the support, that is the cost against the benefits, however, is an important aspect that will affect acceptance and utilisation of the support in practice. Not only those interested in the support need to be convinced that it is worth the investment, but also those that funded the research. If information cannot be obtained directly from the evaluation, we suggest estimating efficiency based on the evaluation findings and data in the literature.

Rossi suggests two types of measurement: cost-benefit and cost-effectiveness. In a *cost-benefit* calculation, both aspects are usually expressed in monetary value. An example in design research would be a reduction of lead-time of 20% translated into a financial benefit using data available in industry. This is placed against the cost of using the support, including the necessary investments, to determine the overall benefit of using the support. A *cost-effectiveness* calculation determines the cost of achieving a particular result. Cost is again expressed in a monetary value, but the effectiveness is measured in outcome units. Using the above example, the resulting statement could express the cost spent to reduce lead-time by X%. Other examples that are not easy to translate into a monetary value are: cost against average satisfaction, or against number of warranty claims.

6.6.2 Evaluation Plan

The Evaluation Plan for DS-II is similar to the Research Plan developed in DS-I, details of which can be found in Section 4.6. This section highlights the specific characteristics of the Evaluation Plan used in the DS-II stage.

Research Questions and Hypotheses

The aim of the evaluation finds its expression in the research questions and hypotheses that are formulated. If, due to practical constraints, certain factors and aspects cannot be adequately measured, it might be necessary to limit the number of research questions and hypotheses or rephrase these. Care has to be taken that the evaluation still fulfils its aim, that is, allows useful conclusions to be drawn about the proposed support. Nothing is as frustrating as having gathered large amounts of data, only to find out that useful conclusions cannot be drawn because certain aspects were not taken into account.

An example will illustrate the intricate nature of the links between aspects, research questions, hypotheses, Evaluation Plan and practical constraints, as well as the necessity to address all of these together when planning the evaluation. Furthermore, the example highlights that evaluation of the support has to be considered while developing the support.

A tool is proposed to support the variety of communication channels used in design. A full software implementation is not possible within the timeframe of the research project. It is decided to focus on the core assumptions on which the support is based, reduce and rephrase the original research questions and hypotheses accordingly, and realise the support only as far as necessary to be able to evaluate these assumptions. However, the limited functionality of the Actual Support makes it difficult to evaluate the tool in a practical setting, while a laboratory setting does not provide the variety of communication channels that exist in practice. A possible approach is to select another practical setting – for example an industrial environment with very short projects – or to create a suitable setting, such as a laboratory environment that imitates the variety of communications. Another approach is to question the premise that an evaluation involving a smaller variety of communication channels will affect the findings to such an extent that conclusions cannot be drawn about communications channels in general. The question is: is the core contribution of the support the improvement of communication (a human aspect), or is it the variety of channels that can be handled (a technical aspect)? A few channels may be sufficient to evaluate the former. Whereas the latter requires a large number of channels. This decision will affect the functionalities and features of the Actual Support, as well as the evaluation setting, the research questions and hypotheses, the additional aspects to be considered and the extensiveness of the evaluation.

Since the researcher knows exactly what he or she is looking for, namely verifying the assumptions and claims related to the support, it is easier to formulate hypotheses in DS-II than in DS-I. The empirical study can be much more focused. In such studies, bias may be a serious problem. For this reason some authors (*e.g.*, Scriven in Patton (1987)) propose goal-free evaluation. This requires the evaluator to suspend judgement about the goals of the support and to focus instead on finding out what actually happens. This is not easy when the researcher also developed the support, which is often the case in design research.

One possibility to reduce bias is to consider alternative explanations for the findings while planning the evaluation, by reasoning backwards from potential findings. One should ask, for instance, what the causes could be if the stated hypothesis is negated, and what the other causes could be if the stated hypothesis is indeed confirmed. In order to be certain about such alternative causes, additional aspects may have to be measured. Typical examples are the type of product, the experience of the participants, the environment, *etc*. It may also be helpful to think about the questions that may be asked by critics to whom the results are presented. Although not all eventualities can be covered, it is important to think about alternative explanations and to measure more than strictly derived from the claims of the support.

According to Rossi *et al.* (1999), *evaluation questions* must be:

- *realistic*, in the light of what the support is trying to achieve;
- *appropriate*, that is, questions should be consistent with experience in similar supports and none of the expected effects should be based on assumptions that other studies show cannot be held;
- *answerable*, that is, unambiguous terms should be used, observable indicators should be available, as well as relevant criteria and sufficient resources.

The *impact theory* in Rossi's book consists of two sets of assumptions:

- *action* hypotheses related to the direct, or *proximal outcomes*;
- *conceptual* hypotheses related to the indirect, or *distal outcomes*.

We make the same distinction in DRM, but divide the distal outcomes further:

- action hypotheses related to the application, *i.e.*, the desired values of the Key Factors;
- conceptual hypotheses that can be evaluated as part of the research project, *i.e.*, the factors and links up to the Measurable Success Criteria;
- conceptual hypotheses that cannot be evaluated as part of the research project, *i.e.*, the factors and links up from Measurable Success Criteria to the Success Criteria.

Many questions and hypotheses can be formulated; too many usually to be realised within the project constraints. To avoid 'wandering off' into interesting aspects that are not of direct relevance to the research problems at hand, the study needs to be focused. In Section 4.5.4 the following questions were suggested to prioritise the questions and hypotheses:

- 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.

For an evaluation, prioritisation will also be determined by arguments related to the unique, essential and core features of the Actual Support and the specific purpose of the evaluation (see Section 6.5.2).

Selecting Research Methods

The selection of research methods on the basis of the research questions and hypotheses has been described in Section 4.6.

The possibility to draw valid and useful conclusions about the impact of the support depends on the quality of the evaluation and of the methods used. The outcome of the evaluation should be a valid description of the performance of the support such that the impact of the support can be assessed using the formulated criteria. A valid description is one that accurately represents what the support actually accomplishes (adapted from Rossi *et al.* (1999)). It should be "distinct and precise", that is, "sufficiently definite and discriminating for meaningful variations in level of performance to be detected". As to what level of performance is to be detected depends on the formulated criteria and on the factors and links to be assessed. Of course, the level and type of performance that *can* be detected depends on the extent to which the support has been realised. Essential is the selection of a combination of methods and a setup that allow a differentiation between the effects of the support and other aspects (see Section 6.5.1): even if meaningful variations can be detected in the performance, these might be due to a combination of causes. This will be discussed in more detail in Section 6.7.

Usually methods can be used for different evaluation strategies by varying the type of user involvement, the setting and the task. The support can:

- be used by the researcher or the future user in the following ways;
  - the user uses the support;
  - the user is observing the researcher using the support;
  - the user is shown application scenarios or application simulations;
- be used in a contrived but realistic setting, or in a natural setting;
- be used for an artificial but realistic task, or for a real task;
- involve users (experimental group) or users and non-users (experimental and control group).

Each evaluation study involves a combination of these options. Not all combinations are possible. Some are not possible at all, while others are not possible in a particular context or not relevant. Given a particular evaluation aim, each combination results in a set of statements, but of different strength. At the same time, each combination requires a different effort from the user and from the researcher, *e.g.*, in the preparation of the evaluation. Different strategies can be used in subsequent evaluations to address different questions and hypotheses or to evaluate improvements made to the support based on earlier evaluation results.

As to which strategy is the most suitable depends on the type of research questions and hypotheses the support aims to address and on the existing project constraints. In general, hypotheses referring to changes brought about by the support require a comparative study, unless data about the situation without the support is available. Examples of such hypotheses are: 'Designers using the method will generate *more* variants that those that do not use the method. 'The use of the tool will *reduce* the time spent on calculations'. In these cases, the evaluation could involve an experimental and a control group, a pre-post measurement of the same group, or a time-sequence series (see Appendix A.4.3). The selection of a strategy is also linked to the level of realisation of the support. For instance, evaluation in a natural setting involving a real task usually requires the support to be fairly robust and complete. The type of support also has an influence on the strategy: to evaluate a largely automated procedure, rather than a manual procedure, it might not be necessary to involve future users.

In Chapter 4 and Appendix A.4 we discussed various types of empirical studies. Experimental approaches are the most rigorous because they control bias most effectively, but they are considered less suitable for evaluating support that has several (side) effects other than the intended effects. As we have argued earlier, many different aspects are involved in an evaluation, all strongly linked. "Experiments, even field experiments, trade realism for precision" (Rossi *et al.* 1999), but they "may be useful in single programmes in relatively stable environments" (Schmid in Rossi *et al.* (1999)).

A potential practical problem is that, in contrast to DS-I, an evaluation involving users is likely to be more intrusive and requires more effort than DS-1 because a change in the working situation is required: mere observation is not possible. Depending on the required introduction time and the ease of use of the support, considerably more time may be required of the users, increasing the difficulty to find users willing to participate. This will strongly affect the choice of evaluation methods. It is important to involve future users as early as possible in the project, at the latest when the Evaluation Plan is being developed.

As discussed in Chapter 4, while choosing research methods, one has to be aware of the paradigms underlying the methods. Evaluation literature can be divided into publications that follow a more traditional, positivist, structuralist approach, those that are in favour of a more interpretative approach, and more recently, those favouring a mixed approach. All will give arguments as to why the methods they use are more suitable than others. It is up to the researcher to choose the (combination of) methods that are most suitable for achieving the aims of the evaluation and for addressing the particular questions and hypotheses.

The literature on evaluation research also discusses ethical and political issues, as well as sensitivity towards participants, see for example the American Evaluation Association's Guiding Principles for Evaluators in Rossi *et al.* (1999). In design research too, these issues have to be considered, but in general will be less problematic because of the restricted nature of our evaluations – what we evaluate is a preliminary version, the participants are aware of the state of development, the introduction is on a small scale, design research usually does not deal with controversial issues, and in the case of failure the impact on the users and their environment should be negligible if the evaluation is carefully planned. Securing anonymity, we found, is the main issue for most participants, in particular when findings are linked to success or quality. Cost is the main issue for an organisation. This includes the effort required from the participants as well as potential future cost savings when the support is introduced. This can become an issue for the individual participants who may fear loss of status or jobs as a consequence of the introduction of the support.

Methods for an Initial DS-II

A very suitable first indication of the applicability and suitability of newly proposed support can be obtained from running through an example illustrating, as a researcher, what and how the support would work. If this example is based on relevant, real-life situations, this is likely to show how, in theory, the support would deal with the quirks of real life. The real-life example may come from a study, such as data collected during DS-I. For obvious reasons, the example should not have been used to develop the support. An example is given in Blessing (1994) , see where a tape-recorded session between designers was taken as the starting point to describe how a single designer would work with the proposed support. Step by step, it was described how the sequence of events, at the level of granularity provided by the recordings, would have taken place if the support were used. This run-through illustrated and supported the concept, and highlighted several detailed issues that needed addressing before an evaluation with users could take place.

Another method for an Initial DS-II in which the researcher acts as the user, is the use of old case data as a starting point and evaluation benchmark. Stephenson (1995), see Appendix C.8, applied the method he developed to a design problem that had been solved in the past. The evaluation involved a comparison of his results with the results of the designers who were originally given the problem. This test is similar to a comparison between a control group and an experimental group, where there is a delay between the two tests (which in itself may affect the outcome of the comparison).

These initial evaluations of the proposed support only address some aspects, and are based on one case only. The conclusions that can be drawn are therefore strictly limited. In the first case, it is still an assumption that the support can deal with the given situation. In both cases, no conclusions can be drawn as to whether future users can or are willing to apply the support, whether training is effective, *etc*., because no users were involved. Despite the limitations of these initial evaluations, in all cases the researchers gained useful insights into the problems and issues related to the proposed support. Important is that the findings are treated as the results of an initial evaluation, and not as a proof.

We recommend involvement of users and other stakeholders in the initial evaluation to obtain stronger statements about the support. The minimum evaluation would be to demonstrate and illustrate the support and solicit feedback. This feedback can be organised in various ways, for example through questionnaires (anonymous or otherwise), group discussions, or more in-depth interviews.

Reflection

At this point in the research project, it is important to ask oneself the following reflective questions:

- Why do I believe this Evaluation Plan leads to fulfilment of the aims of the evaluation?
- What is my contribution?
- Why do I believe it to be academically worthwhile?
- Why do I believe it to be practically worthwhile, or to contribute to a practical goal?
- Why do I believe that I have the competences or can obtain the competences to execute this evaluation?

6.6.3 Pilot Study

The usefulness of a pilot study has been described in detail in Section 4.6.4. When undertaking an evaluation, the process of introduction of the support should be part of the pilot study.

6.7 Undertaking Evaluation

The various steps for undertaking an empirical study have been described in Chapter 4. This section focuses on the specific aspects related to the use of an empirical study for evaluation.

Sampling

It is important to ensure that the *intended* participants are involved. If this is not possible, the potential impact on the outcome of the evaluation needs to be assessed. It might be necessary to adapt the Research Plan.

Collecting Data

Focusing the data collection on the formulated questions and hypotheses is important in any study, as is keeping an open mind to other unexpected effects that may influence the outcome. This is particularly important in DS-II where the potential of bias in favour of the support is large when the researcher has been involved in the development of the support.

Analysing and Interpreting Data

The interpretation of the collected data may not be straightforward. In an evaluation it is important to determine which of the findings can be assigned to the support (including the introduction), and which of the findings, both positive and negative, might have other causes. It is important to always think of other explanations for the findings. It is easy to be blind to negative findings, to see these as exceptions, or to find causes other than the support to explain these findings. At the same time, one is inclined to pick up any piece of supportive evidence, no matter how small, and to find causes external to the support to explain why this evidence is not so strong.

Rossi *et al.* (1999) divide the effects in three groups as shown in Figure 6.7 (terminology has been adapted).

Gross outcome	=	Effect of support (net effect)	+	Effect of other processes	+	Effect of research

Figure 6.7 Effects on the outcome of an evaluation (terminology adapted from Rossi *et al.* (1999))

Findings can be the result of processes and aspects other than those related to the support. Examples are effects due to the background of the participants. Their needs may be different from the needs of those for whom the support has been developed. They can be very motivated because they are involved in something new and thus be very positive about the support. Participants can also have their own agenda, which may have a positive or negative effect. These attitudes are difficult to prevent, irrespective of whether the participants have been chosen at random or by screening. In studies running over a longer period of time, attitudes may also change.

Endogenous change is another possible source of effects. Design is a dynamic situation and things will change all the time. Take for example the evaluation of a new computer support, developed to be more user friendly than existing support. We know that acceptance of computers in design is increasing and people are becoming more and more accustomed to using computers for a variety of tasks. Hence, positive evaluation results, such as a reduced effort and positive user reactions, could be caused by a combination of the support and an increasingly experienced and accepting environment. The same support, evaluated at an earlier point in time might not have given the same results.

Research itself also has an effect on the outcome. The influence of the research method, the sampling process, *etc*., has been discussed in Chapter 4. The aim of the support will determine whether a *Type 1 error* – incorrectly concluding that the support had an effect – or a *Type 2 error* – incorrectly concluding that the support did not have an effect – is more acceptable. That is, whether it is better to be optimistic or to be on the safe side. As to which approach to take depends on the risk involved in being wrong.

The effect of research may also link back to the theoretical foundations on which the support was based. The need may have been based on limited evidence because of a limited number of empirical studies or the scope of the support may not have been correctly determined, *i.e*., the chosen application may not be suitable.

Verifying Results

The results of an evaluation can be verified in the usual ways, such as collecting data for multiple cases, collecting data in different situations, involving other researchers in collecting and analysing data, *etc*. Verification of every step in the evaluation process is important to reduce the chances of the evaluation itself and other processes affecting the outcome. Verification of the results should focus on the research questions and hypotheses that formed the basis of the evaluation. Verification will allow for stronger conclusions to be drawn, as alternative explanations may be possible to be excluded. Verification may also lead to further evaluations based on a modified Evaluation Plan.

6.8 Drawing Overall Conclusions

The evaluation focuses on the Actual Support and it should be possible to draw firm conclusions regarding this support, including its introduction and use. Drawing conclusions about the Intended Support, however, requires extrapolation. Whether this is at all possible depends very much on the strength of the evaluation statements, the differences between the Actual Support and the Intended Support, and the differences between the actual context (the evaluation setup) and the intended context.

The evaluation aspects listed in Section 6.1.5 influence the evaluation results, and play an important role in finding possible, alternative explanations for the findings. For example, the outcome of using a support can be due to any of the aspects: the process of identifying the need, the assumptions or theory underlying the support, the conceptualisation and scope of the support, the way it was realised, the way it was introduced, the way it was used, the type of users, the time delay between use and effect and the level of resources required. This will affect the conclusions about the Actual Support as well as the Intended Support, and determine the subsequent steps in the research process. For example, it is not useful to improve the realisation of a computer tool that is based on a weak conceptualisation, or to change a method that did not produce the expected results when the introduction was weak.

The search for possible explanations will result in conclusions and recommendations related to modifications in:

- the realisation of the support, its introduction, use or maintenance;
- the conceptualisation of the support, *i.e.*, the translation of the need (the Impact Model and Introduction Plan) into the support;
- the Impact Model and/or the Introduction Plan;
- the Reference Model, derived in the DS-1 stage.

These will determine the subsequent stage of the research process.

Occasionally, it might be necessary to revisit the research goals and the criteria used, if the findings have raised doubts about the assumptions, taken from the literature and reasoning processes, upon which the links between the Success and Measurable Success Criteria were based.

When the results are documented, care should be taken to describe all aspects of the evaluation in such a way that the process can not only be understood, but also be reproduced and the conclusions be used in other research. For the latter, the context in which the evaluation took place is of particular importance, as this needs to be compared with the context in which other research takes place, to determine whether the conclusions can be transferred. The checklist for empiricial studies in Table 4.1 can be useful.

6.9 Main Points

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

- This chapter focuses on how empirical studies can be used to evaluate the application and impact of design support.
- The objectives of DS-II are to identify whether the proposed support has the desired impact and, if not, why; to identify necessary improvements to its concept, elaboration, realisation, introduction and context; and to verify the Reference and Impact Models.
- Both DS-I and DS-II involve empirical studies but have different aims. DS-I is non-interventional; it aims to understand an existing situation to provide suggestions for support to improve this situation. DS-II is interventional; it aims to evaluate whether the introduction of a support improves the existing situation.
- Evaluating design support is essential because its effects can not be known while developing the support: the support is a creation based on assumptions, and the introduction of the support creates a new situation in a dynamic context.
- Evaluation of design support is difficult as: the Actual Support only partially cover the functionalities of the Intended Support; the effects of its

use may be hard to assess; it may take time till the effects occur; the outcome may depend on many aspects other than the support.

- Evaluation should focus on the outcome as well as the process of applying the support so as to be able to interpret the evaluation results.
- DS-II is likely to focus on a proof-of-concept only as it evaluates the Actual Support rather than the Intended Support.
- Support Evaluation takes place in the PS stage, Application Evaluation and Success Evaluation in DS-II.
- Application Evaluation aims at assessing the applicability and usability of the support against the desired values of the Key Factors.
- Success Evaluation aims at assessing the usefulness of the support, as defined by the Measurable Success Criteria in the Actual Impact Model. This involves the evaluation of all factors connecting Key Factors to Measurable Success Criteria.
- Multiple studies may be required to address the factors and links.
- The proposed Systematic DS-II Process contains five iterative steps: reviewing existing documentation, determining evaluation focus, developing Evaluation Plan, undertaking evaluation, and drawing conclusions.
- DS-II can be Initial or Comprehensive. Both cover all steps in the Systematic DS-II process, but an Initial Study does so in less detail and often only focuses on Application Evaluation. A Comprehensive DS-II is required to assess the effects of the support and suggest improvements.
- The Evaluation Plan and the Actual Support influence each other. Therefore, the first three steps of the Systematic DS-II Process start in PS to produce an Outline Evaluation Plan.
- To ensure an efficient, effective and meaningful evaluation, the following are needed: the Actual Support and its Documentation, the Intended Support Documentation, and the Reference Model. When existing support is to be evaluated, these documents have to be collected or generated.
- Evaluation should not only focus on the factors and links in the Actual Impact Model but also on other aspects, related to the specific context of the evaluation, to be able to assess whether the effect can be attributed to the support or not.
- It is important to focus on the *core*, *essential* and *unique* features of the Actual Support and to evaluate these thoroughly.
- The factors and links in the Actual Impact Model and Actual Introduction Plan that should ideally be addressed, are placed against the features of the Actual Support to determine their importance, and against project constraints to determine whether they can be evaluated. The resulting set of factors and links should include at least one complete chain between the Support and the Measurable Success Criteria.
- Research questions and hypotheses are formulated to cover the factors and links in the Actual Impact Model up to the Measurable Success Factor(s) and any other aspects that may influence the evaluation and its results. Each

factor and link is operationalised. This determines what and how to measure.

- The types of user involvement, the setting and the task involved in the evaluation determine the strength of the statements resulting from the use of a particular evaluation method, and the effort required.
- We recommend involvement of users and other stakeholders in the initial evaluation to obtain stronger statements about the support. The minimum evaluation would be to demonstrate and illustrate the support and solicit feedback.
- When undertaking an evaluation, attention should be paid to the introduction of the support.
- Focusing the data collection on the questions and hypotheses is important, as is keeping an open mind to other unexpected effects, including those caused by the research setup.
- It is important to search for alternative explanations for the findings, as the potential of bias in favour of the support is large when the researcher has been involved in the development of this support.
- Verification of every step in the evaluation process is important to reduce the chances of the evaluation itself and other processes affecting the outcome.
- DS-II should allow firm conclusions regarding the Actual Support. Generalising these conclusions to the Intended Support has to be done carefully, taking into account the strength of the evaluation statements, the differences between the Actual Support and the Intended Support, and the differences between the actual context (the evaluation setup) and the intended context.
- When documenting the results, care should be taken to describe all aspects of the evaluation in such a way that the process can be understood, reproduced and the conclusions used in other research.
- The deliverables of DS-II stage are: the Application and Success Evaluation results, as well as the implications and suggestions for improvement of: the Actual Support; the Intended Support, its concept, elaboration and underlying assumptions; the Actual and Intended Introduction Plans; the Actual and Intended Impact Models; the Reference Model; and the criteria used.

7

Writing Up: Publishing Results

Research is “inherently a social enterprise…”, “..a communal achievement, for in learning something new the discoverer both draws on and contributes to the body of knowledge held in common to all scientists.” (Griffiths 1995) Publications are the means through which knowledge is disseminated and evaluated by the larger research community, and through which new directions for research are found. Publications have to have value to others and are essential to obtain feedback. They should not only be considered as a documentation of results. In particular, the reviewing process prior to publication can be extremely worthwhile to improve not only the publication but also the research work itself. Getting colleagues to read the writing can be a very effective means of obtaining first feedback, in particular when they are less familiar with the work; they will ask questions that those that are all too familiar will not ask. As Cuba (1993) suggests,

> *“You must be willing to part with your words [...] Incorporating the constructive criticism of others gives you an additional chance to ‘get it right’ [...] It is our responsibility as writers to allow enough time so that our first is never our final draft [...] Allowing time for revising is no less important than logging hours in the library for preparing to write.”*

Importantly, publishing or presenting research is not only communicating ideas and results to others, but also forces one to clarify and make explicit one’s own ideas and the results found. Cuba (1993) points this out very clearly:

> *“When we write a paper or even a sentence, we objectify our thoughts. [...] Writing provides a constant opportunity to deepen your understanding and sharpen your insights. Taking advantage of this opportunity requires that writing be followed by revision and rewriting to reflect on these insights [...] Think of writing as a process – not an outcome – involving an ongoing dialogue with yourself and with your imagined readers. As you organise observations, fashion arguments, and articulate conclusions, new ideas will emerge.”*

In this chapter we first list the various forms of publications (Section 7.1) as this determines the way in which a publication is written and structured (Section 7.2). The focus of this chapter is on writing a PhD thesis. Some methods to help structure a dissertation can be found in Section 7.3 and tips on writing specific chapters in Section 7.4. Section 7.6 and Appendix C focus on writing papers for refereed journals and conferences. Papers in general follow the same structure and underlie the same quality criteria as PhD theses, but are shorter and as a consequence focus on delivering one or a few messages only. The thesis-related sections are thus equally relevant to writing papers and *vice versa*. The chapter closes with some general guidelines on content and form.

7.1 Various Forms of Publication and Their Intent

There are various forms of publication with varying intent.The primary ones are: research theses, papers, research proposals and popular articles (Ashby 2005).

A *thesis* is a detailed account of a piece of research undertaken for the purpose of obtaining a research degree – a Masters or a Doctoral degree. The readers of a thesis are the examiners. They expect details about all relevant parts of the research process: the goals, the state-of-the-art, the thinking behind the research, the findings and deliverables, the conclusions, and a view on future directions of the research. The intent of a thesis is to enable examiners to judge the work for its contribution to the body of knowledge in the area, so as to evaluate the suitability of the researcher to be awarded the degree.

A *paper* is an account of a piece of research, typically published in a journal, conference proceeding or a book. The intent is twofold. First, the intent is to enable reviewers, who are experts in the areas covered in the paper, to judge the originality and quality of the work and of the paper itself, as well as its suitability for the journal, conference or book concerned. The second intention is to allow the reader to extract information for the purposes of carrying out further research, teaching, or practice.

A *research proposal* also addresses two types of readers. The first are the reviewers that the funding agency will use; they are charged with judging background, quality, promise, relevance and cost/benefit ratio. The other is the funding agency, who will look for a match between their priorities and the focus of the proposal.

Not to be underestimated is the writing of a *popular article*, addressing an audience who is intelligent but who may know nothing about the subject of research. Here, style, always important, must be fine tuned to meet the needs of the audience, which are to be introduced to a new field and to be entertained.

Note that even though a diary or lab book is also meant to document research, and are indeed immensely helpful, a diary is not a publication. A diary is a chronological, *i.e.*, historical account of how research took place. A publication should be a logical account of the research: a text with a clear thread that at the same time does not do injustice to the way the research took place. If necessary, the chronological account can be suitably placed in a figure in the introduction of the publication.

7.2 Overall Structure of a Thesis

Various books have been written on how to write a thesis. In addition, the charters of a university usually specify the broad goal of a thesis and the constraints on its form (*e.g.*, structure, format and size). In general, the specifics vary widely reflecting the typical research approaches and traditions of the disciplines and institutions involved. It is therefore important to consult not only books but also existing theses in one's own discipline and institution to see what the norm is.

The content of a thesis contains the following elements in one or more chapters. It may be necessary to repeat the elements for each of the stages of DRM carried out in the research.

Introduction

This first chapter introduces the area of focus, the motivation for the research, the main problems, research questions and hypotheses that the researcher has explored, and why these are important, *i.e.*, the main results of the RC Stage will be presented here. Furthermore, this chapter provides an overview of the content of the thesis.

Literature Review or State-of-the-Art

The literature review provides a review of the relevant contributions from the existing body of the literature. The literature review should identify the theoretical foundation for the research, identify the level of novelty and relevance of the research described in the thesis, and help to clarify and refine the focus, research questions and hypotheses to be addressed. The literature review should also provide the justification for the research focus. The literature reviews of various stages of DRM are documented in this chapter. The ARC diagram and the Reference Model can provide a structure for the literature review and its results, respectively. Note that depending on the stages covered by the thesis, a second literature review chapter or section may be required in another part of the thesis. Usually, however, the literature review is the second chapter of the thesis.

Research Approach

This part of the thesis provides detail about the research methodology and the method(s) used to address each question and hypothesis, as well as the argumentation behind the selected methodology and methods. The detail should be sufficient for another researcher in the area to understand the findings, to judge the interpretation of the author, and to repeat the research in the given context. Each DRM stage that involved a Comprehensive Study is likely to require a separate chapter describing the research approach for that particular stage. The Research Plan is the basis for this. For example, the thesis can have one chapter for the research approach of DS-I and, after the presentation of its results, a chapter on the research approach taken in PS. The (Initial) Reference and Impact Models can be used to illustrate the motivation behind the approach and the factors and links on which the approach focuses.

Results

This chapter presents the results of applying the research approach by structuring the findings according to the research questions and hypotheses that the research addressed. Again this chapter can appear for each stage of DRM. Depending on the stage, the results chapter covers new insights, theoretical frameworks, models, support (concept or realisation) or support evaluations. The Reference and Impact Models as well as the Support Documentation can provide additional structure to the chapters.

Discussion

This includes the interpretation and generalisation of the findings as well as a reflection on the applied research methods. If several research methods are used within one stage, this will include a comparison between and the confluence of results obtained by the different methods. The specific contribution of the research is highlighted through a comparison with relevant literature. The structure can again be based on the Reference and Impact Models or the Support Description, depending on the DRM stage concerned. The results can be presented and discussed in the same chapter as the results, but only if particular care is taken that the two can be clearly distinguished.

Conclusions

The conclusions are a summing up of the advances in knowledge that have emerged from the research work reported and its limitations, putting the results in the wider scientific and or practical context, and envisaging the future directions that emerge from this research. Note that each stage needs some conclusions. The final conclusions chapter, however, will combine all the results and link them back to the goals, research questions and hypotheses set out in the first chapter, literally rounding of the thesis.

References

References list the sources of all ideas, prior results and data used in the thesis. It is a conventional courtesy to also reference the originators of key ideas, theories or models that have been an inspiration, rather than explicitly used. Depending on the conventions used, these references may be listed in a separate bibliography. The format for references used in a thesis differs per country, discipline, institute and publisher. No guideline can be given other than that it is important that all such sources are provided with all detail required for the reader to find these.

Appendices

Appendices contain material that provides the background and details that contribute to the understanding of the research reported in the main text, but that would interrupt its flow. Examples are details of the background of participants, of the developed support, of the material used in a research method (such as a questionnaire), of the individual findings, of calculation methods, *etc*. The main text usually contains the most relevant information, often in a summarised form. Appendices can also contain information that might be unfamiliar to some of the readers, such as details of particular statistical methods.

7.3 Approaches to Help Structure a Thesis

Even though it may be possible to write a thesis without much planning, it is far more effective and efficient to write a thesis in a structured way. We have identified four approaches that were found helpful in clarifying the structure of the thesis and supporting the writing. The approaches can be used together.

7.3.1 Table of Content Approach

This approach is commonly recommended. The researcher should start writing a thesis from very early on in the research process, by creating an outline thesis document, *i.e.*, the table of content, covering the intended research work. Analysing existing, successful theses within the same area can help setting up a first outline. In the beginning, the exact content of the chapters might still be unclear. As research progresses, the content will be progressively added to and modified. The outline may also change, as the topic becomes clearer and the first results emerge. This approach not only encourages writing up throughout the process, but also helps the researcher plan his or her work by showing the elements that need further attention.

7.3.2 Presentation Approach

An alternative approach is the use of presentation software, such as Powerpoint rather than word processing software, to remain focused on the key points of the content and the logic behind the structure, rather than on the details of the content, *e.g.*, well-formulated sentences. Slides are produced as and when particular work has been done, starting from and structured around the outline of the thesis. The necessary shortness of the messages on each slide will force effort to be spent on thinking about the main points of what has been achieved. The overview and the possibility to move individual slides around, supports the development and maintenance of the logical structure: text blocks are far more difficult to move around because of a lack of overview and because of the fact that text blocks tend to contain words that link them to other text blocks. As supervisors, we experienced that such a presentation mode allows an excellent basis for discussion by providing a quick overview of the main findings and contributions, even though the final set may contain up to a few hundred slides. The presentation mode helps identify any inconsistencies in the logic and argumentation and allows structural changes to be made and evaluated directly. Text documents, as they grow, lack these possibilities and involve a far larger effort from author and supervisor in writing, reading, discussing, identifying problems and correcting these. When developed consistently, the final set of slides is a summary version of the thesis emphasising the main messages, which – as we observed – greatly facilitates the writing of the thesis: one 'only' needs to add the connecting text.

7.3.3 Methodical Design Approach

Ashby (2005) recommends a methodical design process for writing papers, which can be adapted for the purpose of writing a thesis. Along the lines of the systematic design methodology of Pahl and Beitz (2007), he recommends four steps in writing up:

- Establishing market need: This involves the clarification of the intent of the publication: who the readers are and what they are looking for in the publication.
- Conceptual design: This involves developing a plan for the whole publication. The essential elements of the content are specified and related using an A3 sheet of paper – the concept sheet – to structure the thinking by means of drawing boxes, circles, text, arrows, *etc*.
- Embodiment Design: The outcome of this stage is the first draft of the publication, containing all major facts, calculations and conclusions in appropriate chapters without worrying about style.
- Detail Design: In this stage, the first draft is refined and the content completed. The focus is on clarity, balance and readability. The end product is the publication.

We would like to support the recommendation of Ashby to use graphics, graphs, large sheets of papers, *etc*., to generate, review and maintain the overview of the thesis. Generally such overviews are easier to create by hand than using graphical software. When fewer graphics are used, Mindmapping software can be useful. In general, however, screen or printing size can quickly become a limitation when detailing the structure.

7.3.4 Question and Answer Approach

To support the general line of argumentation in the thesis or more detailed lines of argumentation in individual chapters or sections, a two-column sheet can be useful in which the left column is used to record questions, and the right column to record their answers in the form of keywords. The answers often give rise to new questions, which can be added to the left column at the appropriate place. These questions can be divided into potential chapters, sections and paragraphs, where each paragraph addresses a set of questions that logically connect together. The idea behind this approach is that each chapter, each section and each paragraph should have a purpose, *i.e.*, should answer a question. If it does not do so, the respective text can usually be left out.

Once the structure has been established, the questions and answers can be moved to a thesis document, by writing these at the position of the future text blocks. In this way, the focus of each text block and their combination is determined. As writing progresses the paragraphs, sections and chapters 'only' have to be filled with those details that answer the respective questions. The

questions can also be used in the introduction of each chapter or section to introduce its structure.
The approach of formulating questions can also be applied *after* a text has been written to analyse the purpose of and logical links between text blocks (paragraphs, sections and chapters). Once a text has been written, each text block is labelled with the question this text blocks answers. Looking at the questions and their sequence, it often becomes clear that text blocks related to the same question are not currently together, that the sequence is not logical, that aspects are missing, that the text block does not answer a (relevant) question, that questions are not answered, *etc*.

A slightly simpler but still effective technique at the level of individual paragraphs is the labelling of each paragraph with a keyword. The analysis of these keywords and their sequence provides an overview of the structure and any inconsistencies.

7.4 Tips on Writing Specific Sections

This section provides a number of tips on writing specific sections of a thesis, but apply equally well to other scientific publications. These tips are in addition to those provided in Appendices C and should be used in conjunction with each other.

7.4.1 Writing the Table of Content and Other Lists

The table of content should provide an overview of the content and structure of the thesis. Care should therefore be taken to choose chapter and section headings that are informative, without being too descriptive. For example, a heading ‘Methods’ is not as informative as ‘Decision making methods’, whereas a heading such as ‘Methods that help designers take decisions’ would contain unnecessary detail.

In general, the table of content should not contain more than 3 heading levels. An analysis of the table of content will reveal any imbalance in the size of the chapters and the number of sections in each chapter. Preferably the main chapters are of similar size. Never should a chapter or section have one (sub)section only.

If many symbols are used in the text, a list of symbols (nomenclature) is obligatory. It can be useful to present lists of abbreviations, figures and tables after the table of content. In such cases, clear and concise labelling of the figures and tables is essential: the purpose of the lists is to help the reader find a particular figure or table easily.

7.4.2 Writing the Introduction

The introduction starts with the background and motivation of the thesis. Arguments can come from own prior experience, prior work of the research group or practice, and are backed up by the literature (or the fact that the literature is lacking). Next, the overall problem, the goal(s) of the thesis, the main research questions and/or hypotheses, as well as the underlying assumptions are derived or generated. The formulation should be clear and the link with the background justified.

It is important to clarify what the readers can expect in the thesis. The focus and the scope of the thesis, describing what will be covered and what not, and for which applications, helps in doing so. The references to the literature used in the introduction should be limited to those that are essential to back up the choice of focus. This is not the place to discuss the literature in all detail.

The chapter should conclude with an overview of the content of the different chapters and the structure of the thesis. The formulation of the main question(s) addressed in each chapter can help to show the logic behind the structure. If the relation between the chapters is not straightforward, a graphical overview of the structure of the chapters can be very useful (for an example, see Appendix C.2, Figure C.2.1)

7.4.3 Writing the Literature Review

A literature review is used to present the current *relevant* state-of-the-art to inform the reader about the context in which the research has taken place and the reasons for the focus of the research. A literature review should thus be written such that the reader can understand why particular literature is mentioned, can follow the argumentation, and would draw the same conclusions as the author about the areas that need addressing and the problems that need solving.

This requires that the literature is summarised and findings classified and compared (not just listed) such that it becomes clear where the gaps are, what the problems are (identified by the author or in the literature) and what conclusions can be drawn that are relevant for the research described in the thesis. It is important not to leave the drawing of conclusions to the readers: their conclusions may differ from yours. The summaries of a particular publication should contain enough detail to understand the work to the extent that its relevance or contribution to the thesis can be judged. Often, several publications can be summarised together in one or a few sentences. One reference can be taken as an example to illustrate the relevance of its content for the thesis work. The other references can be simply cited within one additional sentence referring to the fact that they to provide the same evidence, use the same approach, *etc*. For example, ‘Author [reference] showed that ……. Similar results were found by [reference, reference, ….]’. The latter way of referring to a group of sources can also be used to indicate that the author is familiar with a particular area, which the reader may expect on the basis of the research topic, but which the author does not consider relevant enough to discuss in detail. Obviously an argument is required when doing so. For example: ‘A different approach is the process-based approach described by [reference, reference, …]. This approach is not considered applicable to the work at hand, because ….’

The literature review is thus used to provide the justification for the research. A simple collection or listing of descriptions and critiques of existing work is insufficient and this is not the place to show what one has read. A well-argued line of thought established through the description and critique of current literature is required. It is a mark of good writing when, through this chapter, the author is able to not only inform the readers but also raise their interest in the topic and the rest of the publication, and convince with the chosen line of thought and argumentation.

A proper literature review will involve an interpretation of and reflection on what is found in the literature. Care should be taken that, at all times, readers can distinguish the interpretations from the original statements. The reflection on what is found is likely to involve criticism. In fact, quality research is about critically reviewing existing results and the processes used to arrive at these results. When writing down critical words about a reference, this should not be formulated in an offending manner and must be supported by arguments that are relevant for one's own work. A critique should also state, if applicable, how it positively influenced one's own research.

All the work that the researcher has used in the research should be cited. The instructions for citing references from the Open University Library Service (U.K.) mentions some of the reasons why it is important to give complete, accurate references (Taylor 2008):

- Your references show you have read around the subject.
- Your argument will be stronger if supported by evidence from other's research.
- You enable others to find and use the sources that informed your work.
- If you don't include references, you will be guilty of plagiarism, *i.e.*, passing off someone else's work as your own.

One can quote directly, retain some of the words but embed these in one's own sentences, or paraphrase the work in one's own words. As Cuba (1993) points out "Whether you quote directly or summarise in your own words the ideas of someone else, you must acknowledge your debts" otherwise you are plagiarising.

Information about plagiarism, quoting text and citing references can be found on the websites of many scientific institutions and organisations and funding agencies. For example, Berkeley Campus Code of Student Conduct (N.N. 2007) defines plagiarism as "the use of intellectual material produced by another person without acknowledging its source. This includes, but is not limited to:

- Copying from the writings or works of others into one's academic assignment without attribution, or submitting such work as if it were one's own;
- Using the views, opinions, or insights of another without acknowledgment;
- Paraphrasing the characteristic or original phraseology, metaphor, or other literary device of another without proper attribution."

Different types of plagiarism, are discussed in (N.N. 2008), indicating that even when sources are cited, plagiarism can occur knowingly or unknowingly. An example is "The Too-Perfect Paraphrase" type of plagiarism. "The writer properly cites a source, but neglects to put in quotation marks text that has been copied word-for-word, or close to it. Although attributing the basic ideas to the source, the writer is falsely claiming original presentation and interpretation of the information."

It is important to consult the relevant websites and other sources to get familiar with the standards used in a specific discipline.

Common knowledge does not require a citation to the original source. Common knowledge is "anything that is repeatedly mentioned in published materials but never cited. When in doubt, cite the reference." (Cuba 1993)

Extensive use of quotations can render a text unreadable because of the large number of quotation marks. The original should be quoted only "if one really likes the ideas and the wording of the original this much, if it is important to your paper, and if it is stated more concisely in the original than it would be in your paraphrase or summary" (N.N. 2006). When quoting a text, no changes should be made unless needed for understanding, which should then be explicitly indicated.

When writing the thesis, it is often hard to trace back the relevant segments from particular publications about which one has made notes, if at all. In order to prevent this, it is important to take notes at the time of reading that *include* page numbers and to clearly mark in the notes whether the text is a quotation or one's own summary, paraphrase or interpretation of the original.

7.4.4 Writing the Research Approach

The description of the research approach(es) should be accompanied by an argument as to why this particular approach, these methods and this context have been chosen to address the research problems, questions and hypotheses. The reader should be convinced that this choice is likely to solve the research problem, answer the questions and verify the hypotheses, under the given constraints. The different types of research listed in Figure 2.2 can be used as a starting point.

The description of each of the methods should be detailed, whether it involves an empirical study, support development or evaluation. The list of characteristics of descriptive studies presented in Table 4.1 can be used to specify each of the empirical studies and evaluations. As discussed earlier, the description should have sufficient detail for the reader to be able to determine whether the findings are relevant for his/her own work. If, *e.g.*, the context of the empirical study described in the thesis differs considerably from the context in which the reader is involved, he or she may not be able to transfer the results of the study to his own work. Details, such as background information about the participants, companies, projects or products involved, the details of the operationalisation of the concepts, the material used or the algorithms applied, are important but would interrupt the main text flow and should therefore be placed in appendices.

Note that a detailed description is not meant to include a historical account of the development of the approach. Rather it should give the approach taken and its justification. A short discussion of some of the alternatives that were considered can be useful.

7.4.5 Writing the Outcomes

The outcomes for each stage are best presented in a way that is structured according the goals, research questions and hypotheses related to this stage, as the outcomes are intended to address these. This focus will avoid a mere listing of all results that would leave it up to the reader to determine the relevance of the results in the context of the work.

In writing down the results of empirical studies, it is important to make a clear distinction between findings, interpretations, generalisations and conclusions. Interpretations and generalisations should be clearly argued to convince the reader and alternative explanations should always be presented. It is important that the results are discussed in relation to the goals and focus of the research and the identified problems, and that they are compared with the state-of-the-art to find supporting evidence and to establish what is novel.

Results can be easily formulated in a way that suggests a trend or conclusion favoured by the researcher, even though the data does not provide sufficient evidence (see also the discussion about bias in Chapter 4). For example, if in 50% of the cases the support worked and in 50% it did not work, a statement that '50% worked' has a positive touch as compared to '50% did not work'. Also, a statement that 'two out of three participants found the method useful' suggests strong support. Knowing that only three participants were involved renders the statement very weak. In this case, the statement should have been 'two of *the* three participants found the method useful'. The news is full of such examples. Care should be taken that the presentation of the data is not biased and that statistics are used only where this makes sense.

When describing support developed as part of the thesis, a clear distinction should be made between what the support actually does and the expected consequences. In particular, it is important to distinguish between the description of the structure and flow in the method or tool, and the description of its use. Furthermore, a distinction should be made between what the Intended Support is expected to achieve and what the Actual Support is expected to achieve.

The Support Documentation should help the reader understand the assumptions, scope and limitations of the system. Using the checklist of scope and assumptions presented in Appendix B.4 could be helpful to structure the description. To describe the process used, one or more scenarios, preferably based on an existing situation, have a strong illustrative effect, in particular because the situation, the actions and the actors are linked. Particular attention should be paid to the description of the input and output data, as this is a strong indicator for the situation in which the support can be used. A user manual should not be part of the main text, but can be a useful appendix or additional publication. Explanations of the support can be given not only in text but also using models and examples. Any evaluation of the support should be described as discussed above for empirical studies.

In general, a description using graphs or tables is better than using flowing text. Being creative in finding ways to present outcomes (obviously without biasing these) can underline the essential contributions and keep the reader interested. Providing information and explanation should always be the main goal – graphics that are nice but do not add information should be avoided, and any graphics should be fully explained, rather than left to the interpretation of the reader.

7.4.6 Writing the Conclusions and Acknowledgements

It is important not to duplicate the 'Abstract' as 'Conclusions', and *vice versa*, as Ashby points out (Ashby 2005). A summary provides an overview of the entire thesis, reminding the reader of what has been presented in the thesis, while the conclusions are a summing up of the advances in knowledge that have emerged from the work. The conclusion chapter can start with a summary, but should not be limited to this. The main elements are the main contributions, reflections on the research undertaken and its limitations, as well as an outlook.

Typically, the conclusions should include the main points about both the problems addressed and answers found or solutions proposed, while highlighting the contribution of the researcher in comparison to that of others. Particular emphasis should be on the scientific and – if applicable – practical contribution of the research: a mere reference to the developed support as a contribution is insufficient. The conclusions should include a comparison of the outcomes with the goals, problems and objectives that the researcher set out to fulfil. The easiest way is to go back to the first chapter and discuss what was promised and said in this chapter, so that the 'circle' is closed.

The reflection on the limitations of the approach should include their possible effects on the results, and the potential implications on the interpretations and generalisations that were made. The limitations should cover the specific methods that were applied, as well as the overall methodology. The chapter will close with suggestions for further research, for improvements to the research process and its outcomes, and for the application of the results. The importance of these sections should not be underestimated as any individual research is a small but potentially significant step in an ongoing process of knowledge creation.

In a paper, the main text is followed by a section 'Acknowledgements' – thanking all who supported the work, scientifically, technically, financially and editorially. In a PhD thesis or book, this is usually placed in a preface.

7.4.7 Writing the Reference List and Bibliography

The reference list contains detailed information about all the references used in the other chapters. References are listed according to either the numerical order in which they are used in the main text (if a number system is followed) or in alphabetical order of the first authors' surnames (if other systems are followed). The information provided should be adequate for another researcher to identify the reference. Various standards exist. These can be country and discipline specific. It is important to know the relevant standard early on, in order to avoid having to retrace the original sources to find the required information. Software such as EndNote can help maintain reference lists and produce citations in the text and reference lists at the end of the main text in various formats. The reference list is usually placed before the appendices.

Apart from a reference list, a bibliography can be included to list all the references that are not quoted in the thesis, but have been influential in the development of the ideas described in the thesis.

A Digital Object Identifier (DOI) is a persistent identifier (such as the ISBN number used for books) which may be used to cite and link to electronic documents. The DOI consists of a unique alpha-numeric character string that is assigned to a document by the publisher upon the initial electronic publication. The DOI will never change and is therefore the correct identifier for referencing to online publications and articles in press, which have not yet received their full bibliographic information. Such articles are referenced using the author, title, DOI and DOI number. DOIs may also be used to create persistent URL hyperlinks to documents on the web, which are guaranteed never to change. An example of a hyperlink URL for a DOI is constructed as follows: http://website/DOI number (*e.g.*, http://dx.doi.org/10.1016/S0006-8993(00)02382-9). The DOI scheme is administered by the International DOI Foundation. Many of the world's learned publishers have come together to build an article linking scheme based on DOIs known as CrossRef. Details can be found on http://www.doi.org.

7.5 Writing Papers

We encourage the readers to use the booklet of Ashby on how to write a paper, which is available online (Ashby 2005). In this section we limit ourselves to highlighting an essential aspect of writing papers, namely to *focus* on one or a few messages that are of interest to the research community. The work described in a thesis covers or may cover many such messages and can thus be the basis for several papers. In most cases, summarising the whole thesis in one paper results in a lack of detail and strength of the statements made. For a paper, and in particular a survey paper, it is hence important to answer the following questions (as suggested by a reviewer of a survey paper and with kind permission for reproduction):

- What are the three or four insights/nuggets that you would like to share with the community?
- How will you tell your story so that these nuggets are clearly recognised for value by the readers?
- Is it worth our effort (in writing and reading)?

If papers are reviewed, common criteria are: novelty, significance, correctness and readability. A further criterion is the relevance of the content to the scope of the journal or conference to which it is submitted. Reviewer instructions can therefore be valuable sources of information on how to write a paper. Smith (1990) wrote an excellent article, which is also available online, about the task of a referee for different kinds of publications, providing the criteria by which a referee should evaluate such publications and hence providing the author with requirements that can be used to check one's publication before submission.

7.6 General Guidelines

This section contains some general tips for writing various kinds of publications.

7.6.1 About Content

It is essential that all key terms are defined and, if necessary, explained with an example. If there are many terms, a separate appendix (a glossary) could be used.

All tables, figures and equations should be numbered and referred to in the text. The captions of tables and figures should be brief and preferably self-explanatory. Everything included in a table or figure should be explained in the text.

The whole body of text (across the chapters and sections) should have a clear, logical thread. Anything that breaks the flow should be kept out of the main text and put in an appendix, a footnote – for small remarks – or graphically marked, such as examples as boxed text. Each chapter and section should have a purpose within the structure. Any chapter or section that does not do so, is not part of the thread and should be left out of the main text. If in doubt, leave it out.

All chapters should open with an introduction to the chapter and its link to the previous chapter(s), *e.g.*, 'In the previous chapter we have seen …. In this chapter we will further elaborate on….'. Similarly, all chapters should be closed, *e.g.*, by highlighting the main points or providing a short summary and linking it to the following chapter(s).

Long paragraphs should be avoided. Instead, subsections with unnumbered headings can be used, as well as strong transitional sentences, or a sentence outlining the structure of the paragraph, such as 'Three different approaches can be identified: approach A, B and C' followed by a description of each approach. Such descriptions should follow the same order as used in the introductory sentence (*i.e.*, A, B and then C), and should preferably follow the same pattern and style.

In general it is very helpful to get others to read a draft publication, even though this might provoke criticism. This should not embarrass the researcher but should be taken as positive clues as to how to improve the content and form of the work.

Further inspirations can be derived from analysing other, similar kinds of publication, identifying what one likes and dislikes about the style and structure of these publications. Further sources are the instructions for authors in various journals, and instructions for reviewers – as they will eventually judge the publication.

7.6.2 About Form

For non-native speakers a thesaurus, in particular online, is often more useful than a dictionary as it provides several alternative meanings of a word, each of which is again listed with its alternative meanings. This makes it easier to identify the appropriate word. When looking for translations of specific terms, normal dictionaries may not provide the correct terms. The literature written in the intended language by native speakers is a far more reliable source.

In general, short sentences are better than long ones. Lists with bullet points generally are clearer than flowing text, if the constituents of a theme or concept (like the steps in a method) or possible alternatives are described. The order in which the items appear in the list has to have a structure: *e.g.*, ranked according to importance or size, following the process of execution, grouped according to type, *etc*. If the list is followed by a description of the items, all listed items should be described and in the same order in which they appear in the list. When lists are repeated, the items should always appear in the same order. We have seen many lists that did not follow these basic rules and as a consequence considerably reduced the clarity of the text.

It is essential that the same terminology is used throughout the text. If more terms with the same meaning are inevitable, this must be clarified upfront.

If 'first' is used, a 'second' should follow. The 'one hand' also has an 'other hand'.

It is worth trying to avoid 'I' and 'We', as long as the source of the statement remains clear.

Male and female terms should be avoided if possible; one simple way in the English language is to use plurals. For example, 'When writing down *their* results, researchers should….' rather than 'When writing down *his or her* work, a researcher should….'. In reporting findings of an empirical study, the actual gender of the participants can be used. In generalisations and conclusions personal pronouns should not be included unless the gender distinction is relevant.

Ashby (2005) provides a summary of the main style rules and English grammar, spelling and punctuations rules.

Writing has to be checked for clarity, conciseness and grammar, as well as organisation, balance and style. Despite the effort involved, reading the final text out *aloud* is very worthwhile. This will immediately reveal inconsistencies in the text, sentences that are too complex, punctuations that are missing, and disruptions in the flow. All of these are very difficult to identify by just reading through the text.

A file 'rest.doc' can be very useful as a deposit for all those beautifully crafted sentences and sections one is reluctant to throw away because of the effort that was put in writing them, although they are no longer relevant for the publication. The file ensures that the text is not lost for future use (although we found that reuse hardly happens).

7.7 Main Points

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

- Research is "inherently a social enterprise". Publications are the means through which knowledge is proposed, evaluated and disseminated into the larger research community.
- There are various forms of publication with varying intent: research theses, papers, research proposals and popular articles. This chapter focus on writing theses and papers.
- A publication should not be a historical account of the research process, nor a dumping place for what the author has read, the knowledge he or she gained, the tasks done, and the results found. A publication is written for an audience, should have a purposeful, logical structure, convey clear messages, and highlight its contribution to the wider community.
- A thesis typically contains an introduction, literature survey, research approach, results, discussion, conclusions, references and appendices. Some of these may appear several times depending on the research process.
- A thesis may further have a preface/acknowledgements, an abstract, a bibliography and lists of figures, tables and nomenclature.

- The structure of a paper is similar to that of a thesis but focuses on one or a few messages only.
- The introduction introduces the area of focus, and the main problems, research questions and hypotheses that the researcher explores in the thesis, as well as the reasons as to why these are important to be addressed.
- In the literature survey, the contributions of the related, existing body of the literature is reviewed to clarify the questions remaining to be answered, and their novelty and importance.
- The research approach details the chosen research methodology and the method(s), as well as the argumentation for this choice.
- The results and discussion chapters present the findings of the research, their interpretation and generalisation, and compare these with the existing literature.
- The conclusions sum up the advances in knowledge that emerged from the research, and emerging future directions.
- The references lists all sources of ideas, prior results and data used in the research and referred to in the thesis.
- Appendices contain background material and details that contribute to the understanding of the research, but would interrupt the flow of the main text.
- Four approaches are recommended for clarifying and detailing the structure and content of a thesis: Table of Content Approach, Presentation Approach, Methodical Design Approach, and Question and Answer Approach.
- In the Table of Content Approach, the researcher creates an outline document of the thesis early on, which is gradually developed as research progresses.
- The Presentation Approach uses a continuously extending slide presentation describing the key messages for each section of the thesis.
- The Methodical Design Approach has four steps: establishing the market need for the thesis, its conceptual design, embodiment design and detailed design.
- In the Question and Answer Approach, a two-column sheet is used as an overview to record the research questions and their answers that form the basis for each chapter, section and paragraph in the thesis.
- Writing is a process, involving an ongoing dialogue with oneself and with imagined readers, allowing ample time for revision. Getting others to read the writing can provide useful feedback for improvement.
- Similar kinds of publications, the instructions for authors in relevant journals, and instructions for reviewers can provide useful guidance.
- The checklists, the models and deliverables proposed in this book can help structure and illustrate the chapters and sections.
- Writing should be checked for clarity, conciseness, grammar, organisation, balance and style, for example by reading aloud.
- A file 'rest.doc' is useful for sentences and sections that cannot be used in a particular publication, but may have 'reuse' value.
- The booklet by Ashby provides an excellent summary about style, grammar, spelling and punctuations in English (Ashby 2005).

8

Summary and Conclusions

As the final chapter in this book, we would like to share our experiences and that of our students and colleagues regarding the impact of DRM.

8.1 Experience of Using DRM

Often, young researchers are worried about the contribution of their work, have a tendency to work alone to protect their findings, and to hide their worries. As mentioned earlier, science, however, is best seen as a social enterprise. As King *et al.* (1994.) say:

> *"Every researcher or team of researchers labours under limitations of knowledge and insight, and mistakes are unavoidable, yet such errors will likely be pointed out by others. Understanding the social character of science can be liberating since it means that our work need not be beyond criticism to make an important contribution [...]. As long as our work explicitly addresses (or attempt to redirect) the concerns of the community of scholars and uses public methods to arrive at inferences that are consistent with rules of science and the information at our disposal, it is likely to make a contribution."*

With this as the broad view, we wish to reflect on the impacts of DRM as we experienced through our teaching and research projects. To what extent does DRM do its job and fulfil its promises?

For this, let us refer back to the specific objectives of DRM (see Section 2.1). The objectives are to:

- provide a framework for design research – Chapter 2 describes the DRM framework and seven different types of research in design; Appendix C provides examples of research projects that did not use DRM and yet can be characterised using the identified types of research and placed within the framework;

- help identify research areas – Chapter 3 is intended to support this;
- allow a variety of research approaches and methods – many approaches and methods are linked to DRM in Chapters 3–6 and Appendices A and B, many others can be linked;
- provide guidelines for research planning – Chapter 3 provides guidelines for overall planning, Chapters 4–6 provide guidelines specific to each DRM stage;
- provide guidelines for rigorous research – the overall DRM framework described in Chapter 2 and the stage-specific guidelines in Chapters 3–6 support this;
- help develop a solid line of argumentation – the DRM framework and the Reference and Impact Models are intended to support this;
- provide new methods and pointers to existing methods – research processes specific to each DRM stage and new methods are introduced and pointers to existing methods provided in Chapters 2–6 and Appendices A and B;
- help select research methods – Chapters 4–6 provide specific guidelines, while general guidelines are given in Appendices A and B;
- provide a context for positioning research projects and programmes – the inter-linked stages of DRM and the seven research types (Chapter 2) are intended to provide this; the examples in Appendix C show that a variety of existing research projects can be positioned in the framework and, in doing so, can be related to each other and the wider context of design research;
- encourage reflection on the approach – with the logical links between the DRM stages, the use of Reference and Impact Models to distinguish the current and desired situations, the distinction between Intended and Actual Support and the different types of evaluation, the emphasis on linking the research to previous and subsequent stages and the specific guidelines on selection of research methods, we hope to support this.

We hope that fulfilment of these objectives will bring more rigour to design research, which in turn would make design research more relevant, effective and efficient. We further hope that the use of DRM is able to reduce the gap between what design researchers develop and what people use in practice, making design research more utilisable.

All these objectives can, in some sense, be considered the elements of the 'Impact Model of DRM', showing the impact of our envisaged support (DRM) to doing design research. This raises the question: what are the DS-II results? In other words, has DRM been evaluated and what was the outcome?

Formal evaluation of DRM – in the sense of DS-II – is about to begin with your help. The book, we hope, will be a means of widespread introduction of this support, and feedback received from the users will provide us with the data for its evaluation. Right now, however, we can only speak from initial evaluation based on the introduction of DRM through our research papers, courses and projects, and from subsequent feedback received from students and colleagues who used it in their work. We base our evaluation on the following three sources, each of which is discussed in the subsequent sections:

- feedback and reflection reports from over 200 PhD students from various European countries attending the ten Summer Schools on Engineering Design Research, held annually for two weeks, where the first author is one of the instructors, and DRM forms the basis for the course (Blessing and Andreasen 2005);
- feedback and performance reports of a total of about 30 Masters and PhD students from the five batches of a graduate level course on Methodology for Design Research, offered annually during one semester (5 hours per week, 18 weeks) in India where the second author is the instructor, and DRM forms the basis for the course (Chakrabarti 2008);
- the various research theses and research papers that used, or discussed the use of DRM in their work, including those of our own Masters and PhD students.

8.1.1 Feedback from the Summer School on Engineering Design Research

The following questions asked by PhD students triggered the idea of the Annual Summer School on Engineering Design Research for PhD students (Blessing and Andreasen 2005):

- How do I make my work scientific and not just a consultancy report or a report of my own learning?
- What are the methods and approaches to be used? How to do empirical research? Is it sufficient to do only theoretical research? How to provide or justify my results?
- How to obtain insight into the state-of-the-art, to find a theoretical foundation and to judge competing contributions?
- How to actually plan my research?

The experience of running the Summer Schools and interacting with the PhD students have led its teachers to learn the following (Blessing and Andreasen 2005):

- The view of the current situation in design research as detailed in this book – that the structure, contribution, coherence and research methodology in design research are rather poor, has been corroborated.
- Many research projects are impossibly broad, and students are often unable to articulate arguments for the research questions and hypotheses.
- Many projects have no proper formulation of metrics, which makes verifications or conclusions weak.

Flanagan and Jänsch, who were Summer School participants, identified the common challenges of PhD research in design (Flanagan and Jänsch 2004) as:

- Focusing research on the topic relevant to industry and realistic within the constraints imposed by a PhD project, and associated challenges of structuring the project with appropriate research methodology;

- The sheer range of factors that influence design. Research challenges are often closely related to the Social Sciences; poorly defined terminology causes additional problems.

The above is a reflection of the problems faced by at least the young design researchers, and can be considered elements of the Reference Model – the current situation – which the support (DRM) is intended to respond to and change. For more details on these issues, see Chapters 1 and 2.

Flanagan and Jänsch (2004) provide a number of observations about how DRM and associated methods help students in the Summer School. In their experience, DRM provides a framework for a wide variety of research "by allowing the researcher to determine which stages are given the most attention and indeed how many of the research phases are covered throughout the research project." They also found that the emphasis on establishing the Success and Measurable Success Criteria focuses attention on the problem of assessment, and "several students proposed changes to planned investigations in order to deal specifically with this issue." The use of the ARC diagram, they observed, "clearly illustrated the intended scope of the PhD project"; in many cases, this revealed that the project scope was "extremely broad and ambitious for the approved timeframe", which helped to "increase focus and illuminate problems." Use of Reference and Impact Models, in their opinion, helped the students "to identify the industrial value of a given PhD project, by showing the connection between successful project outcome and enhanced industrial performance." The core value of these models, they feel "is that they cause students to (re-)consider the questions *is my work useful?* and, if so, *how is it useful?* This in turn helps them re-evaluate their work and identify simple changes which can improve their research."

What are the lessons learnt by the Summer School participants? Here are some highlights from their reflection reports (Blessing and Andreasen 2005):

- "…the course has forced me to awaken and at least start the process (of questioning points and issues of research) off explicitly, so that my focus becomes more directed and refined at an earlier stage of my PhD."
- "I changed my approach after the course (more rigour, focus, use of literature, start writing now)."
- "I have self-confidence now and can structure my research."
- "Now I know that creating something is not science; the reasoning must be scientific."
- "I thought I could be original by not reading literature."

The feedback session on the last day of the Summer School provides more material on evaluation. The participants are asked to comment on what they *liked*, and what they *would have liked.* The major comments are that the PhD students liked: the guidance provided for developing the research plan and structuring their research; experiencing that most have similar problems; the use of DRM to describe their work and helping to make it more consistent. Many would have liked another week after a year to obtain feedback on their progress and most would have liked more details in written form. This is echoed in a comment made by Flanagan and Jänsch (2004): "The lectures provided the students with a much wider view of the scope and interpretation of the DRM methodology than can be obtained from literature."

8.1.2 Feedback from the Design Research Methodology Course

The one-semester Masters level course on Methodology for Design Research is a core course for research students at the Centre for Product Design and Manufacturing at the Indian Institute of Science, Bangalore. An analysis of the course spanning 5 years offers further indications about the benefits of using DRM (Chakrabarti 2008).

The students who did the course expressed, in obligatory, but anonymous, post-course questionnaires, that the content of DRM had been helpful in clarifying their research process and in equipping them for doing design research. The course is consistently rated between 4 and 5 on a scale of 0–5, where 5 is considered 'excellent'. Some of the comments from the feedbacks received from the students are:

- "I found the course difficult to start with, but eventually the structure of DRM became part of my regular research."
- "Retrospectively I see the structure of DRM as logically obvious."
- "It was useful in relating apparently disparate pieces of research papers."
- "I find myself confident in carrying out research in this area after going through the course."
- "I use DRM as a framework for my thinking, especially for developing and testing design support."

In addition to this, the efficacy of DRM was evaluated in a subsequent research course called 'Design & Society', by comparing the grades of students who were taught DRM to those who were not, over a period of 5 years, involving a total of 30 Masters students. In this course, the students have to take up independent design research under the supervision of any of the departmental faculty members for an 18 week (9 hours per week) period. Those students who were taught DRM, either formally through the above 'Methodology for Design Research' course, or as part of their research for 'Design & Society' under the supervision of the instructor of the 'Methodology for Design Research' course (who ensured they understood and used the structure of DRM as part of their research) were, in general, found to do better than those who were not taught DRM.

The students who were taught DRM had an average grade point of 6.7 for their 'Design & Society' course (on a scale of 1–8, where 8 is considered outstanding, 7 excellent, 6 very good, 5 good, 4 satisfactory and 0–3 fail). Those who were not taught DRM had an average grade point of 5.8. Furthermore, those who were taught DRM had their 'Design & Society' grade *identical* to their average grade (6.7) over their entire Masters course, while those who were not taught DRM had their 'Design & Society' grade *well below* their average Masters grade (6.3). We argue that this is an indication of the positive influence of the students' knowledge and use of DRM on the quality of research they carried out. For students who attended both the 'Methodology for Design Research' course and the 'Design & Society' course (in that order), there is a strong correlation (Pearson's $r = 0.82$)

between their grades in these two courses ($p < 0.02$). The work of several students who were taught DRM resulted in a research publication. If these are initial indicators of rigour in design research, they point towards DRM as a positive influence.

8.1.3 Other Sources

There are several instances that suggest that DRM provides a useful context for positioning research projects and programmes in design. We used it (Blessing *et al.* 1998) to categorise the studies described in a book on empirical design studies (Frankenberger *et al.* 1998), similar to the way we categorised the PhD projects in Appendix C. Wood and Greer (2001) who were not involved in the development of DRM, used DRM to categorise all the papers in their review of Functional Reasoning Methodologies. Several researchers used the DRM framework and associated methods in their PhD, *e.g.*, Almefeldt (2005); Dagman (2005); Eilmus (2007); López-Mesa (2004); Nidamarthi (1999); Preiss (2006); Sarkar (2007); Weinert (2001). Other resarchers and research groups have used the DRM framework in project proposals to outline the research approach. The examples show that DRM can be used for a variety of design research topics, covering the various facets of design mentioned in Figure 1.1.

The Reference Model has also been used in other areas of research. Our students have applied this model in engineering science research: *e.g.*, for mapping the factors that potentially influence the fatigue limit of serrated shaft–hub connections in order to focus the experiments and integrate the research results (Blessing *et al.* (2006), and for mapping the factors involved in testing seat comfort in order to identify requirements for new test equipment (Eilmus 2007). In an interdisciplinary research group at the Technical University of Berlin, working on micro-energy systems in structurally weak regions, the development of Reference Models by the individual PhD students and the subsequent comparison of these models lead to the identification of potential synergies with respect to, *e.g.*, data collection and triangulation.

One of the consistent observations by the students in the Summer School and DRM course is the lack of material on DRM other than the lecture notes. The papers we wrote and the lecture notes are seen as helpful but inadequate in understanding and using DRM in depth. This could have been why some aspects and terms were sometimes misinterpreted. This concerns in particular the term Measurable Success Criteria, which has been misinterpreted as focusing on quantitative methods, and thereby devaluing qualitative methods. Also, the DRM process has sometimes been interpreted as linear, focusing only on the process aspects of design and seen applicable only to individual research projects rather than research programmes (see our Preface for a more detailed discussion). These interpretations not only do not represent our view on design research, but are also in direct contradiction to what we have written and to our own research activities. Nevertheless, these critiques have emphasised the need to be more explicit about our view and to provide more examples.

8.2 Further Research

Observing the application of our approach in the Summer School, in the DRM course, and by our own students, as well as the many discussions, critiques and questions, have been invaluable for the development of DRM. They have pointed out which aspects the research community finds particularly important. As a result we have tried, amongst others, to focus more strongly on the most problematic issues; to bring clarity into the terminology and concepts used in DRM; to explain the underlying assumptions; and to refine our methods. We hope that this book will help strengthen the judgment and ability of researchers to deal with the complexities of design research, and encourage reflection about the ways of doing research in design.

Nevertheless, DRM is by no means complete. The book intends not only to provide some of the methodological materials currently lacking in design research, but foremost to stimulate discussions about methodologies and methods for design research.

Note that one area of design research is not covered in this book, namely research into DRM and research methods. As Reich (1995) states: “Researchers may find it fruitful to study: the objectives or goals of engineering design research; how these objectives can be fulfilled through research; how progress towards research goals can be tested; and how this overall process can be improved. Such study will evolve a repository of methods with their assumptions, interpretations, successes and failures.” This involves addressing methodological questions such as: ‘How can we generalise our findings given the uniqueness of each design project?’ ‘How is the quality of the product and process to be measured in order to evaluate a particular support? ‘How can a situation in which support has been introduced be compared with a situation in which this support has not been used?’ Answering these types of question is very important for improving the quality and determining the applicability of approaches, methods and guidelines used in design research. We intend to continue our research into DRM as much work is still to be done. We therefore welcome any feedback from our readers on DRM and its methods.[20]

---

[20] For those readers interested in what happened to the protagonists of our examples: they met at a design research conference, and with her synthesis and his reliability competences, they lived happily ever after.

Appendix A

Descriptive Study Methods

The focus of this appendix is on the methods used to obtain a better understanding of design through empirical studies, as in the DS-I and DS-II stages. As a researcher embarking on empirical studies, it is necessary to understand that methods are often based on particular assumptions about how to execute research. Section A.1 contains a short discussion about paradigms and assumptions. In Section A.2 the checklist for reviewing empirical studies discussed in Section 4.4.2, is described in more detail and closes with an example in Section A.2.21. Because of its relevance, the differences between empirical studies in laboratory and industrial environments are described in Section A.3. An important characteristic of empirical studies is the method by which data is collected. Section A.4 provides short descriptions of some of the main data-collection methods, based on the literature and our own experiences. Some basics about statistics are given in Section A.5.

In all cases, further literature and experts in the field have to be consulted when setting up an empirical study. This appendix is only an introduction, intended to raise awareness of the issues.

A.1 Paradigms and Assumptions

In design research a variety of topics can be studied for a variety of reasons with a variety of methods. The research approach and methods applied should be chosen such that they are suitable for the topic, the aim and the existing understanding, and that they result in valid statements. For that purpose, methodologies for research have been (and are still being) developed in and across disciplines based on what is viewed as valid research. Design research has many facets as shown in the first figure in this book (Figure 1.1), representing various disciplines. Some disciplines have well-established research methodologies, others have several, sometimes conflicting approaches causing heated discussions about which approach is more likely to produce the best, most valid understanding. This is clearly visible in the debates around quantitative versus qualitative research, and theory-driven versus data-driven research (see also Section 4.1).

As a researcher, it is important to be aware of the fact that:

- different schools of thought exist based on underlying paradigms;
- every school has (or should have) a consistent methodology that links the problem to the methods applied and the ways of validation; and
- every methodology has certain premises.

Furthermore, multiple standpoints are found with 'followers' of well-established approaches, and taken-for-granted categories and methods of data collection have become problematic (Coffey *et al.* 1996). As discussed in the main text of this book (Section 4.1), paradigms do change over time and new ones emerge and competing paradigms may exist simultaneously, specifically in immature sciences (Kuhn 1970). For example, the view that it is not possible to be completely objective while observing (which was the traditional scientific pre-requisite) has become more and more accepted in all sciences. The so-called givenness of data, whether numerical or not, is questioned. That is, the view that sources of data can be treated as independent of, and as imposing themselves on, the researcher, is generally rejected (Hammersley and Gomm 1997).

Independent of a particular paradigm, groups of researchers or individuals are likely to have particular views (assumptions) about the topic of investigation – even if this is just the belief that a particular factor is relevant. This influences the interpretation of the findings. These assumptions can hardly be avoided, but should be made explicit, and an attempt should be made to consciously use alternative views to find alternative explanations.

A.1.1 Paradigms

The differences in research approaches and methodologies are due to the underlying paradigms (also called worldviews or belief systems), that are used by a domain or by groups of researchers within a domain, as the accepted perspective at a given time. The paradigms express the basic assumptions upon which the research is built and should therefore be known when applying the related approach or its methods.

To illustrate this, we give two examples of particular views shared by a larger research community and both referring to the role of experience. The philosophical theory of empiricism has been the basis for several paradigms. According to Bullock *et al.* (1988) empiricism states "(1) that all concepts are derived from experience, *i.e.*, that a linguistic expression can be significant only if it is associated by rule with something that can be experienced, and (2) that all statements claiming to express knowledge depend for their justification on experience". Empiricists assume that this gives "access to the neutral or unalloyed access to a realm of pure 'facts'" (Bullock *et al.* 1988). The focus of empiricism is on objectivity. "Any statement of the empiricist theory, to be consistent with itself, must be empirical or, if not, analytical. An empirical basis for the theory is provided by elementary facts about the way in which the meaning of words is learned." (Bullock *et al.* 1988).

A different view, also referring to the importance of experience, but resulting in a very different research approach and methods, is ethnomethodology: personal experience of a situation is considered crucial to developing knowledge about the situation. The focus is on "how the individual experiences and makes sense of social interaction" (Bullock *et al.* 1988). To obtain understanding it is important to immerse oneself in a situation; to live it. This refers to a meaning of experience quite different from that from an empiricist view: namely referring to a more subjective view. The focus is on details of everyday life, in particular on details of the practices through which action and interaction are accomplished (Button and Dourish 1996). The determination of what is going on and what is appropriate is considered to be the outcome of local, occasioned and situated action (Anderson 1996).

These examples show that world views are not necessarily applicable to all situations. Ethnomethodology, *e.g.*, would not be suitable to gain understanding about the reliability of products from a technical point of view. However, to understand the reliability of products from the point of user experience, this approach could be very suitable.

Paradigms might also specify the main concepts used. Cook and Campbell (1979), *e.g.*, present some of the different views on the concept of cause.

Even if one is not aware of such paradigms, or does not wish to follow a particular paradigm, one's view on how to do research, on the suitability of methods and on the topics to be studied, is coloured by beliefs; beliefs about the right, scientific way of obtaining understanding. Despite the best intentions, no researcher is free from 'coloured glasses'. The least researchers should do is to acknowledge this fact and make deliberate attempts to make their views, assumptions and beliefs explicit. This will improve the quality of one's own research and help others to decide whether or not to take up a certain view and resulting approaches.

"A world view has a profound interaction with the research questions we choose to investigate. Some questions are interesting or even meaningful only within a particular world view, and some questions limit our horizon to seek alternative world views" (Reich 1994). Cook and Campbell, *e.g.*, state in their book on quasi-experimentation that they adopt "a critical-realist perspective, positing that causal connections are 'real' but imperfectly perceived, and particularly stress epistemological[21] theories that restrict the analysis of causation to the analysis of manipulable causes – factors that can be varied 'at will'" (Cook and Campbell 1979). Denzin states in his book on social methods that he "offers a symbolic-interactionist view of sociological theory and research methodology …… which stresses the self-reflexive nature of everyday and scientific conduct" (Denzin 1978). Patton (1990) describes his particular view on the most suitable approach to research as follows: "the qualitative-naturalistic-formative approach is especially appropriate for programmes that are developing, innovative or changing, where the focus is on programme improvement". These views not only affect the research questions, but also the choice and intended application of the methods they propose.

---

[21] Epistemology is the philosophical theory of knowledge, which seeks to define it, distinguish its principal varieties, identify its sources, and establish its limits (Bullock *et al.* 1988)

Many researchers consider the combination of different paradigms impossible, as many debates about paradigms show. Research groups tend to hold on strongly to the paradigm of their research field or even their own group. Some researchers advocate the combination of different paradigms (which in itself is a paradigm!). Tashakkori and Teddlie (1998), *e.g.*, propose “a pragmatic view, combining the quantitative positivist paradigm with the constructivist qualitative paradigm” – which they consider compatible – as the basis of a mixed methodology to be applied across all phases of the research process. In their view, methods should be chosen such that they are most appropriate for the research question. This question should predominate over the paradigm. This very much reflects our view.

As a design researcher it is not necessary to join in the debates in the various disciplines about the best methodology, but it is important to read multiple sources before choosing a particular approach to ensure that the data obtained using this approach is valid for the purpose intended and that pitfalls that make data invalid are avoided. Also in the area of design research, certain approaches, methods, types of questions and issues will be considered particularly relevant. Here too, one should be aware that – even if not explicitly stated – relevance strongly depends on the paradigm used by the researcher, the particular facet of design studied, and the research questions and hypotheses addressed.

In design research one can observe at least two views of design that play a role in the research approach taken: a process view and a product view (the first two project descriptions in Appendix C of this book provide an example of each. These views are often considered opposite and mutually exclusive. A process view assumes that the process (the context in which the product is being created) determines the product, and therefore that a better understanding of the process enables process improvement that will then lead to more successful products. The product view can be described as assuming that the product determines the process. Better understanding of the product will make it possible to determine the most suitable process for generating more successful products. We argue that our methodology is applicable in both cases.

A.1.2 Assumptions

To clarify one’s own view and assumptions, one can search for and look into existing theories (*i.e.*, views of reality) that seem appropriate to one’s own research, discuss ideas and approaches with others, and ask the ‘why’ question: ‘Why do I think this is the case?’ ‘Why do I think this is relevant?’. Researchers have to be reflective and self-critical about their approach as a whole, and about each individual step. That is, apart from paradigms, or world views, on the research approach, one also has views (assumptions) on the topic of interest. For example, in interpreting findings, a researcher should consciously try to generate alternative, or rival explanations, in the same way as designers are encouraged to generate alternative concepts for the task they are given. It has to be said that it is unlikely that all assumptions can be made explicit, and many may not become apparent until much later in the research process, for instance while generating these alternative explanations.

A simple example of how certain assumptions might influence our interpretation of findings is the following. Assume that, in our view, designers identify themselves with the product they are developing – it is 'their child' – and, as a consequence, always strive to generate the best solution and tend to experience constraints such as time as impeding their process. In a study we collect evidence that those designers who complain about time pressure also produce products of low quality. Based on our assumption, a plausible conclusion would be that shortage of time relates to or even causes low-quality products. Another possible explanation is that the designers were not motivated and used time constraints as an excuse, but this does not fit our assumption. This second explanation would, however, have been plausible had the basic assumption been that employees – and the designers in the study were employees – tend to consider their work as an activity to earn money, *i.e.*, do not have an intrinsic motivation nor identify themselves with their work. A further possible explanation is that another factor played a role, namely inexperience. If designers are too inexperienced for the type of product they have to develop, the quality might be low and they might have problems with time constraints. This would be based on the assumption that designers are not necessarily experienced.

It is therefore crucial to identify and make explicit any assumptions one has made during any of the stages of the research.

A.2 Reviewing Empirical Studies

In Section 4.4.2, the checklist shown in Table A.1 was proposed for reviewing descriptive studies.

Table A.1 Checklist for determining the characteristics of empirical studies (Table 4.1) (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 Measured Success Criteria and possible constraints.
Nature of the study	Observational or interventional, comparative or non-comparative, (*i.e.*, whether the study involved intervention in the design process by the researcher).
Theoretical basis	Paradigms, methodologies, theories, views, assumptions, *etc.*, that guided the researchers.
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 using video, participant observation, diary keeping, archival research, questionnaire, interview.

Table A.1 *(continued)*

Dimensions	**Options**
Role of researcher	Involvement of the researcher in the research process.
Time constraint	Time constraints imposed by the researcher, *e.g.*, available design time, available time to answer a questionnaire, time of the observation (when the phenomena observed lasts longer).
Continuation	Continuous data collection or sampling
Duration	Length of the part 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.
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 analysed
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, models, 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 dimensions in this table offer various options; those listed in the table are not exhaustive. None of the options can be said to be right or wrong, or providing stronger or weaker statements, as this depends on the research questions and hypotheses addressed with the particular study, and on the combination of options that are chosen. Note that the dimensions are not independent and the specifications can overlap. Furthermore, not all categories may apply to each type of study.

If the checklist is used to set up an empirical study, it is important to be creative in specifying the options; the options and their realisation have to be adapted to the aims of the study, without compromising the validity of the study and its results.
To illustrate the use of the checklist an example is given in Section A.2.21. For an overview of empirical studies using earlier versions of the checklist, see Blessing (1994) reviewing 66 studies and Dwarakanath *et al.* (1995) reviewing 90 studies.

The following sections provide a short explanation of each of the dimensions in this checklist.

A.2.1 Aim, Research Questions, Hypotheses

This dimension lists the aim(s) and objectives of the research project and of the empirical study itself, as well as the the main research questions and hypotheses. If Success Criteria and/or Measured Success Criteria have been defined, these can be added here, as well as the main constraints relevant for the study.

A.2.2 Nature of the Study

The nature of the study reflects the two types of empirical studies in DRM: DS-I and DS-II. A distinction is made between observational and interventional research, each of which can be comparative. *Observational* research here does not refer to observation as a research method, but to a hands-off approach by the researcher. Observational research is typical for DS-I. The researcher does not intervene, *e.g.*, by introducing design support. The objective is to identify the influencing factors and their links by studying design 'as is'. Methods and setting used for empirical studies will often influence what is observed, but that is not considered to be an intervention in the sense used here. The method and setting should be chosen such that the effects are minimal or not relevant, or that the effects can be taken into account in the interpretation of the findings. *Interventional* research deliberately influences the existing situation in order to investigate the effects of this intervention. This type of research is typical for the evaluation studies in DS-II.

A study is called *comparative* when the cases are divided into groups with different characteristics, *e.g.*, different tasks, different settings, or different subject backgrounds, *before* data has been collected. The objective is to find the effect of these differences so as to better identify the influencing factors and their links. In the case of interventional research, a comparative study might divide the cases into those that use the support (the so-called *experimental group*) and those that do not (the *control group*) to better identify the effects of the support (see also Section A.4.3). A comparison can also involve the same users, comparing before and after the use of a particular support. The latter, however, is not very common in design research because of the difficulty to compare the situation before and after a design problem has been solved.

A.2.3 Theoretical Basis

Any paradigm, methodology, theory, *etc*., that guided the researcher, as well as any relevant views and assumptions have to be made explicit. Unless explicitly mentioned, it may be difficult to determine these from research papers, other than indirectly through the research approach and methods that were used.

A.2.4 Unit of Analysis

A central concept in research is the *unit of analysis*. It is the main element of the study about which the researcher wants to obtain information, wishes to draw conclusions and make generalisations, *i.e.*, the unit on which the analysis focuses (hence the name). The chosen unit of analysis should allow independent observations. The product can for example be the unit of analysis, if the researcher aims to identify the factors that contribute to its attribute 'quality' or its attribute 'reliability'. The designer would be the unit of analysis, if the aim is to draw conclusions about differences between designers, *e.g.*, how they approach a design problem. The design team would be the unit of analysis, if we aim to draw conclusions about the behaviour of design teams, for example the way in which team members collaborate. Note that in the last example, data is collected at the individual level, *i.e.*, the *unit of measurement* (or *unit of data collection*) is the individual. This data is aggregated and analysed at the team level. Hence, the unit of *analysis* is the team. The unit of measurement can thus differ from the unit of analysis. The unit of measurement relates to data collection, in this case the individual; the unit of analysis relates to data analysis, in this case the team about which we want to draw conclusions.

There are no limits to what can be the unit of analysis, but once chosen, this affects the other dimensions of the study, such as the most suitable data-collection methods, the nature of the study, *etc*. The unit of analysis also determines which conclusions can be drawn and should thus be chosen carefully. For example, the fact that a design team has generated many solutions does not imply that each designer in the team has generated many solutions. That is, the conclusions drawn from an analysis at team level may not apply at the individual level and *vice versa* (doing so is referred to as *ecological fallacy*).

The unit of analysis has to be chosen such that the research questions can be answered and the hypotheses verified with as little interpretation as possible. Units of analysis used in design research are manifold and include: design team, requirements, product module, design process, decision making, human–machine interfaces, information exchange, collaboration, documentation, and organisational strategy.

A.2.5 Data-collection Method

Literature provides a variety of methods for data collection and recording, each with its specifics and limitations, *e.g.*, direct observation, participant observation thinking aloud, introspection, diary keeping, document/drawing analysis, questionnaire, interview, product analysis. Section A.4 discusses the most common ones. In most studies a combination of methods is used in parallel or sequence, *e.g.*, a questionnaire about the participants' background, followed by the observation of a task using video and note taking, and ending with a reflective interview. Drawing up a matrix showing the research questions and hypotheses against the methods used, as shown in Figure 4.6, can be useful.

A.2.6 Role of Researcher

The way in which the researcher is involved in an empirical study can influence the outcome, even in pure observational studies, and should thus be specified in detail.

For example, in a study in which video cameras are used to observe individual designers designing a product against a particular specification, the role of the researcher is essentially that of an observer. However, experience shows that designers tend to ask questions about the problem, about the background of the assignment, *etc*. A decision has to be taken as to how to deal with such a situation: not giving an answer may block the process and take away some of the reality (in practice designers do ask others), whereas giving an answer will influence the course of the process, but will be closer to reality. More importantly, answering will provide the particular designer(s) with information that the other designers do not have. One solution is to write down, prior to the first observation, all possible questions one could think of by trying to anticipate what designers would wish to know when designing this product, and to formulate an answer for each question. When a question is asked that appears in the list, the answer is read out without rephrasing, explaining or adding. If the list does not show the question, the researcher tells the participant that he or she does not know the answer. Importantly, a note is made about when which questions were asked, as this may be important in the analysis and interpretation of the data. In this way, all participants will be able to get the same information. Obviously, if the same question is asked over and over again, it can be useful to inform each following participant.

Another method, in which the researcher plays a clear role, is participant observation. Here too, it is important to specify the details of the role of the participant as the following example shows. Hales, in the example used in Section A.2.21, was the main consultant designer in the project he observed. He was continuously involved in most of the issues. This allowed insight he would not have been able to obtain had he only been in a supporting role dealing with a particular design issue, such as was the case in the industrial study undertaken by the first author (Blessing 1994). The specifics of the role thus influence the results and conclusions that can be drawn.

The role of the researcher at the start of the study can also have an influence. It is important to think about and make explicit the way in which participants are contacted and informed, for instance by asking questions such as: Are the participants aware of the aims of the study? Has the information been provided to all participants in an identical manner? Has the contact between the researcher and participants been direct or has, *e.g.*, the company selected the persons? Instructions should be in writing, irrespective of whether the instructions are handed out or read out aloud by the researcher.

A.2.7 Time Constraint

This dimension provides information about any time constraints that were imposed by the researcher. This includes times given to designers to solve a particular problem or to interviewees to answer the questions, as well as the time period during which data about product or process was collected when the phenomena studied covered a longer period.

Any time constraint can influence the process and the results. For example, the last questions on long questionnaires often have no or few answers, or are answered in a hurry, without much thought. As a result, parts of the questionnaire might not be useful. Limiting the time used to analyse the consequences of new features in a product, may only allow detection of failures that occur early on in the product life-cycle.

When the period in which data can be collected is constrained in the sense that it is shorter than the period in which the phenomena observed, it has to be clearly specified what is covered (see also Section A.2.9).

A.2.8 Continuation

This dimension clarifies whether data was collected continuously or whether sampling took place. For example, if data is sampled at certain intervals, the intermediate periods are not covered and it may be difficult to draw conclusions about the whole period (additional methods may be used to fill these gaps). This dimension is also relevant when the process is in principle observed continuously, but involves breaks that are not captured. Breaks, whether in observation or interviewing, allow participants to reflect, to meet others, *etc*. When observing designers at work, the design process is likely to continue during the breaks ('I got that idea when I was taking a shower/met my colleague over lunch/saw the vending machine when I got a coffee'). Whether it is problematic or not if this is not captured, depends on the aims and specific research questions of the study. If discontinuity is problematic, observation of shorter processes, a different setup (*e.g.*, lunch in the room), or the addition of interviews after the breaks, are possible solutions.

A.2.9 Duration

The duration is the length of the process or project analysed in terms of time. A difference may exist between the length actually observed and the total length, *e.g.*, due to breaks, in which case both have to be specified.

A.2.10 Observed Process

This dimension mainly applies to observational or archival studies. In order to determine what the results of a study relate to, it is important to know what the starting point (input) and required deliverables (output) of the analysed processes are. In an artificial environment, a requirements list could be the input and a layout drawing requested as output. In a real environment, when only part of an ongoing process can be studied, a distinction can be made between the input and output of the actual process and those of the observed process.

A.2.11 Setting

This dimension specifies the setting in which the study takes place. This includes the location (the actual physical environment), as well as whether the setting is contrived or natural. In a survey, for example, the answers and their length may differ, depending on whether people are interviewed on the street or at home. The effects of a laboratory or practical setting are manifold, as discussed in Section A.3.

Studies that take place in laboratory environments are often contrived, but not necessarily so. Similarly, a study in which designers are observed within their company is a study in an industrial, but not necessarily natural setting. For example, if designers are observed in their company, but in a room in which they do not normally work and without access to the materials and people in their normal environment, this setting can be considered contrived, if this setting may affect the conclusions that can be drawn from the study. The setting has therefore to be chosen carefully: it should either have no influence, or the influence should be known such that it can be 'filtered' out or taken into account when analysing the data and drawing the conclusions.

A.2.12 Task

In many cases it is important to specify the task(s) to which the data refer. In an observation in practice, which will often concentrate on few people, it is important to know their particular tasks within a project; for example, if the persons observed are the analysis experts, one should not be surprised that the findings show an emphasis on analysis activities. The particular task and related materials determine what can be expected and what generalisations can be made. The task also plays a role in archival analysis of product or process documents, as these often only give insight into the results of particular tasks or specific views: it might, *e.g.*, be difficult to find alternatives that were considered.

The complexity of the task can also be relevant. In a contrived setting, indicators might be the number of hours required to complete the task and the knowledge and skills required. In a natural setting, the indicators may be: the number of hours and people involved; the number and types of disciplines involved; the novelty of the product, *etc*. The task should be compatible with the aims of the study and – in particular when the task is given – with the participants involved.

Furthermore, it is important to determine whether the task is *real*, *realistic* or *artificial*. A real task is one in which the participants are normally involved and is the basis for direct observation in a natural setting, for archival analysis, and for interviews and questionnaires. In our reliability example, studying how designers assess reliability while working in their own project represents a real task. When the researcher provides a task, introduces constraints, or changes the setting – for example to obtain data sets that are more easily comparable or to reduce the length of the process – the task can be called realistic, provided that the task is a derivative from a real task. In our reliability example, a task could have been devised that is similar but more focused than the one on which the designers normally work. A task would be artificial, when such a task does not occur at all in practice or not in this form. In our reliability example, this would be the case when the designers are given various product drawings and asked to assess reliability. Which type of task is most suitable depends on the aim of the study and the specific research questions and hypotheses to be addressed. Note that the way in which the task is described should be considered carefully, as this can influence the process adopted by the participants (see, *e.g.* Fricke (1993); Guindon (1990)).

Furthermore, it should be taken into account that a realistic or artificial task may not be taken seriously as the usual motivation and pressure are missing, in particular because there are no consequences if the task is not done properly, and the tasks given are often relatively small and not very challenging. In a real task, the consequences of incorrect decisions can be serious and the tasks can be very challenging. A lack of motivation can, *e.g.*, occur when the participants have been asked to participate by the company. On the other hand, we have also seen an increased level of motivation and pressure caused by the fact that participants are observed and their work and working analysed. It is important to realise that motivation can change during the data collection period, in particular when the task (design task, interview, questionnaire, *etc.*) takes very long to fulfil. A short interview or questionnaire at the beginning and at the end of the study may reveal some of the motivations and state-of-mind of the participant.

A.2.13 Number of Cases

The number of cases refers to the number of data sets *collected*, *e.g.*, the number of experiments, interviews, projects, documents, products, individuals, groups or companies. The number of cases is important to determine statistical significance. In the analysis, the number of cases may differ depending on the research question or hypothesis addressed. For example, in a study of design teams, one research question may be about the differences in quality of the work of design teams. The number of cases for the analysis is the number of teams. Another question may relate to the individuals within the teams, such as differences in the contributions of each of the group members. This analysis would involve many more cases, namely the number of individuals involved. It is important that the data allows analysis at different levels (see discussion on unit of analysis in Section A2.4).

A.2.14 Case Size

A collected case (*e.g.*, a team or product) can consist of one or more elements. (*e.g.*, individuals or components). The case size is the number of elements within each case. This dimension can be relevant for the analysis, interpretation and generalisation of the results. For example, groups of two persons will work differently from groups of ten persons and the situation in small companies will differ from that of large companies. Within an empirical study, the sizes of the cases can differ: the number of people in each team or the number of components in each product may not be identical.

A.2.15 Participants

It is very important to know who participated in a study. Their experience, education, current position, motivation, culture, *etc.*, are of interest, as these could help explain the findings. If students have participated in a study, which is very common in design research, generalisation to practising designers might not be possible, and the answers in a questionnaire sent to a company may differ, depending on whether they come from a manager, a designer or a marketing specialist.

The characteristics of the participants that are relevant to know are those that could influence the outcome of a particular study. Experience is a typical influencing factor in design, but needs to be operationalised carefully. The number of years a designer has been working is an oft-used measure of experience. However, if the task used in the study is from a completely different domain, this experience may be of little use. Furthermore, we have found that senior managers can have a long experience, but might not have been designing for some time and thus lost some of the knowledge and skills necessary for a particular task.

A.2.16 Object

Apart from participants, the cases can (also) involve products, projects, companies and documents, as well as design support. Their characteristics have to be specified too, as far as these might influence the outcome.

For products to be designed or analysed, this includes: the name, the number of components, the type (original, variant, redesign), the batch size (mass, large batch, small batch, one-off) and the disciplines involved (mechanical, electro-mechanical, mechatronic, medical, *etc.*). In a similar way, the relevant details of companies, projects, documents, and other objects have to be specified. These details are necessary to generalise across objects. The required details of a design support to be evaluated in DS-II, can be found in Chapter 5.

In interviews and questionnaires it is important to collect details about the products, projects or companies on which the individual answers are based, in order to correlate the answers to different questions. Most correlations are only possible if the answers relate to the same object (see also Section A.4.9).

A.2.17 Coding and Analysis Method(s)

In most publications in the area of design research, the coding and analysis methods used are not specified, although these are essential for an interpretation of the data. Details of the processing, coding and analysis methods have to be documented. Some of the relevant aspects can be found in Sections 4.7.2 and 4.7.3.

A.2.18 Verification Method(s)

Methods used to verify the results of the study have to be specified (see details in Section 4.7.4). In the area of design research, this is unfortunately not often done.

A.2.19 Findings

To complete the description of an existing empirical study, it is useful to summarise the main statements, models, theories, and conclusions resulting from the study. For the planning of an empirical study, this dimension is used to document the expected (types of) statements, models, theories, and conclusions.

A.2.20 Notes

While reading about a particular study, questions may arise about particular details, information may be missing, inconsistencies may be found, certain methods may be particularly interesting, or differences and similarities with one's own study may be noticed. This section is intended to make any such notes. If the study seems relevant for one's own research, contact should be made with the authors to clarify any outstanding issues. Most authors will be pleased to explain their work.

A.2.21 Example

Table A.2 illustrates the use of the checklist to summarise the study of Hales (1987), who analysed an engineering design process in an industrial context.

Table A.2 Example use of the checklist for empirical studies based on the study described in Hales (1987)

Dimensions	**Options**
Aim, research questions, hypotheses	Better understanding of an engineering design process in industry in which a systematic approach was introduced. Several research questions, in particular the identification of the factors that influence the design process.
Nature of the study	Non-comparative, single case study, interventional (systematic approach of Pahl and Beitz was introduced).
Theoretical basis	In particular Grounded Theory (Glaser) and Systematic Design (Pahl and Beitz).
Unit of analysis	The stages in the design process (unit of measurement, a.o. the individual participants).
Data collection and recording	Participant observation, using diary notes, audio tape recordings, weekly reports and design reports.
Role of researcher	Researcher was the main contract designer employed in the project.
Time constraint	Time constraints were not set by the researcher, but by the company as project deadlines.
Continuation	The work on the project was restricted to one day a week. Data was collected every day on which work on the project took place.

Table A.2 *(continued)*

Dimensions	**Options**
Duration	The whole design project was studied, covering 36 months, and 2368 hours. At that point the company decided to stop the design project.
Observed process	From initial proposal (planning stage) to near completion of detail design.
Setting	The project took place in a large company in a natural setting.
Task	The task was real.
Number of cases	One project in a large company.
Case size	The project team consisted of 37 people.
Participants	Varying backgrounds and positions within the company and a contract engineer (the researcher) who was the main engineering designer.
Object	The project involved the development of a high-pressure and high-temperature system for evaluation of materials in a simulated slagging coal gasifier environment, to be used within the company. The design was original and one-off.
Coding and analysis method(s)	Notes were colour coded according to participant. Data was continuously entered in a database for indexing, sorting and categorising (using the structure of systematic design as proposed by Pahl and Beitz, and a list of 103 factors likely to influence the engineering design process drawn from the literature). Data was then transferred to spreadsheets for numerical and graphical analysis. Quantitative and qualitative analysis methods were applied.
Validation method(s)	Comparison with the literature.
Findings (main)	Ideal engineering design projects may be characterised by a series of overlapping, bell-shaped 'phase curves' each representing the work effort in a particular phase of the project along a time axis. Setting up an ideal phase diagram for a project provides a model against which to measure actual performance. Design work not completed within the envelope of the ideal phase curves for a particular project will have to be completed outside the envelope. This causes diversion of effort and thus increases the cost. The use of methods and aids commonly described in the literature accounted for less than one-quarter of the observed engineering design effort. A further 13 categories of design-related communicating, working, and motivating techniques were identified that accounted for the rest.
Notes	An extensive list of influencing factors was derived from the literature and this study. The fact that the project only run one day a week allowed time for processing the data in much detail.

A.3 Laboratory Versus Industrial Environment

Analysing real design processes in a practical environment is a type of field research. Field research in general has many advantages. The main advantage is that results are based on reality. The major disadvantage of field research is that the activities of the researcher can hardly be planned and defined when starting the research project, and that there is no guarantee that results will be useful, or the observed situation will continue and without interruption. As the example in Section A.2 showed, a design project might be cancelled. Furthermore, qualitative field research is fundamentally of an improvising nature and depends on the specific situation and on knowledge that has yet to be processed (ten Have 1977).

The difficulties of research in professional fields such as design, is that these fields are relatively closed and have a high degree of organisation, making them difficult to enter (Friedrichs and Luedthe 1975). The literature discusses the many obstacles encountered during the introduction in such a professional field. A major obstacle in design research, not mentioned in the literature on field research, can be the time required from the observed designers in an industrial setting. From the company's point of view this implies a financial commitment: time is money.

In Tables A.3 and A.4, we have summarised the differences between a laboratory and an industrial setting for design research, as we and our colleagues have experienced these: not to discourage field work, but to raise awareness of the issues that have to be taken into account in the preparation of such study. Which setting is more suitable depends on the aim and the research questions and hypotheses. Several of the differences also apply to other practical settings, such as class rooms or design situations outside industrial contexts.

A.4 Data-collection Methods

Several books exist in the social sciences that provide an overview of available strategies and methods for collecting, analysing, interpreting, and evaluating data. These specialist books greatly facilitate the selection of potentially suitable methods and help clarify the underlying paradigms and assumptions, which will influence the recommended use. We found the following books particularly useful: (Cook and Campbell 1979; Frankfort-Nachmias and Nachmias 1996; Patton 2002). The latest Handbook of Qualitative Research of Denzin and Lincoln (2005) contains interesting but advanced discussions by various authors of different qualitative approaches of inquiry and their methods of data collection, analysis, interpretation, evaluation and presentation.

Care should be taken to consult primary and not secondary sources. Secondary sources, such as this appendix and the literature describing specific empirical studies, can be useful to find out about the experiences of those who applied the methods, but cannot be considered authorised texts on those methods. For similar reasons, care should be taken using the Internet, and the background of the authors checked for their experience and research area.

Table A.3 Differences in process of observing: industrial versus laboratory setting

Process of observing	
Industry	**Laboratory**
Difficult to plan.	Time and location can be determined in advance.
Difficult to control • interrupts from others; • usually no explicit starting or end point (when did the first idea come up?); • topic cannot necessarily be chosen.	Easy to control • interrupts can be avoided; • at least starting point is known; • topic can be chosen.
Continuous observation may be difficult. Tasks may be abandoned or stopped for a short while or for months.	No or hardly any interrupts if carefully planned. However, no positive effects of interrupts either (talking to others, sudden insights through other activities, gestation of ideas)
Interference with existing processes can cause problems with obtaining allowance for the study (time = money). In particular, recording equipment can interfere or may not be allowed.	No interference with existing processes.
Confidentiality of observed processes and results may hinder data collection and publication of results. Anonymity of participants to the outside is possible, but difficult to achieve internally.	No confidentiality issues regarding the task, although intellectual property rights may have to be clarified beforehand. If these do play a role, publication of the results may be subject to limitations. Anonymity of participants plays a role.
Multiple studies on the same topic impossible or very difficult.	Multiple studies using the same topic but other participants is possible.
Existing environment not easily optimised for observation.	Environment can be optimised for observation.
Work on one task can extend over months requiring specific data-collection methods. Very difficult to capture all work on a particular topic, as many people may be involved and work may be discontinuous. Duration, *i.e.*, involvement of participants, difficult to predict.	Limited duration possible. Task can be chosen to fit within a certain amount of time. Participants do not spend (much) more time on the topic than the recorded time.
Time consuming for researcher and potentially for participants.	Observation time often much less than analysis time.

Table A.4 Differences in the observed process: industrial versus laboratory setting

Observed process	
Industry	**Laboratory**
The product to be designed has a history and a future within the company, which has to be taken into account. This may require additional data-collection methods.	Task is self-contained, therefore easier to analyse (all data available), although the designers bring in their own history. This may require an additional data-collection method.
Level of complexity can be high: the number of components of the design object can be very high (thousands of components). Not easy to single out a particular assembly or component on which to focus, because other parts of the design may become relevant.	Possible to focus on a low level of complexity. Usually restricted to designs with tens of components.
Data will show effects of influences from personal to macro-economic level, such as company goals, costs, availability of components, disagreements, suppliers, *etc*.	The participants determine the design. The focus of the design process can be very functional.
Observation of individuals cannot take place without considering influences of others in the project. Only part of a project is captured, as not all the work of all participants can be captured.	Analysis of individual is possible.
Problem definition and requirements will change, due to duration, relation with the market, *etc*.	Essentially a frozen assignment unless a change is deliberately introduced by the researcher.
Only parts of the process can be covered.	Whole process can be covered.
Observed process can be chaotic and complex due to interrupts and the fact that designers may need to attend to issues that are not related to the design project.	A 'smooth' process, determined by the participants.
Results show reality.	Results may not relate to reality.
Difficult to determine correlations, causes and effects.	Setup can be chosen such that correlations and causalities can be determined.

This section focuses on the main characteristics of a variety of common data-collection methods, their application in design research, our experiences and that of our students with these methods, and some references to the literature. None of the methods is generally more suitable than the others: it all depends on the research questions and hypotheses that were formulated, the context in which the method is to be applied, and how the method is tailored to and applied within this context.

Many books and articles have been written on each of these methods. In particular the books in the Applied Social Research Methods Series of Sage Publishers contain good introductions to many of the methods and provide useful pointers to more detailed literature. Reading this literature is essential to ensure the effective and efficient use of the chosen method and to avoid bias and unexpected problems, when tailoring and applying the method.

A.4.1 Observation

Observational methods involve the researcher recording what is actually taking place either by hand or using recording or measuring equipment. Observational methods are real-time methods. Observation, whether in a laboratory or practical setting, is one of the most common ways of data collection. An experiment is a classical observational method (see Section A.4.3).

The quality of observational data is highly dependent on the skill, training and competency of the observer. In the words of Patton "The trained observer is skilled in identifying and accurately describing meaningful human interactions and processes. In addition to training and practice, the fieldworker[22] needs concentration, patience, alertness, sensitivity and physical stamina." (Patton 1987). He also gives a useful account of the training requirements. Careful preparation is essential as chance favours the prepared mind.

Patton discussed six dimensions[23] of observational studies, which we discuss in the context of design research. Although specific for fieldwork, these dimensions are useful for the planning of the majority of observational studies.

- Role of observer: from full participant, to partial observation, to onlooker observation. The role may vary and evolve over time. In design research, researchers took a variety of roles, although full participation has been limited.
- Insider (emic) versus outsider (etic) perspective: the categories used to classify the data are those used by the participants, those created by the researchers, or a combination of both. In design research both perspectives have been taken, with a preference for the outsider perspective.
- Degree of collaboration of the participants: from individual or teams of researchers, to partial or periodic involvement of participants, to full collaboration and participation.[24] Design research has involved all types of collaboration.

---

[22] Patton focuses on studies in a natural setting (fieldwork), not in a laboratory context

[23] In earlier publications (such as Patton (1987) five dimensions were used, some of which have now been split or merged. The third dimension is newly included. The changes clearly show a shift in the role of the participants in field research, from passive to active.

[24] Note that 'participatory research' refers to the role of the participant, where as 'participant observation' refers to the role of the researcher (see first bullet point). 'Action research' (see Section A4.9) combines both: collaboration exists throughout the research project, with varying levels of involvement from each side depending on the phase of the project.

- Overt versus covert observations: this involves two issues. First, do participants know about the observation? Options are: all involved know that observations are being made and who the observer is (overt observation); the observer role is known by some, not by others; those involved do not know they are being observed (covert observation). Second, in the case of overt or partial overt observations, do the participants know the *purpose* of the evaluation? Options are: full explanation of real purpose to everyone, partial explanations, no explanations, or false explanations. The tendency in design research is overt observations but without providing a full explanation of the purpose, so as not to influence the design process to be studied. Explaining, *e.g.*, that the purpose is to study the use of decisions making in the design process, might lead to more frequent, explicit decision making than would normally have occurred.
- Duration: single observation with limited duration to long-term, multiple observations. Long-term or repeated observations can be useful in evaluations of design support or design teaching (such as Bender (2004)) to capture learning and motivational effects, but are not often applied.
- Focus of the observation: narrow focus, observing a single element, to broad focus seeking for a holistic view of the entire process and all its elements. Again, design research, often being explorative, has covered the whole range.

Pure Observation

Typical for pure observational methods is that the researcher is not involved in the process and does not interfere with the process while the process is ongoing (although interference caused by the observer's presence cannot be ruled out). In order to observe particular phenomena, the researcher might or might not have created the context in which the process takes place. This approach is considered the most objective.

Participant Observation

The term participant observation is used in fieldwork when the role of the researcher is not restricted to that of an onlooker: the researcher participates in the process. Participant observation can help gain acceptance and increase familiarity with the field and the problems. As an insider the researcher might also be able to collect more in-depth data and is in a better position to interpret these, such as Hales (1987) discussed in the example in Section A.2.21.

Some authors, such as Denzin (1978) view participant observation as a *research strategy* that simultaneously combines several data-collection methods, such as document analysis, interviewing, direct participation, observation and introspection. Other authors, such as Yin (1994) view participant observation as a *method of data collection* based on a special mode of observation – namely one in which the observer participates in the observed process. This data can be used as one source of evidence in, *e.g.*, a case study. In our view, participation is a role following specific playing rules that better allows collecting certain types of data in a natural setting and often involves a variety of data-collection methods.

Participant observation requires the researcher to have “the commitment to adopt the perspective of those studied by sharing in their day-to-day experiences” (Denzin 1978). The observer not only shares the subject’s world, but also takes on their language, rules and behaviour, and takes actively part in what is happening. Patton gives a useful account of the training requirements (Patton 1987).

An important issue in participant observation is that ‘experiencing what the observed experience’ is considered essential for obtaining insight, but at the same time, brings with it the danger that researchers lose their research perspective and see the world too much from the point of view of those with whom they are identifying themselves. For some schools of thought, this ‘going native’ is not acceptable: they consider the collected data invalid because of bias through subjectivity. However, a general trend towards a more involved perspective can be observed. For others, such as ethnomethodologists, this is the only way to gain true understanding. In our view, the roles of observer and participant can be combined, but this involves an awareness of the dilemma between objectivity and subjectivity, and of observing a field and being part of it. The challenge, according to Patton (2002) is “to combine participation and observation so as to become capable of understanding the setting as an insider, while describing it to and for outsiders”. The different types of notes described later in this section can be used to separate data accordingly.

Combining the roles of researcher and participant also gives rise to a very practical problem: increased participation reduces observation possibilities, observation time and interaction, and *vice versa*. Participating in a design process requires focusing on a specific design problem, which means working individually for part of the time. Hence, the time to observe the other ongoing processes reduces. A further practical problem is that researchers might not have the necessary (design) qualifications to be a complete participant. In many studies, therefore, researchers have a supporting role only.

In design research, participants are normally aware that they are being observed: to enter and work unnoticed in a professional field is highly unlikely, if at all desired. An interesting alternative is the study described by Eckert (1997). Those observed knew she was working on her thesis, but not the details. Because the researcher’s professional qualification inhibited participant observation, the researcher ‘disguised’ herself as a mixture of a placement student and a “child visiting an aunt”, *i.e.*, watching an expert doing her daily activities and chatting to her. She experienced that she got the best answers when she told the designers that she had difficulties with a particular task and got them to tell her how to do it.

Obviously, in participant observation, the background of the researcher, the role of the researcher and the possible research questions are closely linked.

Observing Participant

In participant observation, the researcher is designing, *i.e.*, temporarily taking on the role of designer. In some studies, designers observe and document their own process. We call this ‘observing participant’: the designer is observing, *i.e.*, temporarily taking on the role of researcher. Researchers might have asked designers to do so, or researchers – with the necessary qualifications – take on the role of main designer in their own design process and observe themselves. An example of the latter is a study in which researchers with design qualifications observed themselves designing a product in their academic environment (Waldron and Waldron 1988). Various methods can be used by the observing participants to capture the data. As a real-time method, data collection takes place continuously, or very regularly, to avoid the problems inherent to retrospective methods such as questionnaires and interviews. If the designer is not the researcher it may be difficult to obtain the designers' commitment if a long period of time is involved.

The terms 'observer as participant' and 'participant as observer' used by Denzin (1978) look similar to our terms, but have a different meaning. The 'observer as participant' role refers to a situation in which typically only one visit or interview is included in which no relationship builds up. The 'participant as observer' does form relationships with the other participants. In our terminology, both variants of Denzin belong to participant observation, although we do not consider the first variant, a one-visit participation, likely in the context of design.

Non-occurrences

Apart from recording what has been observed, it can be useful to make explicit what has *not* been observed, if prior knowledge suggests that certain things ought to have occurred or when in a particular case – in contrast to other cases – something did not occur. "Making informed judgements about the significance of non-occurrences can be among the most important contributions" but "this clearly calls for judgement, common sense, and experience" (Patton 2002). Such non-occurrences indicate, *e.g.*, that the finding cannot be generalised and will raise the question as to what did occur instead and why. Making explicit statements about non-occurrences in one or mores cases can emphasise the occurrences of this in the other cases and confirm correlations. An example is the observation that experienced designers, prior to undertaking a particular action, considered whether it was worthwhile pursuing this action (Ahmed 2001). This was reinforced by the observation that this behaviour did not occur in the processes of the novice designers. Here, the statement about the non-occurrence was simply the opposite of the statement of the occurrence. Note that this is not always the case (see also the discussion at the end of Section 2.4.1).

Recording

Technological developments and general availability of affordable recording equipment such as video camera's, have made it much easier to capture large amounts of rich data, but analysing these recording is still time consuming and not very easy, despite progress made in software packages for quantitative as well as qualitative data analysis.

Despite the advances in technology, taking notes has not lost its importance. Many thoughts and context details occurring during observation cannot be recalled later and might not have been captured by the recording equipment used. It is essential to make notes and go through these as soon as possible after the observation to add where necessary to obtain notes that will remain meaningful.

While observing, it is important to distinguish different types of notes to avoid confusion as to whether the note refers to, *e.g.*, what someone has said, what has been observed, or an interpretation of something that was observed. ten Have (1977) suggested 4 types of notes: observational notes, methodological notes, reflective notes and theoretical notes. In Blessing (1994) two more types of notes were added: interview notes and organisational notes.

- *Observational notes*. These notes describe pure observation of events without interpretation. The data can be biased because the underlying mental processes cannot be observed and because data depend on the view of the observer only. Observarional notes can be supported by recording equipment, taking away some of the bias. It is important to be as detailed and concrete as possible and avoid general terms describing actions and conditions such as 'the sketches were very detailed'. The level of detail has to be specified, *e.g.*, 'contained comments', 'used drawing conventions and views', 'showed many of the shape details', *etc*. Instead of writing 'She was not happy with the design' one should write 'She shook her head, did more calculations, erased parts of the drawing and corrected several times while regularly saying "oh dear"'. A distinction has to be made between what was said and what was observed, *e.g.*, using a particular coding scheme, such as using quotation marks for what was said, or using different colour pens. The source, date and context have to be recorded too.
- *Interview notes*. These notes do not contain an interpretation either (although some interpretation cannot be avoided as part of the process of interviewing (Ackroyd and Hughes 1981)). Interview notes are more reliable than observational notes because they do not rely on observation or what was observed to be said (see also Section A.4.8 on interviews). The source of the data has to be recorded as well as date and context.
- *Methodological notes:* Methodological notes are descriptions of the research approach, *i.e.*, the planned data-collecting process, adjustments that were made in the data-collection method before or during data collection, the way of analysing the data, and the experiences with the methods used.
- *Reflective notes*. These notes contain the reactions and feelings of the researcher, reactions of the participants and others, and thoughts about the role of observer and researcher. These notes enrich the observations because they can make the observer more conscious of his or her own behaviour and of the changes this causes in the field. These reflections may result in changes to the research approach, which are then described in the methodological notes.
- *Theoretical notes:* These notes contain thoughts about the collected data, such as interpretations, comparisons and characterisations that come up during the observations. These notes require verification, once all data have been collected. They can lead to further research questions and hypotheses.
- *Organisational notes:* These are notes about the organisation, the role of the participants in the organisation, aspects of their background that emerge

during the observations, and events that do not directly involve the observed process, but could have an effect.

The observational and interview notes have to be strictly separated from the other notes to avoid bias in the analysis of the resulting data.

Before planning to use any recording equipment in an industrial context it is essential to ensure that this is allowed. A confidentiality agreement might not cover this type of recording. In addition, each of the participants needs to agree to the use of recording equipment and should be allowed to request the recording to be stopped for a certain period of time, if necessary.

A.4.2 Simultaneous Verbalisation

Simultaneous verbalisation refers to the situation in which participants speak aloud while working. Participants may have been asked to do so, or this may be a natural part of their work. The aim is to provide insight into the cognitive behaviour of participants, which may not be obtained through normal observation. Simultaneous verbalisation has often been used to analyse problem solving behaviour based on the theory of information processing, which makes fundamental assumptions about the nature of cognition. The most important characteristic of simultaneous verbalisation is the real-time aspect: while working on the problem the problem solver is thinking aloud. Under controlled conditions it appeared that individuals (when trained to concurrently verbalise their thoughts) could reveal a remarkably accurate picture of their cognitive processes while engaged in problem solving (Eckersley 1988).

When participants are asked to verbalise their thoughts while working, this is called *think aloud*. The term *talk aloud* is also used, in particular if participants are just asked to speak while working, as if no one is observing. When subjects are asked to *reflect* on their thoughts, this is called *introspection*. This is more intrusive than thinking aloud and not often used in design research (an exception is Visser (1990). The least intrusive method is capturing the utterances of two or more participants who work together on a problem; less of the individual thought processes might be captured, but more of the reflections when participants explain to each other their thoughts. Note that the definitions of these terms differ depending on the sources used. Some authors consider thinking aloud the same as introspection and different from talking aloud. Patton (2002) views the think-aloud protocol approach as a specific kind of qualitative interviewing, that is, a retrospective rather than a real-time approach. The interviewing should take place "as close to the action as possible" to illuminate "what's going on in a person's head during the performance of a task". He refers to concurrent use (as common in design research) as an interesting exception. Other authors, such as we, consider thinking aloud a real-time approach, distinct from introspection, but similar to talking aloud.

Whatever terminology is chosen, important is the careful consideration of the wording used to ask the participants to speak while working, as this will affect the resulting data and hence the suitability of the data to address the research questions and hypotheses.

Simultaneous verbalisation sessions are usually a few hours and not longer than a day, not only because of the effort required by the participants, but in particular because of the effort required for analysis.

Audio tapes were found to be of limited use as the sole source for a detailed analysis of a process such as design, involving drawings and gestures. Notes taken during observation may prove insufficient to provide the missing details. This is illustrated by the following bite of an audio recording of a designer who pointed at his drawing and explained to his colleague how he imagined that the component should be cast: “and this is sand, sand, sand, sand, sand” (source: own research). Video recordings were found to be far more useful for this type of design research: the context is captured, and the data can be analysed by others and long after the recordings were made.

Protocols are the transcribed recordings of the utterances and activities of the participants. These protocols are the basis for the analysis. The collected data is very rich, but consists of fragments of sentences and seemingly inconsistent lines of thought as shown in Figure A.1. The most extensive book on protocol analysis is that of Ericsson and Simon (1996), but various other books describe qualitative and some quantitative methods for analysing and deriving meaning from such data.

design time	text	researcher	act	focus	write	trans focus 1	trans write 1	crit
00:56:14	I am sure you can do this		m	s3		a1wg		fun
00:56:21	we are getting 15 degrees movement in there		m	s3		a1wg		fun
00:56:33	really it can be quite crude		s	s3	s3	a1wg	a1wg	
00:56:37	I mean, how primitive the machine technology is		s	s3	s3	a1wg	a1wg	man
00:56:41	provided the sockets is got to wear a bit		s	s3	s3	a1wg	a1wg	env
00:56:57	if you wanted to be really sort of terribly that makes so many clamps dosn't it		m	nn		nn		com
00:56:59	because we have all these clamps off here		m	s2		pwg		
00:57:02	clamping to that		m	s2		pwg		
00:57:14	could this thing actually move at all		m	s2		pwg		
00:57:23	if somebody tried to lift this out,		m	s2		pwg		
00:57:28	there still would quite some, some of the strain would be on that joint there		m	s2		pwg		mech
00:57:36	I think that is the way to do it		m	s2		pwg		
00:57:41	how does it actually look like		o	o		o		

Figure A.1 Part of a think-aloud protocol of a designer designing a small mechanical device and categorisations of his utterances and activities used to address the research questions (Blessing 1994)

The effort involved in protocol analysis should not be underestimated. Our colleagues and we have found that detailed transcription of the video recordings of design sessions into protocols for analysis, such as the first two columns in Figure A.1, can easily take 8 hours per hour of video. A detailed analysis of the protocol using several classification schemes, resulting in codes such as those in the last six columns in Figure A.1, can take up as much as 40 hours per hour of video.

The advantage of a detailed transcription, including some information about the context (such as the material that was used or paper on which the participant wrote), is that further analysis can rely largely on the protocol, without the need to consult the original recordings to understand what was going on. Moreover, detailed transcription facilitates the use of the transcripts by other researchers.

When few aspects have to be analysed and the data can be divided into large chunks, it may be more efficient to make a summary protocol, describing rather than transcribing each segment, or not to transcribe but watch the recording to analyse each aspect. Which level of transcription and analysis is required, depends on the research questions and hypotheses that need addressing.

It is worthwhile to look for the possibility to use software packages specifically developed to support the analysis of video recordings, which could remove the need for detailed transcription and reduce the effort required for analysing the data and representing the findings. The website of the American Evaluation Association[25] lists several such software packages. Selection has to be done carefully, based on the analyses required. The currently available packages focus on particular types of analyses typical for certain domains. We found that several packages have limited possibilities for using multiple classification schemes, and hence do not allow analyses such as that shown in Figure A.1.

Using simultaneous verbalisation in a natural, industrial setting and on a real task can be problematic. We found that designers considered it difficult and even embarrassing to think aloud while designing in the design office amongst their colleagues. Furthermore, companies may not allow such detailed data to be captured.

Specific problems in transcribing and analysing the recordings of *teamwork* we have encountered are: more words (data) per time unit compared to recordings of individuals; overlapping data 'streams' because people interrupt each other and talk at the same time (a specific notation in the transcription is necessary); parallel processes when one or more team members become engaged in another issue than the rest of the team members; and team members 'doing their own thing' in silence.

A.4.3 Experiments, Quasi-experiments and Non-experiments

In the context of design the term experiment is often used incorrectly. Our field rarely provides the possibility to do true experiments. This section aims at clarifying the terminology.

Experiments

Classical experimental research is comparative research in which:

- the researcher has control over the context in which the phenomena to be studied occur;
- the participants or objects are randomly assigned to the groups involved;
- the participants or objects are representative of the target population.

Note that *randomly assigned* is not the same as *randomly selected*. Randomly assigned refers to how the participants or objects involved in the study are divided into groups to be compared, and thus relates to the setup of the study and the internal validity (see Section 4.7.4). Randomly selected refers to the way in which

[25] http://www.eval.org/Resources/QDA.htm, accessed 13 December 2008

the participants or objects were chosen to participate in the study, and thus relate to sampling and external validity (see section 4.7.4).

Classical experiments can be repeated under controlled conditions and one or more independent variables can be manipulated to test the underlying hypothesis about the effects on the *dependent variable*. The dependent variable is the variable the researcher wishes to explain. The variable that is expected to change the dependent variable is the *independent variable* (see also Section 4.5.2).

These types of experiments are the most rigorous methods available to determine *causality*, that is, to determine the time order between concepts, covariance, and the exclusion of rival factors (see Section 4.7.3). Experiments are well known in the field of natural and engineering sciences. For a discussion about experiments in social sciences see Denzin (1978) and Cook and Campbell (1979).

In experiments use is made of at least one control and one experimental group. A group is a set of cases or objects of study, each of which is observed separately. A case in design can be as diverse as an artefact, a designer, a design team, a company, or a drawing. The cases in the *experimental group* are exposed to the independent variable, they are 'treated' or 'trained' (or treatment/training is withheld, depending on what the 'normal' situation is). In design, the treatment could be the training of a design method or the introduction of a new tool. The cases in the *control group* are not exposed to the independent variable, that is, they represent the normal situation. The groups are equivalent, or at least assumed to be, because the cases are assigned to the groups at random.

Each group is observed at least using the same measurement method(s): once before (the pre-test) and once after the experimental group has been treated (the post-test). The aim of the pre-test is to detect any differences between the groups before the exposure to the independent variables. The aim of the post-test is to detect the effects. Statistics are used to determine the significance of any differences in findings. This depends on the number of cases in each group, for which reason the sample size is an important factor. This type of research is represented schematically in Figure A.2a. The 'O' stands for an observational test as a measurement method; the 'X' for a treatment or exposure to the independent variable.

Denzin (1978) suggests that if, *and only if*, pre-tests are not possible, the pre-tests may be left out, because the fact that the cases are assigned at random will often suffice as a control for the pre-test. An advantage is that it removes any problems with changes that might occur between the two tests. This is illustrated in Figure A.2b. The "purest of the experimental models" (Denzin 1978) is the Solomon Four-Group Design in which the groups differ as to whether they are exposed to the independent variable and undergo a pre-test. This is illustrated in Figure A.2c.

Quasi-experiments

In many cases it is very difficult to fulfil all requirements for experimental research (control over setting, random assignment, and representative of target population). In *quasi-experiments*, the researcher has less control over the experiment than in classical experiments, but still enough to allow the logic of the experiment to apply.[26] Quasi-experiments are experiments that have treatments, pre-tests, post-tests, and cases, but do not use random assignment. Instead, the comparisons depend on *non-equivalent groups*. To be able to interpret the data, the effects of the treatment have to be separated from those due to the initial differences between the groups (Cook and Campbell 1979).

Experimental group	O_1	X	O_2
Control group	O_1		O_2

a. Classic experimental research plan:
equivalent groups, identical pre- and post-test

Experimental group		X	O_2
Control group			O_2

b. Experimental research plan:
equivalent groups, pre-testing not possible

Experimental group 1	O_1	X	O_2
Control group 1	O_1		O_2
Experimental group 2		X	O_2
Control group 2			O_2

c. Experimental research plan:
Solomon-four-group design

Figure A.2 Experimental research plans based on random assignment of cases to groups. 'O' denotes observational tests; 'X' denotes exposure to the independent variable (or treatment)

Cook and Campbell discuss a large number of different quasi-experiments, based on varying the various elements in Figure A.2 to take into account any deviations from the conditions necessary for a classical experiment. Some of the research plans do permit reasonable causal inference, others do not. Those that do, include research plans that use different methods for pre- or post-testing, involve several treatments, or are observed several times, see Figure A.3. In quasi-experimental designs, statistical techniques substitute for the experimental method of control. In other words, because the data is not 'as hard' as in a classical experiment, statistics is used to 'filter out' these uncertainties. For example, differences between groups might have to be larger in order to be significant. See Frankfort-Nachmias and Nachmias (1996) for a discussion.

[26] The term 'quasi-experiment' is not generally accepted. King *et al.*, e.g., argue that the researcher either has control over the observations and values of the key causal variables (experimental research) or not (non-experimental research) (King *et al.* 1994).

Experimental group	O_1	X	O_2
Control group	O_1		O_2

a. Typical quasi-experimental research plan:
non-equivalent groups, identical pre-test O_1 and post-tests O_2

Experimental group	O_{A1}	X	O_{B2}
Control group	O_{A1}		O_{B2}

b. quasi-experimental research plan:
non-equivalent groups, different pre-test O_A and post-test O_B

Experimental group	O_1	O_2	X	O_3
Control group	O_1	O_2		O_3

c. quasi-experimental research plan:
non-equivalent groups, repeated pre-test

Figure A.3 Examples of quasi-experimental research plans assuming non-equivalent groups but permitting reasonable causal inferences to be drawn: 'O' denotes observational tests; 'X' denotes exposure to the independent variable

Non-experiments

Those research plans that are "*normally* not sufficient for permitting strong tests of causal hypotheses because they fail to rule out a number of plausible alternative interpretations", are useful for suggesting new ideas (Cook and Campbell 1979). These so-called *non-experimental designs* (Denzin 1978) consist of: pre-test and post-test, but only a single, experimental, group; two (non-equivalent) groups but only a post-test; and – the weakest of all – only one group and one post-test. These research plans are illustrated in Figure A.4.

An example of non-experimental research of type *c* in Figure A.4 is the analysis (post-test) of companies (experimental group) who have introduced a particular design support (treatment or exposure to the independent variable). Valuable information can be obtained from observing the use of the method and the outcome. However, nothing can be said from such an observation about the differences in process and outcome compared to the situation before the introduction of the support (although often such conclusions are drawn). Interviewing participants or using questionnaires can provide some indications about these differences, in particular if the various companies and participants involved express the same opinion. The non-experimental design type *a* requires a pre-test (such as the observation of the particular task and outcome before the introduction of the method) and would allow more statements. However, clear statements about causal inference are not possible because rival or alternative explanations cannot be ruled out. Examples of such rival explanations are an increase of experience, differences between the design tasks (the design tasks have to differ, if the same designers are involved).

Experimental group	O_1	X	O_2

a. Non-experimental research plan: no control group

Experimental group	X	O_2
Control group		O_2

b. Non-experimental research plan: Non-equivalent groups, only post-test

Experimental group	X	O_3

c. Non-experimental research plan: No control group, only post-test

Figure A.4 Three non-experimental research plans: 'O' denotes observational tests; 'X' denotes exposure to the independent variable

Design research usually involves non-experimental and quasi-experimental research, in particular for DS-I, although some product-focused research may allow experimental research plans. For empirical studies to evaluate design support (DS-II) quasi-experimental or even experimental studies are more common.

We are not certain whether design research can ever be truly experimental other than by approximation; either we conduct experiments in a contrived setting or resort to quasi- or non-experiments in a natural setting. The decision to go for an experimental or non-experimental approach involves a trade-off between realism and precision (Rossi *et al.* 1999). Currently, in particular because of the lack of understanding in design, much research will be non-experimental, but not less worthwhile or necessarily easier; "both experimental and non-experimental research have their advantages and drawbacks; one is not better in all research situations than the other" (King *et al.* 1994) and "paradoxically, the 'softer' a research strategy, the harder it is to do" (Yin 1994).

A.4.4 Case Study

The term case-study is often used to describe a study that involves data from a real setting (in our case often a setting in practice), and is seen as equivalent to an observational study in which only one or very few cases are involved. For obvious reasons, a one-shot-case study cannot be used for testing causal relationships, but can provide very valuable information. Such case studies are primarily used for exploratory research or for pre-testing some research hypotheses. Note that a one-shot case-study is not the same as the non-experimental research plan *c* in Figure A.4. The latter involves multiple cases. The earlier mentioned participant observation studies are examples of case studies. A useful introductory book about case studies is Yin (1994).

A.4.5 Collecting Documents

Retrieving documents related to a particular project, topic or product, from a variety of sources can be very useful as an additional data-collection method. Often, a study in industry will start with the collection and analysis of documents to understand the organisation, the background of the project, and the experience of the designers. Documents will be created by designers at all times. Examples are sketches, drawings, notes, calculations, minutes of meetings, emails, *etc*., and will therefore be part of the data in most observational studies.

Documents can also be the main source of data. Examples in design research are the collection of maintenance and service data to determine the reliability of various products and the effects of improvements (Stephenson 1995) and the collection of documents that were used and produced in order to study the information flow in design projects (Vroom 2001). It is often useful, but for historical data not always possible, to support document collection with other methods such as interviewing. This helps to overcome one of the main limitations of documents, namely the usual lack of data about the context in which they were created, and the rationale behind the contents.

Special methods to analyse textual data are available under the keywords of 'document analysis' or 'content analysis'. For the analysis of the other types of data produced during design, such as sketches, no guidelines are available. We can only refer to the publications of design researchers who worked with such data.

A.4.6 Collecting Products

Physical products and any mock-ups, prototypes, and other physical models can be part of the collected data, *e.g.*, to trace the development of a product. The products could be variants, members of a product family, versions that were developed over time, modules, *etc*. The focus can be on a particular aspect, or on the product as a whole. An example is the evaluation of the suitability of current product configuration methods for products with large numbers of variants (Hami-Nobari 2007). For traditional engineering research, focusing for example on the analysis of product behaviour, products are the main source of data.

A.4.7 Questionnaires

Questionnaires are used to collect thoughts, beliefs, opinions, reasons, *etc*., from people about past, present or future facts and events, by asking questions. A particular focus is on data that cannot be captured using observation or simultaneous verbalisation, and on data about the past that was not captured. "They allow us to enter into the other person's perspective" (Patton 2002).

Disadvantages of these methods are the time required from the participants and the potential bias of the results due to the fact that people are forgetful, see things from their present point of view, or give an answer that is coloured by what they conceive as more desired with respect to the purpose of the interview, social standards (*e.g.*, political correctness), or their own behaviour and that of others. A design process, *e.g.*, may be represented as more systematically and reflective than it really was.

Questionnaires may seem easier than real-time methods and to ensure data from a larger number of cases. However, the effects of poorly formulated questions on the resulting data, and the effort required for a proper analysis should not be underestimated, nor should the return rate be overestimated. Pilot studies are always necessary, which should include the actual analysis of the data.

In their book on mail and Internet surveys, Dillman *et al.* (2008) provide a very useful set of guidelines on how to formulate good questions and construct open and closed questions. To conduct a survey, questionnaires are used to address a large number of subjects. They are often sent by mail and increasingly by Internet, but also used in telephone surveys and surveys in public places.

Questions should be unambiguous, interesting, and quick to answer, because there is little incentive for people to spend their time and effort on answering a questionnaire. This is reflected in the notoriously low return rate (often less than 5%) unless specific measures are taken. Companies sometimes receive several survey questionnaires per week from research students, often on topics in which they are not interested. If at all, they will only answer those that are of interest; a low return rate thus also implies that the returned questionnaires often are not representative. Questions should not be biased, that is, should not suggest an answer. A questionnaire should also be self-explanatory, as the researcher is not at hand, and should be answerable by one person. We have regularly seen questionnaires combining questions about different areas of expertise, *e.g.*, on company strategy, methods to generate ideas, methods to assess market needs. Few persons, if any in a company, can answer all these questions. No company will put the effort in involving multiple people in order to answer the survey.

Apart from questions about the topic of interest, the following additional questions are necessary.

- Questions about relevant characteristics (*e.g.*, function or background) of the person who answered the questions, as this can influence the answers.
- Questions related to the characteristics that were used to select the sample, in order to verify whether those that returned the questionnaires are representative of the target group. These can be the characteristics of the person, project, company, *etc.*
- Questions about relevant characteristics of the context (product, project, situation, or other) to which the answers refer, in order to be able to correlate answers of different questions within one questionnaire. Determining correlations between answers makes sense, only if the answers refer or apply to the same context. A simple but effective solution is to ask the participants to answer the questions for, *e.g.*, a particular project, and then to repeat the questions for another project. Specifying a particular feature (the last, the current, a particularly successful or problematic project) can help the participants in their choice, result in more focused answers (rather than 'it depends') and provide useful additional information for analysis. To know how typical a particular answer is, additional questions are necessary.

As a supporting method, questionnaires are often used to obtain information about the participants or organisations involved and their opinion about the study. Questions can be asked prior to the study (*e.g.*, to obtain data about background and expectations of the participants) and afterwards (*e.g.*, to obtain their opinion about the effect of the research environment on their way of working). The return rate is usually near 100%, because the questionnaire is part of a larger study. This also implies that the questionnaires can generally be more extensive and can contain more open-ended questions then survey questionnaires. Nevertheless, to obtain valid, comparative and useful data, the questions should be formulated and the questionnaire constructed in an equally thorough way as survey questionnaires. Although the researcher may be available, help should not be necessary.

A.4.8 Interviewing

The purpose of interviewing is similar to that of questionnaires (see previous section) but are done carried out face to face. They are used to collect thoughts, beliefs, opinions, *etc*., about past, present or future facts and events, with a focus on data that cannot be observed or was not captured in the past. The interview should "provide a framework within which respondents can express their own understandings in their own terms. One of the greatest obstacles to overcome is unlearning the bad habits practiced and reinforced daily in our ordinary conversations, such as lack of depth, miscommunication, lack of clarity in our questions, interrupting the answers given, and lack of direction in the dialogue" (Patton 1987). Not only the questions asked, but also the interviewer has a large influence on the process and outcome of the interview. Patton (2002) provides useful guidelines for formulating interview questions and conducting in-depth interviews. He distinquishes the following types of interview:

- The *informal conversational interview*, in which questions are generated spontaneously. This is often part of participant observations.
- The *interview guide*, which is a list of questions or issues generated prior to the interview. These issues are to be explored during the interview and do not prescribe the precise questions. The list allows the same topics to be covered across interviews, and ensures coverage.
- The *standardised open-ended interview*, which consists of carefully worded questions (a questionnaire) that are asked to all interviewees in the order given in the list. This allows easier comparison, and can be very useful when multiple interviewers are employed. When time is limited, it might also ensure that all topics are covered. The flexibility to pursue topics that emerge unexpectedly is limited.
- The *closed quantitative interview* (multiple-choice questionnaire) in which questions and answers are determined in advance. The interviewee only chooses an answer. Although this eases analysis, it does not necessarily capture the experiences and opinions of the interviewees.

The latter three are often referred to as *unstructured*, *semi-structured* and *structured*, respectively. The informal conversational interview is usually not considered a real interview. Structured interviews are easier to analyse and compare. Unstructured interviews are more suitable for an exploratory study. Interviews can be used as the main data-collection method and in conjunction with real-time studies, as discussed in the previous section on questionnaires. Compared to the anonymous questionnaires, interviews are potentially more confrontational and might require more effort.

Focus groups are group interviews that focus on a specific topic. The group dynamics can enhance the overall outcome of the interview, but may have a negative effect on the contribution of some participants, depending on the person, the topic, and the differences in status of the participants.

The questions and the structure of the interview should follow the same careful preparation as questionnaires, and the same additional questions are necessary to obtain information about the context (see previous section).

Tape recording interviews supports analysis because it captures what has actually been said and the intonation used, but also supports the interviewer who can concentrate on the interviewee and the direction of the interview, rather than on taking notes. Note taking will remain important to record potentially relevant observations and – in semi- or unstructured interviews – to have an overview of the direction of the interview, which can act as a reminder of aspects that need to be addressed or clarified. The latter will avoid the need to interrupt the interviewee ('Before I forget …').

Interviewing can reveal many of the reasons for behaviour that cannot be observed and only partially be derived from simultaneous verbalisation, team discussions and surveys. Disadvantages are similar to those of questionnaires, but with interviews, the ability to verbalise one's thoughts also plays an important role. According to Mintzberg (in Bessant (1979)) there is no evidence to suggest that people can effectively translate complex reality into meaningful abstraction. The results rely on the verbalising capacities of the observed, and the researcher's ability to interpret what has been said. A detailed discussion of this multiple translation problem in interviewing can be found in Ackroyd and Hughes (1981).

There are six types of questions that can be asked to people, in particular in interviews (Patton 2002):

- experience/behaviour;
- opinion/belief;
- feeling;
- knowledge;
- sensory (what you see, *etc*., used to find out the stimuli the interviewee is subject to);
- background and demographic.

Regarding the order of questions, it is useful to start with some non-controversial, easy questions. One should avoid asking a long list of background questions right at the beginning as these are considered particularly boring. Their number should be minimised (what do we really need to know) and distributed throughout the interview. It is also important to keep interviewees focused on relevant issues, as they may drift off into topics they would like to discuss, *e.g.*, because these are particularly important for the interviewee at the time of the interview. A related problem is that interviewees may tell what they think the interviewer will want to hear, rather than their own opinion or experience. This is a particular problem when support is evaluated that has been developed by the researcher. A focus on the research questions, a good ear, and practice are needed to overcome these potential disadvantages.

Methods and techniques for the analysis of spoken data can be found under the keywords ‘conversation analysis’ or ‘discourse analysis’.

A.4.9 Action Research

Action Research is an approach to introducing and evaluating change, originally in organisations and programmes, but increasingly in design (*e.g.*, Björk (2003)). Action research has the dual aims of action and research. Through cycles of action and research a better understanding is obtained, while at the same time the organisation or programme under investigation is gradually changed. Action research is usually qualitative, data driven, participatory, and makes use of multiple data sources. “Action researchers help transform inquiry into praxis, or action. Research subjects become coparticipants and stakeholders in the process of inquiry. Research becomes praxis – practical, reflective, pragmatic action – directed to solving problems in the world. … Together, stakeholders and action researcher co-create knowledge that is pragmatically useful and is grounded in local knowledge” (Denzin and Lincoln 2005).

This approach was developed as a reaction to the failure of social sciences to produce results that were useful in solving society’s problems. The close relationship with practice has its effect in the practical relevance of the work and its similarity with consultancy work. The combined responsibility for actual change as well as for research is demanding and the demands on responsiveness and flexibility are high. Furthermore, action research often emphasises local relevance (that is, responsiveness) at the cost of global relevance (that is, generalisation) (Dick 1997).[27] Some specific developments of Action Research are the Critical Action Research Approach of Carr and Kemmis (1981), the Evaluation Methodology of Guba and Lincoln (1989), and in particular the Soft Systems Methodology (SSM) of Checkland (1999).

A.5 Statistical Analysis

The aim of this section is to introduce the terminology needed to select the most suitable methods for analysing quantitative or quantified qualitative data; in particular when this data is nominal or ordinal, and when there are no assumptions about the distribution of the population out of which this data has been taken. A useful book is Vogt (1999) about statistical concepts and methodological terms in

---

[27] Dick (1997) gives a very practical overview of how to do action research, summarizes the main methodologies that can be applied and provides an extensive list of the literature with short descriptions.

social and behavioural science providing non-technical definitions of these terms and concepts with examples.

Statistics can be divided into *descriptive statistics* that enable the researcher to describe and analyse data without drawing conclusions or inferences about a larger group, and *inferential or inductive statistics* that enable decisions or inferences to be made about a larger group by interpreting data patterns. Because such inferences cannot be absolutely certain, the language of probability is used in stating conclusions. All statistics can be used for interval and ratio scales, but far fewer can be used for nominal and ordinal scales (see Section 4.7.2).

The analysis methods can be divided into univariate, bivariate and multivariate depending on the number of variables taken into account. Which specific method within these three groups are most suitable, depends on the scale of the variable and on the distribution of the data.

Typically, *univariate analysis* methods look at only one variable and are descriptive: frequency distributions, measures of central tendency such as mean and median, basic measures of tendency, such as variance and standard deviation, and type of distribution.

Bivariate analysis methods measure the relationship between two variables. The first step is usually to construct a bivariate table, placing the categories of the variables along the two axes.

Multivariate analysis methods measure relationships between multiple variables. Two types of methods can be distinguished; methods to *verify* relationships between dependent and independent variables, and methods to *discover* relationships between variables when the variables have not been divided into dependent and independent. This reflects the difference between an hypothesis-driven and a data-driven approach. Table A.5 lists the basic multivariate methods to verify relationships. A premise is that the data is numerically coded.

Table A.5 Basic multivariate methods to verify relationships between dependent and independent variables (Erichson *et al.* 2000)

Dependent variable	**Independent variable**	**Method**
interval or ratio	interval or ratio	Regression analysis
interval or ratio	nominal	Variance analysis
nominal	interval or ratio	Discrimination analysis
nominal	nominal	Contingency analysis

Contingency analysis is one of the more frequently used set of methods when verifying hypotheses in design research, since much of the qualitative data can only be coded on a nominal scale. Of these, the χ^2 (Chi-square) method is one of the most widely known. These methods assume that it is possible to observe the variables involved. When use is made of hypothetical constructs that cannot be observed, such as motivation, LISREL (Linear Structural Relationships) is a suitable method. Conjoint Measurement is often used when the dependent variable is measured on the ordinal scale.

The following methods are used to *discover* dependencies:

- *Factor analysis*, to reduce a large set of variables into fewer 'core' variables. These variables are often constructs; it may not be possible to measure these directly. Examples of such core variables are quality, experience or motivation.
- *Cluster analysis* does not group variables, but cases (units of analysis). All cases in a cluster are more similar to each other than to those in other clusters.
- *Multidimensional scaling* is similar to factor analysis but is based on a similarity or dissimilarity assessment between cases using, for example, pairwise comparison, rather than on evaluating each property of a case. It allows one, *e.g.*, to determine how a participant sees a unit of analysis.

The statistical tests included in software such as spreadsheets usually incorporate assumptions about the underlying distribution of data, that is, the parameters of the population from which the sample is drawn (such as normal distribution) and require that the variables are measured at least at an interval scale. *Non-parametric statistics* are so called because they make few or no assumptions about the distributions, and do not require interval scales. There is at least one non-parametric equivalent for each parametric test, as shown in Table A.6.

Among the most often used is the χ^2 (Chi-square) method based on contingency or cross-tables and its variants. Also commonly used is the Spearman's rank correlation coefficient, as the alternative to the standard Pearson product-moment correlation coefficient for which the variables have to be metric. In choosing a method, not only the type of comparison is relevant, but also the type of data and, most importantly, the minimum required number of cases.

A few words about the *sample size* required for a particular study. The relevance of statistics and hence of sample size, is related to the approach taken, *i.e.*, the paradigm chosen. The most suitable sample size depends on many factors, but foremost on the research questions and hypotheses to be addressed. For that reason, a study with a sample size of '1' can be as valuable as a study with a sample size of 1500, but only if the research questions, the research methods, and the conclusions are in line with the sample size.

In particular in quantitative approaches, the calculation of the sample size is essential to be able to draw conclusions. The calculation is mainly based on the expected variability of the data and on the difference or precision one considers relevant. If the testing of an hypothesis requires a certain difference to be measured, *e.g.*, a difference in quality of the products analysed, or an improvement caused by a new support, the study should be designed such that a difference can be detected and that – given a certain level of confidence – this difference is not coincidental and is relevant. The so-called *power* is the probability that the study will successfully detect a difference. When the variability of the data is not known and cannot be estimated, it may be necessary to run a pilot study. In more exploratory studies, the sample size should ensure that the estimates from the study have adequate precision. Many statistical methods demand a minimum sample size (either overall or per category). In various books and websites, methods to calculate sample size or power can be found.

Table A.6 Non-parametric alternatives to parametric statistical tests (Burke 1998)

Types of comparison	**Parametric methods**	**Non-parametric methods**
Differences between independent groups of data	t-test for independent groups	Wald–Wolfowitz runs test Mann–Whitney U test Kolmogorov–Smirnov two-sample test
	(ANOVA/MANOVA)	KruskalvWallis analysis of ranks Median test
Difference between dependent groups of data	t-test for dependent groups	Sign test Wilcoxon's matched pairs test McNemar's test χ^2 (Chi-square) test
	ANOVA with replication	Friedman's two-way ANOVA Cochran Q test
Relationships between continuous variables	Linear regression Correlation coefficient	Spearman R Kendall's Tau B
Homogeneity of variance	Bartlett's test	Levene's test, Brown and Forsythe
Relationships between counted variables		Coefficient Gamma χ^2 (Chi-square) test Phi coefficient Fisher exact probability test Kendall coefficient of concordance

Appendix B

Prescriptive Study Methods

A wide range of methodologies and methods exist to support the development of products. A similarly wide range exists for the development of software. Design support can be seen as a product that primarily deals with information and is increasingly in the form of software. Therefore, both the above methodologies are potentially useful for support development. This appendix outlines some of these methodologies and methods, and provides pointers to more detailed sources. Some are general approaches that can be beneficial throughout the support development process. Others are task specific, or particularly useful for the development of computational design support.

The first two sections of this appendix (Sections B.1 and B.2) focus on product development and software development, respectively. In each section, first an overall development methodology is outlined and then a list of methods for supporting each stage of the methodology is given, with references to further literature. Section B.3 discusses methods specifically for user-interface design. Section B.4 presents a checklist to aid the documentation of the scope and assumptions of the support throughout the development process.

B.1 Product Development Methodologies

Product development methodologies propose that designing should be done in a series of stages, progressively detailing the product under development. An example is the approach of Pahl and Beitz (2007). At each stage a series of steps is proposed to lead the designer from problem understanding to solution. An example is the series of steps suggested by Roozenburg and Eekels (1995): analysis, synthesis, simulation, and evaluation and selection. The following sections provide a list of methods for each of these steps. Some methods are more suitable for the earlier stages of development, while others are for more detailed stages. The listed methods have been selected on the basis of their relevance for design support development. Unless otherwise specified, the methods are taken from the following books: Pahl and Beitz (2007) (abbreviated as PB); Roozenburg and Eekels (1995) (abbreviated as RE); Jones (1970) (abbreviated as J); Cross (1994) (abbreviated as C).

B.1.1 Methods for Analysing Objectives and Establishing Requirements

The following (Table B.1) is a list of methods that help analyse objectives, clarify the requirements that the support should fulfil, the relationships between the requirements, and their relative importance.

Table B.1 Methods for task clarification (PB = Pahl and Beitz (2007), RE = Roozenburg and Eekels (1995), J = Jones (1970), C = Cross (1994))

Method	Aim
Stating objectives [PB, RE, J, C]	To identify the external conditions with which the support must be compatible
Literature search [J]	To find published information that can be useful
Interviewing users [J]	To elicit information known only to the users of the intended or existing support
Questionnaires [J]	To collect information from the members of a large population
Investigating user behaviour [J]	To explore the behaviour patterns and to predict the performance limits of potential users
Interaction matrix [J]	To permit a systematic search for relationships between elements within a problem
Interaction net [J]	To display the pattern of relationships between elements within a problem
Classification of design information [J]	To split a design problem into manageable parts (this should help solve the problem as well as help modularise the support)
Objectives Tree [RE, C]	To clarify objectives and sub-objectives of the support, their relationships and their weightings
Function Analysis [C, PB, RE]	To establish the functions required and the system boundary of the support to be designed
Performance specification [C]	To make an accurate specification of the performance required of a support
Quality Function Deployment (QFD) [RE, C]	To set targets for the characteristics of the support so that they satisfy user requirements
Specification writing [J]	To describe an acceptable outcome of the planned development process (can be useful in writing the future situation expected of a support)
Design specification procedure [RE]	To specify the requirements by listing, analysing and editing objectives

B.1.2 Methods for Synthesising Support Proposals

The list of methods in Table B.2 could be used at various stages of synthesis in support development. The use of some of these may be straightforward in the context of generating design support; others may require adaptation. The methods could help generate a variety of alternative proposals for fulfilling the individual requirements of the support, and help combine these into overall proposals.

Table B.2 Some methods for synthesising proposals (PB = Pahl and Beitz (2007), RE = Roozenburg and Eekels (1995), J = Jones (1970), C = Cross (1994))

Method	**Aim**
Brainstorming [J, RE] , Brainwriting [RE], Checklists [RE]	To stimulate a group of people to produce many ideas quickly
Synectics [J], Random stimulus [RE, C], Intermediate impossible [RE], Concept challenge [RE]	To direct the spontaneous activity of the brain and the nervous system towards the exploration and transformation of development problems
Removing mental blocks (Adams 1993)	To find new directions of search when the space searched has yielded no acceptable solution
Transformation [C] Counter-planning [C], Why?Why?Why? [C]	To enlarge search space
Function-Means Tree (Hubka and Eder 1988)	To develop functions and means to fulfil the functions together.
Morphological charts [J, C, PB, RE]	To widen the area of search for solutions
Value engineering [C]	To increase or maintain the value of a support to its user whilst reducing its cost to its developer (Can be particularly useful when modifying an existing method).
Functional innovation [J]	To find a radically new type of support capable of creating new patterns of behaviour and demand
System transformation [J]	To find ways of transforming an unsatisfactory support so as to remove its inherent faults
Alexander's method of determining components [J]	To find the right components of a structure such that each component can be altered to suit future changes in the environment

According to Pahl and Beitz (2007), synthesis methods can be classified into *intuitive* and *discursive* methods. Cross (1994) calls these *creative* and *systematic* methods respectively.

Roozenburg and Eekels (1995) divide the creative methods into *association* methods and *creative confrontation* methods. Association methods, such as brainstorming, encourage spontaneous reactions to ideas expressed earlier. Creative confrontation methods are – like association methods – characterised by connecting initially unconnected ideas, but such connection is now 'forced' by a particular step in the method. An example is Synectics.

Systematic methods are based on the systematic analysis and description of a problem, the drawing up of a variety of solutions to sub-problems, and systematic variation and combination of these sub-solutions into solution variants.

In Table B.2, the first four rows list some well-known creative methods, while the rest are a sample of available systematic methods.

B.1.3 Methods for Simulating Support Behaviour

The following list of methods (Table B.3) support simulation of designs at various levels. Roozenburg and Eekels (1995) distinguish four kinds of models: structure models (*e.g.*, flow diagrams), iconic models (*e.g.*, dummies, scale models or prototypes), analogue models, and mathematical models.

Table B.3 Some methods for aiding simulation of proposal behaviour (PB = Pahl and Beitz (2007), RE = Roozenburg and Eekels (1995), C = Cross (1994))

Method	**Aim**
Analysis of interconnected decision areas (AIDA) [RE]	To identify and evaluate all compatible sets of sub-solutions
Failure modes and effects analysis (FMEA) [RE, PB], fault-tree analysis [RE]	Methods for analysing reliability of a new support by finding possible causes and effects of failure early in the development process
Simulation of product form [RE]	To simulate the look and feel of the support to provide insight into its semantic and aesthetic functions. Can be particularly useful for user-interface design, the look of a workbook, the layout of a checklist, *etc*.
Business and economic simulation [RE]	To analyse the costs and benefits of a new support
Value analysis [RE, C, PB]	To analyse functions and sub-functions of a support and compare their value to their costs
Ergonomic simulation [RE]	To use a support model and a human model together to simulate ergonomic aspects of using the support

Rules for development and interpretation of *structure models* are often intuitive. Such models may be useful in determining the ways a proposed support should work, especially at the early stages of its development. Structured design methodologies in software development have many such methods, see Section B.2. In *iconic models* the properties of the design are represented in the model using the same properties. In support development, a scaled-down version may be developed (say for developing a constraint propagation algorithm for a hundred constraints rather than the required thousands of constraints). This can be used for testing the feasibility of the idea before developing full-scale prototype software. *Analogue models* use a different property to represent a given property of the original. For instance, a method for finding an optimum may be seen as analogous to finding the best path to move among a tree of paths, which in turn can be seen as *hill climbing*. The hill-climbing algorithm can then be used as an analogue model of the original task, which, once solved, could give an answer analogous to the original. *Mathematical models* are symbol models that can be analysed using rules from mathematics. For instance, the exhaustiveness of combinations produced by an algorithm could be calculated mathematically.

B.1.4 Methods for Evaluating and Selecting Support Proposals

The following is a list of methods (Table B.4) commonly used for evaluation and selection of proposals at various stages of the development process, in particular for identifying the right criteria for evaluation and assigning appropriate relative importance, as these are often non-trivial tasks.

Table B.4 Some methods for aiding evaluation of proposals (PB = Pahl and Beitz (2007), RE = Roozenburg and Eekels (1995), J = Jones (1970))

Method	Aim
Checklists [J, PB]	To enable support developers to use knowledge of types of requirements generally found relevant in similar situations
Estimating weighting factors [PB, RE]	To indirectly assess the importance of criteria by externalising the decision-maker's preference for hypothetical alternatives
Identifying and selecting criteria [PB, J]	To decide on how to recognise acceptable support proposals
Ranking and weighting [PB, RE, J]	To compare a set of alternative support proposals using a common scale of measurement
Estimating evaluation uncertainties [PB]	To estimate the reliability of evaluation results
Searching for weak spots [PB]	To identify the weak areas of a support variant
Ordinal methods (majority rule, Copeland rule, rank-sum rule, lexicographical rule, Pugh Charts) [RE]	To rank alternatives per criterion on an ordinal scale and compare the alternatives against a list of criteria and their importance
Cardinal methods [RE]	To rank alternatives by quantifying judgement of the effectiveness of the alternatives and the importance of the criteria on an interval scale

B.2 Software Development Approaches

Software development approaches have much in common with design approaches. They too have some main stages such as specification, development and validation. However, software development has its own idiosyncrasies over and above those involved in product development, that are relevant for the development of computational design support. Section B.2.1 discusses CaeDRe, a methodology specifically developed for supporting computational design support development that uses DRM as one of its underlying bases. The two generic software development paradigms, functional and object-oriented are outlined in Section B.2.2, and some generic methodologies in Section B.2.3. One of the methodologies, the waterfall model is discussed in more detail in Section B.2.4. Realisation of design support may require use of generic technologies drawn from areas such as artificial intelligence; some of these are discussed in Section B.2.5. The last section (Section B.2.6) focuses on computer-aided software engineering (CASE) tools that are available for some of the methodologies discussed previously. Unless otherwise stated, the information provided here is taken from the classic book of Sommerville (2006).

B.2.1 A Design Support Software Development Methodology

This section describes CaeD (computer-aided engineering design research methodology) (Bracewell *et al.* 2001) and its associated computational environment CaeDRe (computer-aided engineering design research environment) (Bracewell and Shea 2001). Figure 5.1.4 in the main part of this book illustrates the methodology. CaeD is a "practical methodology for computer aided engineering design research" that is intended to enable the development of useable computational design support early on in research projects. Its application is "intended to enable rapid, robust implementation of research design methods suitable for empirical evaluation in academic and industrial settings". The methodology uses DRM as its underlying basis, and makes extensive use of software and social science technologies.

CaeDRe is developed in response to the difficulties in practical software implementation that they see as a major reason why the observed integration and evaluation of computational design support tools resulting from fundamental design research (Culley 1999) is such a problem. Some of the causes behind the difficulties of practical software implementation, according to Bracewell and Shea (2001), are "ignorance of, or failure to apply, fundamental software engineering principles", while others are "specific to the particular nature of computational tool design research". The methodology is a variant of the evolutionary or prototyping methodology (Section B.2.3).

The research process supported by the CaeD methodology follows the four stages of DRM. In the first stage, called 'Criteria'[28], Measurable Success Criteria for the tool are identified linking back to overall business objectives. In the second

[28] Based on the name we earlier used for the Research Clarification stage.

stage, Description I, the existing design process is analysed to discover relations between Measurable Success Criteria and the actual design process, thus identifying where application of a design support could lead to improvements in this process. In the third stage, Prescription, insights gained in Description I are used to create a storyboard for an improved design process that could result from using the new design support. For computational design support, this storyboard creates a starting point for specifying and implementing a prototype system. Finally, in Description II, the design support is tested to determine whether it works as intended and whether it actually impacts the Measurable Success Criteria.[29]

The authors divide the support development process into five activities:

- task definition;
- choice of representations;
- choice of methods;
- definition of visualisation, interaction and distribution strategies;
- theoretical and experimental validation.

CeaDRe uses the product platform concept as "a set of sub-systems and interfaces that form a common structure, from which a stream of derivative products can be efficiently developed and produced". Using the product platform concept emphasises that individual software solutions should have interface definitions that allow them to form a flexible but coherent architecture.

CaeDRe "provides an open, flexible environment for the development of computational design support, allowing progressive code-hardening from scripts to robust efficient compiled software for third party applications, using a choice of development tools (high level languages, tools and integration platforms) that are usable by researchers and programmers with a wide range of software expertise without creating barriers to system integration". Its architecture is a modular system of client, server and extension packages, and allows an iterative, rapid prototyping approach to the solidification and testing of exploratory research ideas.

B.2.2 Software Development Paradigms

There are two major paradigms in software development: function-oriented and object-oriented.

The *function-oriented paradigm* relies on decomposing the system under development into a set of interacting functions with a centralised system state shared by the functions. Functions may also maintain local state information but only for the duration of their execution. A functional approach is most suitable in systems where the amount of system state information is minimised and information sharing is explicit. Systems whose responses depend on a single stimulus or input and that are not affected by input histories are naturally function-oriented. Functional approaches were practised informally since the early days of

---

[29] The framework was based on the earlier terminology used in DRM. What was earlier called the 'Criteria' stage is now called the 'Research Clarification' stage. Similarly, the other stages have been renamed.

programming, but were only developed into a formal paradigm in the seventies (also called 'structured' approaches). A number of analysis, design and programming methods are available within this paradigm that help identify, develop and implement the functions that are necessary. The approach of Yourdon (1989) is a typical example.

The approaches based on the *object-oriented paradigm* differ from the functional approaches in that they view a software system as a set of interacting objects with their own private state, rather than as a set of functions that share a global state. Objects are abstractions of real-world entities that are responsible for managing their own private state and offering services to other objects. They are independent entities that may be readily changed because state and representation information is held within the objects. System functionality is expressed in terms of operations or services associated with the objects. According to Sommerville (2006) object-oriented systems are easier to maintain and change as objects are independent, and there is a clear mapping between real-world entities and their controlling objects in the system. This improves the understandability and hence maintainability of the system. A number of analysis, design and programming methods are available that help identify and develop objects and their interactions (Booch *et al.* 1994; Coad and Yourdon 1990; Jacobsen *et al.* 1993).

Since the late 1980s, object-oriented approaches have become increasingly popular, also with some of the protagonists of the function-oriented approaches, such as Yourdon. He mentions the following three difficulties from which the function-oriented (structured) approaches suffer (Yourdon 1990):

- Function-oriented approaches place an enormous emphasis on the modelling of functions, and little on that of data. Although entity–relationship diagrams have been introduced to alleviate this problem, many project teams ignore this altogether while modelling user requirements. Object-oriented approaches deliberately package both data and functions together into a single container – the object.
- The diagramming notation in function-oriented approaches provides little mechanism for using reusable components. Through the inheritance mechanism, object-oriented methodologies promote reuse of attributes and methods (functions) of existing objects.
- Due to its history of development in an era when graphical user interfaces were unknown, these methodologies offer little to support user interface development.

However, according to Sommerville (2006), function-oriented and object-oriented approaches should not be treated as competing approaches, but chosen or even combined according to their suitability to the application at hand.

B.2.3 Generic Software Development Methodologies

There are various generic methodologies for software development. The *waterfall model* is the most commonly used. It has several stages (Royce 1970), each of which produces its own distinct type of deliverables and associated documentation (see also the next section). The model is a cascade from one stage to another.

The stages in the waterfall model are:

- Requirements establishment: The system's services, goals, and constraints are established and defined in a manner that is understandable for both users and developers.
- System and software design: the requirements are allocated to either hardware or software systems, and an overall system architecture established. Software design involves representing the system functions into a form that may be transformed into executable programs.
- Implementation and unit testing: the software design is realised in terms of a set of program units. Unit testing is used to verify that each unit meets its specification.
- Integration and system testing: the individual program units are integrated and tested as a complete system to ensure that the software requirements are met. After this, the software is delivered to the customer.
- Operation and maintenance: the system is installed and put into practical use. Maintenance involves correcting errors that were not discovered during development.

A strong advantage of the waterfall model is that it provides a very clear-cut method for managerial control. However, practice has shown that it has a number of disadvantages (Schreiber *et al.* 2000) .

- It is sometimes difficult to see progress in the early stages since these are mainly document-oriented; visible and operational results in terms of software can only be tried much later in the software development process.
- Prefixed stages make changes during the project difficult and costly. Changes can arrive from changed external circumstances, new insights gained during the project, or changing needs and requirements.

An alternative methodology, the *V model*, relates each development stage not only to its immediate predecessor and successor, as in the waterfall model, but also to the related testing stage at the same level of detail (Gram and Cockton 1996). The left leg of the *V* represents the development stages – problem analysis, requirments specification, system design, software design, module design – the right leg of the *V* the testing stages – module testing, integration testing, system testing, acceptance testing, and use and maintenance. The coding stage joins the legs. This shows that acceptance tests have to be created as part of the specification, and used during the final installation. Similarly, software design is at the same level as integration testing, and so on. The V-model, however, is still a variant of the waterfall model in that it does not easily allow backtracking to a phase once development has advanced beyond it (Schreiber *et al.* 2000).

The *evolutionary or prototyping approach* (Smith 1991) can be seen as the opposite of the waterfall model: it aims to produce practical results quickly in a number of iterative improvements based on learning from the previous cycle. This approach is therefore highly adaptable and experimental. However, this is also its weakness: due to the lack of structure it is not really possible to generate sound project goals and plans in advance. In other words, it is too flexible.

A model that attempts to combine the good features of both the waterfall and prototyping approaches is *Boehm's model* (Boehm 1988). His model is also known as the spiral approach, combining the linear waterfall approach and the cyclic prototyping approach. The aim is to achieve progress by means of subsequent cycles that may be adapted on the basis of experience gained in earlier cycles. Depending on the situation, one may decide to follow analysis and design as in the waterfall model, or prototyping activities if these are more illuminating or useful.

A variant of the waterfall model, but carried out using formal mathematical methods, is called *Formal System Development Model*. In this model, the goal is to produce a formal mathematical system specification first, and then transform this, using mathematical methods, to construct a program. Verification of the system components is carried out by making mathematical arguments about how these component-functions conform to the specification.

Another methodology gaining increasing acceptance is called *Reuse-based Development*. It is based on the assumption that a significant number of reusable software components exist, and the goal of the system development processes is to integrate these components into a system rather than developing them from scratch. While reuse of available software components is informally encouraged in all software development methodologies, it is formally used in this methodology. The typical steps are: requirement specification, component analysis, requirement modification (to reflect reuse of available components), system design with reuse, systems development and integration, and system validation. Note that the stages are similar to those in the waterfall model, but emphasise maximum reuse of existing software components.

Apart from the methodologies discussed in this section, development methodologies are also available for specific types of software systems, such as knowledge-based systems (Buchanan and Duda 1983) (see Section B.2.5).

B.2.4 The Waterfall Model

The waterfall model is discussed separately in this section, as it provides a highly detailed approach with a logical breakdown of its stages into steps of increasing detail. These steps can be useful to follow, as long as they are followed flexibly, as aimed in Boehm's spiral model. Unless otherwise indicated, the description is taken from Sommerville (2006).

Requirements Establishment

In software engineering two levels of requirements are considered:

- A requirements *definition* is a statement, in natural language and diagrams, of the expected services of the system and the constraints – including potential users – within which it is expected to operate.
- A requirements *specification*, also called a functional specification, is a structured document that sets out the system services in detail and should be precise.

The requirements definition and specification are developed in four steps:

- Feasibility study: An estimate is made whether, or the extent to which, the identified user needs may be satisfied by using the current technology.
- Requirements analysis: The system requirements are derived, *e.g.*, through observation of existing systems and discussions with potential users.
- Requirements definition: The information gathered during the analysis activity is translated into a document that defines a set of requirements, reflecting what the customer wants.
- Requirements specification: A detailed and precise set of requirements is formulated to act as a basis for a contract between client and system developer.

Table B.5 provides a list of methods, some generally applicable, others with a function-oriented or object-oriented flavour, that aid requirements establishment.

Table B.5 Methods for establishing software requirements (S = Sommerville (2006))

Method	**Aim**
Checklist of critical characteristics of software (S)	To provide a list of requirement characteristics to choose from when developing software
Viewpoint-oriented analysis (S)	To analyse requirements taking into account the viewpoints of all parties involved
Method-based analysis (VORD) (S)	To analyse requirements using a structured method to understand the system
Semantic data models (S)	To analyse requirements by identifying the logical form of the data processed by the system
Object-oriented models (S)	To analyse requirements by representing both data and its processing
Data-flow models (S)	To analyse requirements by identifying data flows through a sequence of processing steps
Standard format approach for requirements definition (S)	To define requirements based on a standard format of the requirement and rationale
Requirements specification approaches (Davis 1990)	A number of approaches to add structure to the specification to reduce its ambiguity
Software prototyping techniques (N.N. 1989; Smith 1991)	A number of approaches for rapid prototyping of software in order to understand its requirements
Checklist for requirements validation (S)	To provide a checklist of characteristics that a requirements specification should have

Software and System Design

Software and system design is the stage that leads to the transformation of informal ideas into detailed implementation descriptions. Note that this is not yet the actual implementation into software. The outcome is comparable to technical drawings in product development.

A general model to represent a software design is a directed graph. The target of the development process is the creation of such a graph without inconsistencies, where the nodes represent entities such as processes or functions, and links represent relationships between these such as calls or uses. A software design is iteratively developed through a number of different versions, increasing its formality and detail in each version, making it more consistent and complete. Frequent backtracking is needed to correct earlier, less formal and less detailed designs.

The software and system design stage has six, iteratively related, steps:

- Architectural design: The sub-systems making up the system and their relationships are identified and documented.
- Abstract specification: For each sub-system, an abstract software specification is produced of the services it provides and the operation constraints.
- Interface design: The interface of each sub-system with other sub-systems is designed and documented.
- Component design: Services are allocated to different components and the interfaces of these components are designed.
- Data structure design: The data structures used in the system implementation are designed in detail and specified.
- Algorithm design: The algorithms used to provide services are designed in detail and specified.

The last five steps are repeated for each sub-system until the components identified can be mapped directly into programming language components such as packages, procedures or functions. As in the requirements establishment phase, it is possible to use both function-oriented and object-oriented paradigms in this stage.

The design activities in a function-oriented development process are to:

- model the system using data-flow diagrams showing how data passes through the system and is transformed by each system function;
- model how functions are decomposed into sub-functions using graphical structure charts;
- describe the entities in the design and their interfaces;
- describe the control structure of the design using a program description language that includes conditional statements and looping constructs.

The design activities in the object-oriented development process are to:

- identify the objects in the system along with their attributes and operations;
- organise these objects into an aggregation hierarchy that shows how objects are part of other objects;
- construct dynamic object-use diagrams that show which object services are used by other objects;
- specify object interfaces.

Implementation, Integration and Testing

Software implementation involves the actual realisation of the support into software units, that are integrated into modules, sub-systems, and the overall system. Each unit, module, sub-system and the overall support need to be tested to ensure that they work as intended. This involves the following steps.

- Implementation of units: implementing individual program units and ensuring that they work as intended.
- Implementation of modules: integrating individual units into modules, and ensuring that the modules work as intended and no unanticipated interactions between the units occur.
- Implementation of sub-systems: integrating modules into subsystems, and ensuring that the sub-systems work as intended and no unanticipated interactions between the modules occur.
- System implementation: integrating sub-systems into the overall system, and ensuring that it works as intended by the requirements and no unanticipated interactions between the sub-systems occur.

As these steps show, implementation and testing must go hand in hand. The commonly used cycle is called *debugging*, which has the following steps.

- Implement or modify the program: a provisional version of the intended program is implemented.
- Test the program: the program is run and output collected.
- Evaluate the output: the output is analysed for its correctness, possible errors detected, possible causes hypothesised, and possible remedies proposed.
- Go back to the first step: the remedial proposals generated in the previous step are used for modifying the program in step 1. The cycle is continued until the output of the program is satisfactory.

Software should be tested at each stage. There are two types of testing: verification and validation. *Verification* involves checking that the program conforms to its specification. *Validation* involves checking that the program as implemented meets customer requirements. Boehm (1979) summarises these differences as:

- 'Verification: Are we building the product right?'
- 'Validation: Are we building the right product?'

Verification and validation can be done using static and dynamic techniques. *Static techniques* are concerned with the analysis and checking of system representations such as the requirements document, design diagrams and the program source code. Static techniques can only check correspondence between various levels of specification (verification); they cannot demonstrate that the program is operationally useful (validation). *Dynamic techniques* involve exercising an implementation, and can be applied both for verification and validation as long as an executable program is available. Two types of dynamic testing can be distinguished.

- *Statistical testing* may be used to test the program's performance and reliability.
- *Defect testing* should find areas where a program does not conform to its specification. This testing is most common during program implementation, where each component developed should be checked against its intended functionality, and modified if it does not provide this functionality. Test-debug-test is the usual program implementation cycle.

Except for small programs, systems should not be tested as a single, monolithic program, but stepwise from unit, to module, to sub-system, to the overall system, as discussed earlier. Once the overall system is tested, *acceptance testing* (sometimes called *alpha testing*) can take place. This involves testing of the system with data supplied by the user rather than simulated test data. This may reveal errors and omissions in the requirements definitions or reveal that the system does not really meet the user's need because it introduced additional influences not anticipated at the development stages. A further level of testing is *beta testing*, which involves delivering a system, prior to its marketing, to a number of potential customers who agree to use the system. The results may lead to further modification and beta testing, or to marketing. A number of available testing strategies are summarised in Sommerville (2006).

Operation and Maintenance

Operation and maintenance issues have to be taken into account during support development and testing (see DS-II, Chapter 6), but these are normally not a stage in the development of support as part of academic research.

B.2.5 Generic Technologies

Awareness of generic technologies available for performing various software tasks is necessary for efficient and effective software development. Those discussed briefly in this section are artificial intelligence (AI), expert/knowledge-based systems (KBS), knowledge engineering (KE) and computer-supported collaborative work (CSCW), because of their widespread use in developing computational design support. The technology used has ramifications to the whole research approach. For instance, choosing knowledge-based systems to realise a support might mean using knowledge-acquisition techniques during DS-I and earlier stages of PS.

Artificial Intelligence

The area of AI has two main goals (Schank 1990): to develop methods that will make computers far more intelligent and therefore more useful than they are at the moment, and to find out about the nature of intelligence. Many issues that AI deals with are relevant for developing design support and AI is a major source for potential computational solutions.

The two major issues in AI are representations and methods. AI proclaims that good representations are a key to good problem solving as they support explicit, constraint-exposing descriptions. Methods are procedures that use representations to solve specific problems. In AI, generate and test, mean-ends analysis, and problem reduction are three powerful and generic methods. A wide variety of heuristics is available to be used in conjunction with these basic methods. The methods and their combinations are used for solving a variety of problems in a wide range of applications, such as qualitative and quantitative constraint resolution, optimisation, and learning. Many books on AI have been written: for an excellent introduction, see Winston (1993). The AI in Design Webliography of Brown (2008) provides a collection of potentially useful sources of information on AI in Design, as well as on Knowledge Based Design, Intelligent CAD, Computational Approaches to Design, and Design Theory & Methodology.

Expert Sytems, Knowledge-based Systems and Knowledge Engineering

Expert systems technology is one of the spin-offs of AI that has found many real-world applications. Expert systems are able to execute a task that, if carried out by humans, would require expertise (Schreiber *et al.* 2000). Expert systems are typically rule-based systems, built using a set of *if-then* rules. Kline and Dollins (1989) provide a set of guidelines for designing rule-based systems, helping to answer questions such as:

- What knowledge-representation technique should be used?
- What problem-solving strategy should be employed?

Rule-based systems do certain tasks well, but they do not reason on multiple levels, nor do they use constraint-exposing models. They do not look at problems from multiple perspectives, do not know how and when to break their own rules, and they do not have access to the reasoning behind their rules (Winston 1993).

Expert systems evolved into knowledge-based systems, which refer to programs that require extensive use and explicit representation of knowledge. Unlike expert systems, which are often devoted to automating expert tasks, knowledge-based systems use a variety of ways for representing and processing knowledge, and are often used in supporting rather than automating tasks. In a sense, all information processing systems can be called knowledge systems, since they all use knowledge. However, according to Schreiber *et al.* (2000), the main distinction is that in a knowledge-based system one assumes that there is some explicit representation of knowledge.

Knowledge engineering is the approach taken to acquire and formalise knowledge, and use this for building knowledge-based systems. Knowledge engineers use domain experts to acquire useful knowledge. Knowledge engineering has three benefits (Schreiber *et al.* 2000).

- It helps one to spot the opportunities and bottlenecks in how organisations develop, distribute and apply their knowledge resources, and so gives tools for corporate knowledge management.
- It provides methods for obtaining a thorough understanding of the structures and processes used by knowledge workers, leading to a better integration of information technology in support of knowledge work.

- It helps, as a result, to build better knowledge systems: systems that are easier to use, have a well-structured architecture, and are simpler to maintain.

The classical model of knowledge-based system development was given by Buchanan and Duda (1983). It contains six steps.

- Identification: Together with the domain experts the knowledge engineer identifies the important aspects of the problem, including participants, problem characteristics, resources and goals.
- Conceptualisation: The key concepts and relations resulting from the previous steps are made explicit.
- Formalisation: The concepts identified are represented in a formal language.
- Implementation: Knowledge from the last step is represented in a system shell.
- Testing: Together with domain experts, the completed system is tested on sample cases and weaknesses are identified.
- Revision: The results from testing, may require redesign and re-implementation, which may involve domain experts.

There are two cases in which domain knowledge is particularly important:

- One is where domain knowledge exists with experts, but for the sake of cost or scarcity of experts, it has to be extracted, organised and even optimised, so as to develop a knowledge-based system. For example, knowledge of best practices of developing a product may have to be acquired, reordered and made available to novice designers. In order to do this, expert designers could be used as a resource to draw upon knowledge, and as a means of validating the resulting knowledge-based system.
- The other case is where the knowledge required does not exist, but in order to generate and evaluate this knowledge, it is necessary to have an understanding of the domain. In our reliability example, for instance, the goal was to create a method for estimating, at early design stages, the potential reliability of a product, but the knowledge required for such estimation was not available. In order to develop and evaluate the necessary knowledge, sufficient understanding of the domain of the technical product considered and its designs had to be developed.

In design research, there are two major reasons for aquiring domain knowledge. One is for developing the support, and the other for its evaluation. The knowledge acquired can take various forms and can have been aquired during the DS-I or PS stages.

- *Product descriptions*: These can be extracted from real products, detailed drawings of products, product descriptions, sketches, verbal or written descriptions (functional, behavioural or structural) in design catalogues or designers' documentations. In our synthesis example, for instance, mechanical designs in a number of these forms were collected and analysed

to extract building blocks for synthesis. In the reliability example, detailed drawings of the sub-assemblies having reliability problems and their failure data were collected in order to identify elements in these designs with inherent problems of reliability, and also to test the validity of the model of reliability so developed.
- *Process descriptions*: These are descriptions of processes used to develop products. In the reliablitiy example, for instance, knowledge was collected about the ways in which reliability is currently calculated in order to contrast these with the desired situation and to include these, as far as suitable, into the intended support.

Various techniques for knowledge acquisition are available, especially from the discipline of knowledge engineering. According to Grosso *et al.* (1999), knowledge engineers are required to:

- become familiar with the problem domain;
- characterise the reasoning tasks necessary to solve the problem;
- identify the major domain concepts;
- categorise the type of knowledge necessary to solve the problem;
- identify the reasoning strategies used by experts;
- define an inference structure for the resulting application;
- formalise all these in a generic and reusable way.

Depending on the type of support developed, several of these steps might be useful to follow.

A number of generic knowledge engineering methodologies exist, the most notable of which is the Common KADS methodology (Schreiber *et al.* 2000). This methodology provides a suite of knowledge models and modelling capabilities so as to answer three types of questions:

- Why? These questions help develop the context of the knowledge-based system to be developed: Why is the system a potential help or solution? For which problems? Which benefits, costs, and organisational impacts does it have? These questions are answered by developing models of the organisation, tasks and agents involved.
- What? These questions help develop the conceptual description of the knowledge applied in the tasks to be supported, such as: What is the nature and structure of the knowledge and communication involved? The questions are answered by developing models of the knowledge and communication involved, based on the information available in the models resulting from the 'Why' questions.
- How? These questions help focus on the technical aspects of realising the system: How must the knowledge be implemented in a computer system? How do the software architecture and computational mechanisms look? These questions are answered by developing a design model that provides a technical system specification in terms of architecture, implementation platform, software modules, representational constructs and computational

methods needed to implement the functions laid down in the knowledge and communication models resulting from the 'What' questions.

Computer-supported Collaborative Work

Computer-supported collaborative work (CSCW) or groupware may be defined as "hardware, software and processes designed to aid in group related tasks such as basic communication, information sharing, decision making, scheduling/control, and analysis/design" (Saunders 2008). There may be several individuals or groups involved in a collaboration, located in the same or different location and time, and involved in different kinds of processes or activities. Groupware is a term that encompasses many technologies and business process areas. Specific technologies include electronic mail, digital voice mail systems, text conferencing, video teleconferencing, collaborative databases, workflow, group decision support systems, and living worlds.

Three major points should be considered as the individual groupware technologies are discussed. First, the greatest power from groupware is exhibited when the various technologies can be combined with each other, and be integrated within the business processes of the organisation. Second, success in implementing groupware requires a critical application and a critical mass. Groupware requires group work – the work of all the groups, within an application in which they all participate. And third, the major challenges in the groupware discipline are not technical or economic, but social. This section is based primarily on Saunders (2008). Several books have been written on this topic. An overview of resources can be found in Foraker Design (2005).

B.2.6 CASE Tools

CASE (computer-aided software engineering) systems offer computational support for various aspects of the software development process. They can be classified into (Fuggetta 1993).

- Tools: These support individual tasks such as checking the consistency of a design or compiling a program. They may be stand-alone or grouped into workbenches.
- Workbenches: These support process phases or activities such as specification or design, and normally consist of a set of tools with some degree of integration.
- Environments: These support a substantial part of the software development process, and normally include several workbenches that are integrated in some way, often based on a specific software development methodology.

The most widely used types of workbench are the following (Fuggetta 1993).

- Analysis and design workbenches: These support the creation of models of the system (such as a data flow diagram) in the analysis and design stages of the software development process. These are sometimes called upper-CASE tools. The tools vary from very method specific to general.

- Programming workbenches: These support program development. The main components are assemblers and compilers that translate high-level programming languages to machine code. Other tools are editors, debuggers and printers. These are often referred to as lower-CASE tools.
- Testing workbenches: Testing tends to be application and organisation specific. Consequently, there are not many ready-made testing workbenches. Those that exist tend to be open systems that evolve to suit the system being tested. Tools common in these workbenches include test managers (manages the running and reporting of tests), test-data generators (generates test data for the program to be tested) and simulators (simulates the machines on which the program is to be executed).

The downsides of these CASE supports are the relatively long learning curve and high capital investment. Several books have been written about CASE tools. A list of CASE tools can be found in Wikipedia.

B.3 User-interface Design

This section outlines essential aspects of user-interface design, or human–computer interaction (HCI) which can be a very important part of support development.

Interactive computer systems are built in order to help people achieve some goals as efficiently as possible (Gram and Cockton 1996). The user interface is often the yardstick by which a system is judged, causing at best a high level of user errors to be incurred, at worst, not using the software irrespective of its functionality. Therefore, user interfaces need to be developed sufficiently well so that hypotheses about the functionality of the system that are relevant to the research aims can be evaluated. The more interactive the software, the more important is the role of its user interface.

This section concentrates on the development process for interactive systems, the roles of the users in this process, and the available tool environments.

B.3.1 User-interface Development Issues

A user-interface needs to resolve two key issues:

- How can information given by the user be presented to the support?
- How can the information provided by the support be presented to the user?

Quality of a user interface is measured by two types of properties of the interface and the support. Gram and Cockton (1996) distinguish these two property types as follows. Although they focus on computational support, most of the properties are also relevant to other types of support.

- From the *user's perspective*: The interface should be pleasant, reliable, easily understandable, and have sufficient functionality, so that all tasks can be performed with ease. These external properties fall into five categories:
    - goal and task completeness: you can do what you thought of doing;

  - flexibility: you can do things in several ways;
  - robustness: you can avoid doing things you wish you had not done;
  - learnability: the ease with which novice users can acquire competent performance;
  - user satisfaction: how a system makes users feel in terms of sense of achievement or excitement.

- From the *software engineer's perspective*: The interface, as any other part of the system, should have a number of properties that are defined through software and hardware properties of the system. Particularly relevant are:

  - modifiability: how easy it is to modify the system when facilities have to be extended;
  - portability: how easy it is to change its hardware and software environments;
  - evaluability: how easy it is to evaluate the system against quality goals;
  - maintainability: once installed in an environment, how easy it is to maintain the system;
  - run-time efficiency: whether the system consumes an acceptably low fraction of computer resources;
  - user-interface integratability: how easy is it to integrate the interactive system with existing or new software applications;
  - functional completeness: whether the system has sufficient functionality to support users to do their tasks;
  - development efficiency: whether the most effective use of resources is being made during system development.

Some of these properties are 'soft' and can only be defined and measured by taking the user's cognition and understanding into account; others are 'hard' and can be measured by standard software engineering methods.

According to ISO 9241 (ISO 1998), the main standard for working with computers, defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Effectiveness can be defined as the degree of accuracy and completeness with which the user goals are satisfied. Efficiency is taken as the effectiveness of system usage in relation to its costs in terms of effort or time. Satisfaction relates to user's comfort or acceptance of the system. Ergonomics of human–system interaction enabled by software should satisfy the following (ISO 1998):

- software should enable solving of the specified tasks;
- software must speak the language of users;
- users should be in control of the software;
- the software should present familiar things in a familiar manner;
- users have a right to err;
- users are different from each other;
- software should qualify.

An important issue in user-interface development is to determine the level of interaction required: how should the user (designer) interact with the support? In the case of computational support development, there can be different possible types of interaction, as discussed in Section 5.5.3 and shown in Figure 5.10. The level of interaction required is important for two reasons. It clarifies and constrains the kind of support that needs to be developed, and it indicates the level of implementation necessary for evaluation of the support.

Other useful standards are ISO/IEC 9126 (ISO 2001) on software product evaluation, and in particular ISO 13407 (ISO 1999) on human-centred design processes for interactive systems.

B.3.2 Levels of Abstraction

Gram and Cockton (1996) distinguish four levels of abstraction for interchanges with an interactive system:

- Functional level: This is the highest level of abstraction. At this level the operations or abstract commands and objects provided by the system are described. It is the first level below the 'task' level that is considered outside the interactive system. For example, a functional level command maybe 'start draw program', or 'set date and time'.
- Dialog level: This level is concerned with the temporal behaviour and the interdependencies among the operations and objects. For instance, the above two functional commands are expanded further into: 'open DrawImage window', or 'select month; advance month; select date;…'.
- Logical interaction level: This level expands on how to do the interaction with reference to presentation entities rather than raw device values, and with some generalisation over low-level events. For example, the above two operations are described as: 'move mouse to DrawImage icon; click mouse' or 'move mouse to menu; move mouse to 'month' item; click mouse;…'.
- Physical interaction level: This is the lowest level of abstraction that describes what really happens during interaction. This level may be unnecessary when the underlying system automatically takes care of its details.

B.3.3 User-interface Development Processes and Methods

The software development models discussed in B.2 must be modified to take into account human–computer interaction aspects in interactive systems development. According to Gram and Cockton (1996) HCI approaches:

- model new aspects for system design by introducing task, performance and conceptual models (the latter describe systems at the functional level);
- introduce new detailed design concerns related to output formatting, interaction techniques, and the use of colour and sound as well as other media and modalities in information coding;

- add new software components especially for the dialog, such as help, history, undoing, macros, tailoring, and tutoring;
- produce new development models with different orderings of development phases, *e.g.*, designing the user interface first;
- create new forms of testing, *e.g.*, formative and summative usability testing;
- give rise to new forms of installation plans, *e.g.*, special training plans for dialog-intensive systems;
- introduce new problems of maintenance, *e.g.*, for self-adaptive systems that change the dialog by exploiting the user's emerging pattern of usage.

Schematically, following the abstraction levels of Gram and Cockton discussed in the previous section, the development process starts with a task analysis identifying the tasks to be supported. At the functional level, the task steps are conceptualised as abstract commands applied to objects. These are then refined through the remaining levels into specific sequences of renderings and communication devices (such as speech input or output, graphic displays, mice, tablets, *etc.*) at the physical interaction level.

An important issue in designing interactive systems is keeping the software components for user-interface functions separate from those of the rest of the system, which may be termed the functional core (Gram and Cockton 1996). The functional core provides the computational realisation of the problem domain functionality for an interactive system, while the user-interface components represent this functionality to end-users and support them in the use of these representations. The term UIMS (user-interface management system) was coined in an attempt to promote this concept.

According to Sommerville (2006), an exploratory development is the most effective approach to interface design. This, according to him, must initially lead to creation of paper-based mock-ups before developing into screen-based designs that simulate user interaction. A user-centred approach (Norman and Draper 1986) should be used, with end-users of the system playing an active part in the design process – as evaluators or as co-developers. Sommerville suggests the process for user-interface development shown in Figure B.1.

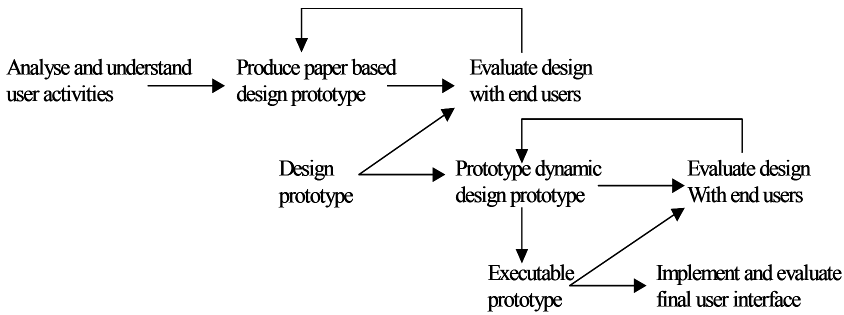


Figure B.1 Process for user-interface development (after Sommerville (2006))

This process contains three iterative, interrelated activities: analyse and understand user activities, produce prototypes and evaluate prototypes, before implementing and evaluating the final user interface. The prototypes may be at various levels of abstraction.

Various authors have listed design principles for user-interface development. Elaborate guidelines are available in Shneiderman (1998). The following is an example from Sommerville (2006).

- User familiarity: Terms and concepts used in the user interface should be drawn from the experience of potential users.
- Consistency: Terms, concepts and operations should be consistent; comparable operations should be activated in the same way.
- Mental surprise: Users should never be surprised by the behaviour of the system.
- Recoverability: Users must be allowed to recover from errors.
- User guidance: Meaningful feedback about errors and context sensitive user help must be provided.
- User diversity: The interface should enable appropriate interaction for different types of users.

Some methods for design of user interfaces are listed in Table B.6. For a more elaborate overview, see Sommerville (2006).

Table B.6 Methods for aiding design of user interfaces (Sommerville 2006)

Method	**Aim**
Direct manipulation (*e.g.*, graphical user interface)	To present users with a model of their information space and allow them to interact *via* direct actions
Desktop metaphor	To provide consistency and user familiarity by making the interface model analogous to some real-world model that users understand
Menu systems	To provide user navigation of a large information space while remaining aware of its current position
Checklist of issues relevant for information presentation	To provide information such that its purpose is fulfilled
Methods for user guidance	To provide user guidance at various levels
Checklist of issues in error message design	To provide useful error messages
Multimedia	To provide multiple media in documents
Issues in help system design	To provide help of various kinds

B.3.4 Interactive Software Development Environments

The CASE environments specifically aimed at supporting development of interactive software are called Interactive Software Development Environments (ISDE). Many attempts have been made to develop practical tools that assist with the development of user interfaces and with management of the interaction between the user interface and the functional core. Typically the support provided by ISDE varies with the hardware and operating system used, the preferred look and feel, the assumptions about the types of interaction required, and assumptions about how interactive system designers work.

B.3.5 User-interface Evaluation

Since user interface development can have far-reaching (indirect) effects on the user evaluation of the software functionality, it is important that care is taken in evaluating the user interface. As systematic evaluation of a user interface can often be an expensive process, a series of simpler, less expensive evaluations, especially during the software development phase should be used. Some of these are the use of questionnaires collecting information about user's thoughts about the interface, observation of users at work with the system, or inclusion in the software of code which collects information about the most-used facilities and most-common errors (see also Chapter 6).

Broadly, there are two classes of methods for evaluating user interfaces:

- Predictive methods that can be used very early in the design process, as soon as a specification or even a low-technology prototype is available.
- Experimental methods that can be used when a running prototype or some mock up of the system is available.

Sommerville (2006) provides list of specific methods under these two classes, see Tables B.7 and B.8.

Table B.7 Some predictive methods (Sommerville 2006)

Method	**Description**
HCI-based design heuristics	These allow inspection by specialists for certain technology aspects (principle-based inspection) or for conformance with published style guides (Style conformance inspection).
Formal methods	These aid in assessing properties, such as using a formal specification of a dialog, to prove that it has some specified properties.
Cognitive-theory-based methods	These allow inspection by specialists for learning problems (Cognitive walkthrough, see Polson *et al.* (1992), or use of a cognitive model using Goals, Operators, Methods and Selection rules for a system to evaluate learnability or efficiency of a dialog (GOMS method).

Table B.8 Some experimental methods (Sommerville 2006)

Method	Description
Participative design	In these, the user interface and the functionality of the developing system is presented to user representatives.
Summative evaluation	These involve structured and planned evaluation of the completed software by usability specialists, with measurement against required targets.
Heuristic evaluation	This involves the informal but planned examination of whether the system fulfils a pre-identified set of heuristic usability criteria (Nielsen 1992).
Usage observation	This involves semi-structured monitoring and observation of real user's interaction with the system.

The purpose of experimental methods, even if based on a prototype user interface, is to obtain objective measures of user difficulties and subjective impressions from the user of ease of use, good and bad features, ease of learning, *etc*. Such user tests need very careful planning and preparation. The inevitable weaknesses of a prototype (*e.g.*, missing functionality) may lead to user frustration if the users are not suitably instructed. Subsequent design must take into account these subjective impressions, as they can be more important than objective measurements.

In contrast to a prototype system, a system functional walkthrough need not be conducted with a computer-based system. It could just as easily be based on low-technology prototyping such as flip charts, recorders or other presentation mechanisms. Developments in participative design have greatly expanded approaches to low technology prototyping (Muller *et al.* 1993).

Whichever mechanism is chosen to derive user impressions and study the usability of the design, it is important that the process is not merely a single step in the development process. Iterations will be necessary until both software engineers and users are content with the proposed user interface.

B.4 Support Outline: Summarising Scope and Assumptions

The use of design support will always change the working situation, and the effects can be intended or unintended. This has to be taken into account from very early on in the development of design support. Several questions need to be answered: What is the scope of the support? Which tasks are supported? What are the desired effects? How does the work environment change? What is the relation to existing support?

It can be argued that most of the design support developed in academia consists of concepts or prototypes to illustrate new ways of working rather than commercially robust products, and that therefore considerations of implementation, use and maintenance do not need to be considered. Potential exploitation, however, is a criterion for success of a research project and often a criterion for funding. The earlier all these aspects are considered, the more likely it is that the ideas will find their way into practice.

Unfortunately, it often shows that, although the researcher has quite a good mental picture of the intended support, this is not made explicit (Blessing 1997).

- The information provided in proposals or reports is insufficient to understand the support, l*et al*one to assess it.
- The overall scope and aim are not expressed.
- The assumptions on which the support is based have gradually been forgotten.
- Only the positive effects of the use of the support and the core functionality are mentioned, not the potential side-effects.
- As a consequence, many support proposals are interesting but unrealistic.

The checklist shown in Table B.9 (Blessing 1997) is intended to help summarizing and illustrating the envisaged design support by identifying its scope and the underlying assumptions, as early as during the planning stage. The resulting description clarifies the problem that is addressed, the approach and the possible implications, and can thus allow the intended support and its concept to be more easily understood and assessed. For the researcher the checklist helps to reveal how realistic the envisaged support and whether its scope has to be narrowed. The checklist can be used for drawing up profiles of existing support and hence allows comparison between different supports.

It is recommended to start using the checklist right at the beginning of the PS stage and to continue updating the resulting description during the support-development process. Note that not all aspects mentioned in the checklist are equally applicable to each kind of support.

Table B.9 Checklist for identifying scope and assumptions of design support

Area of use	
Aims	What are the underlying aims and objectives of the support? What is the ultimate goal? (general, specific, ...) (scientific, social or both).
Product type or domain	What type of product or what domain is being served with the support? (general, mechanical, electrical, ..) (aerospace, rehabilitation, ...) (mass, made-to-order,..).
Process type	What type of design process should be supported? (original, redesign, variant).
Users and tasks	
Tasks or process to be supported	What are the tasks or processes the support is intended to support? The task to be supported is related to the current way of working, not the future situation. Tasks of direct users and of indirect users (those who maintain, install or use the results) should be considered.

Table B.9 *(continued)*

Functions to be fulfilled	What are the specific functions the support has to fulfil in order to support the task? (An outline of its intended behaviour in terms of input and output and the action on the input).
Number of users working in parallel	Is the support intended to be used by a single user, a number of individually working users or a group of interacting users? Include direct and indirect users, human as well as computer software.
User description	Who are the users, including both direct and indirect users. Are the users experts or novices? Are the users familiar with the task supported by the support, or has the support been introduced to provide the knowledge they are lacking or are less familiar with, such as guidelines for manufacturing, or do they only use the data resulting from the support? (users in all life-cycle phases: introduction, customisation, use, maintenance).
Interface	
User's main role, human computer interaction	What are the roles of the various users in applying the support? How and how often does interaction take place? Is the interaction continuous? Who is sender, who is receiver, who takes main initiative for the various tasks? How active is the tool? How much knowledge and support does the support provide?
Input characteristics	What type of input is required? (numerical, verbatim, graphical, symbolic), (complete or incomplete), (specification, drawing, models..). Is a template provided?
Output characteristics	What is the type of output of the support? (Numerical, verbatim, graphical, symbolic, schematic..) (complete, incomplete, ..) (exact, range, fuzzy, ..) (specification, drawing, sketch or other model of the product). What does the output represent? Is it a restructuring of the input, or does it have additional data? Does it provide intermediate answers or a trace of the process? Does the support always provide an output? Is it clear when an output is incorrect?
Implementation	
Customisation	How much is the support tailored to a particular product, process, discipline or company? What has to be customised? What is the expected effort? Who is customising the support?
Maintenance	Does data stored in the system need to be updated? What type of maintenance is involved? What is the expected effort? Who is maintaining the support?
Links with other systems or methods	Links to which other systems or methods are required? What links are possible?Which data is needed as input or output for the link to be effective?

Table B.9 *(continued)*

Effects	
Needs	What needs from practice are addressed? Why would a company or institution want to use (and purchase) the support? Why would a user want to use the support?
Problems	What are the particular problems that are expected to be solved?
Problem-solving method, approach	How is the support expected to solve the problem? What are the procedures involved? What is the rationale behind it? What are the limitations of this approach?
Expected effect on the work situation (assumptions)	What are the expected effects of the use of the support on the work situation? How does the support intend to change the situation?
New work situation	What is the new work situation?
Potential side-effects	What are the potential side-effects of the support?
Validation	What methods were used or can be used to validate the support and what test data is being or can be used?

Appendix C

Example Research Projects

In this appendix we discuss seven design research projects, which are presented by the supervisors or PhD students involved, and place these projects in the context of DRM.

C.1 Overview of the Examples

The seven design research projects presented in this appendix took place before the methodology was fully developed and are thus not examples of our methodology. The aim of this appendix is to provide a glimpse of the variety of issues tackled in design research, and the variety of research methods and techniques used in tackling these issues. Most importantly, the projects are used as examples to illustrate how different research projects can be placed within the framework of DRM. An overview is given in Table C.1; detailed discussions can be found in the Reflection sections at the end of each project description.

Together, the projects cover many important areas of research: synthesis, design for quality, design for reliability, teamwork, process support, developing metrics for design process performance, and generating methodologies. The projects range from the highly theoretical work of Mørup and Andreasen to the largely computational work by Shakeri *et al.* Some have a distinct process-oriented flavour: Blessing *et al.*, Mabogunje *et al.*, Frankenberger and Birkhofer, and Shakeri *et al.* fall in this category. Others carry a substantially product-oriented flavour: Chakrabarti and Bligh, Mørup and Andreasen, and Stephenson and Wallace fall in this category. The aims of the projects range from identifying influences for a particular issue (teamwork (Frankenberger and Birkhofer), quality (Mørup and Andreasen)), through process support (Mabogunje *et al.*, Blessing *et al.*), to support for specific tasks and activities (Shakeri *et al.*, Chakrabarti and Bligh, Stephenson and Wallace). Methods used include, amongst others, protocol analysis, questionnaires, observation, interviewing, agent-based simulation, historical case studies, noun phrase analysis, and product analyses.

The process of classifying these projects as well as others, using the DRM framework, helped us to improve the descriptions of the stages such that the variety of research projects in design can be represented.

Table C.1 Classification of the projects in this appendix in the DRM framework

Example	Research Clarification	Descriptive Study I	Prescriptive Study	Descriptive Study II
Blessing	Review	Comp	Comp (Initial)	Comp
Chakrabarti	Review	Review	Comp Comp Comp	Initial Comp
Frankenberger	Review	Comp	Comp	Comp
Mabogunje	Review	Comp	Comp Comp	Comp Comp
Mørup	Review	Comp	Initial	Initial
Shakeri	Review	Review	Comp	
Stephenson	Review	Comp	Comp	Initial

C.2 A Process-based Approach to Computer-supported Engineering Design

Student: Lucienne Blessing (author)
PhD Dissertation, University of Twente, the Netherlands, 1994 (Blessing 1994a)
Supervisors: Harry van den Kroonenberg, Koos Mars, Cees Terlouw

C.2.1 Introduction and Aim of Research

In this century, in the field of mechanical engineering, both products and the process of their creation have undergone major changes. In order to remain competitive, new approaches to product development are needed to cope with the new characteristics of, and increasing pressure on, product and process. This need was addressed in the *social aim* of the research project:

> *to improve the mechanical engineering design process, i.e., to realise a more conscious, effective and efficient process.*

The following definitions were used. A conscious process is a process that is executed by people who are aware of its structure and use this knowledge in planning and reviewing. An effective design process is a process that results in a product satisfying the actual need. An efficient design process is a process that is effective and in which the applied resources do not exceed the planned resources. Efficiency includes product- and process-related factors; it searches for the best results given the available resources.

The social aim led to an initial investigation into the various means to improve the design process, such as design methodologies and computer tools, in order to determine the scientific aim of the project. The first conclusion was that these means had not had the expected impact on the effectiveness and efficiency of the design process: existing prescriptive models are inadequately exploited; existing knowledge and methods are poorly accessible or unknown; and the range of application and focus of computer-based tools is limited. The second conclusion was that a focus on the design activity *(the design process)* rather than the more common focus on the design deliverable *(the product)*, along with a focus on computer *support* of the entire process rather than *automation* of a specific activity, would be the most promising approach to improve mechanical engineering design. A focus on the design process was expected to increase the impact of existing means; enable the process to be monitored; increase the accessibility of knowledge, methods and tools; and open the way to extend computer support to the entire process.

This led to the decision to address the need to improve design by developing a computer-based support tool based on a model of the design process, combining the advantages of computer processing, the knowledge and capabilities of human designers and the potentials of a methodical approach. The *scientific aim* of the study then became:

the development and evaluation of a model of mechanical engineering design upon which to build a computer-based support tool, *in which the design process is the core.*

The hypothesis that the scientific aim helps achieve the social aim was based on the following *assumptions*, for which the literature was found to provide some support:

- Awareness of a process is considered the first step toward improvement.
- Focusing on the design process (*i.e.*, its activities) influences its effectiveness.
- Focusing on the design process influences its efficiency.
- Computer systems can enhance the effectiveness and efficiency of the design process.

C.2.2 Research Approach

The scientific aim focused on realising a model of mechanical engineering design upon which to base a design support tool. This resulted in the following *central research question*:

What model of design can be used to develop a process-based computer support system for mechanical engineering design that improves the design process?

This research question was then divided into a number of more detailed *research questions*:

- What are the characteristics of effective and efficient design processes according to prescriptive design literature, *i.e.*, what processes are suggested by design researchers to improve design?
- What are the characteristics of the design processes as they actually take place according to descriptive design literature, *i.e.*, what is the context in which the support system should function and, in particular, what are the characteristics of successful design?
- What are the characteristics of the envisaged type of computer support for design?
- What are the requirements for this type of computer support, *i.e.*, at what combination of characteristics from theory and practice should the system aim?
- What are the elements of a model of design that can be used as the core of this system?

The approach applied to answer the research questions is illustrated in Figure C.1.

The first stage focused on the first two research questions. An extensive study of prescriptive and descriptive literature was undertaken to reveal the characteristics of mechanical design that could be of importance for the development and use of a system to support the improvement of the design process.

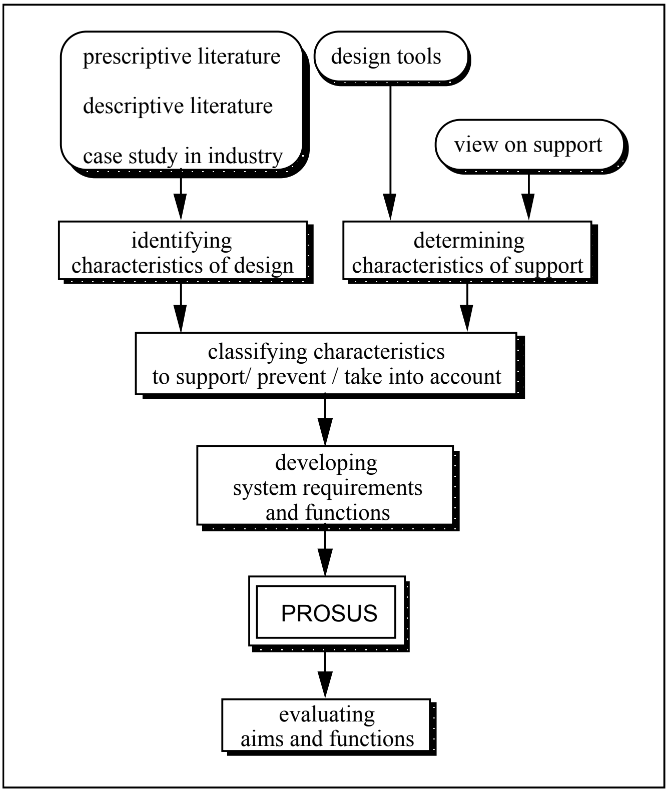


Figure C.1 The research approach

Prescriptive literature, such as Pahl and Beitz (1996); French (1971); Asimov (1962), suggests ways of improving the design process by means of models and methodologies. These offer a broader, albeit less detailed, view of the process than descriptive studies. Nearly 30 models were compared. *Descriptive literature* offers characteristics of design based on observation, revealing the environment in which the developed system would have to function. Over 70 different studies were compared with each other and with prescriptive literature. Because of a lack of detailed studies in industry, a design *process in industry* was observed for 16 months to find additional characteristics. Both sources were consulted to combine the characteristics of effective and efficient processes suggested by prescriptive literature, with the reality of design practice found in descriptive studies. In total, several hundred characteristics were identified and classified (Blessing 1994b).

At the same time, the general characteristics of computer-based support for design were specified, based on the characteristics of existing design tools (Blessing 1991) and on a personal view of the role of computers in design. A design support system was envisaged to involve the designer as an important reasoning component in solving the design problem; to be subordinate to the designer; to permit designers to apply different approaches; and to allow for multiple users and tasks. Furthermore, the system should support the designer throughout the entire process by advising on relevant knowledge, methods, tools and strategies, and it should provide a structure for documenting product and process data. This answered the third research question.

The third stage focused on the fourth research question: the identification of the requirements and the main functions of the support system. The development of the requirements involved a stepwise approach.

First, based on the literature and case study, the identified characteristics of design and support were classified into:

- characteristics to support or stimulate (*e.g.*, the development of a comprehensive requirements list);
- characteristics to prevent or discourage (*e.g.*, the fact that requirements are neglected during evaluation);
- characteristics that cannot be prevented but have to be taken into account (*e.g.*, the fact that requirements change throughout the process).

This provided an initial set of requirements, which was translated into more specific requirements by indicating *how* the system could support, prevent or take these characteristics into account. The compilation of these requirements extended with the requirements that could be derived from them, resulted in the final set of requirements. The requirements focus on the interaction between system and user. They take into account the design process as it actually takes place and how it could be improved. The final set of requirements was used to define the main functions of the envisaged design support system, which were:

- supporting methodical design activity;
- supporting structured documentation of design data;
- supporting retrieval of design data for reuse;
- structuring and supporting retrieval of knowledge, methods, tools and data of past designs;
- providing two types of context-sensitive advice: assistance (suggesting knowledge, methods, tools and past designs) and guidance (suggesting steps in the process);
- supporting communication and teamwork.

Design data includes:

- *product data*: data describing the product as it evolved, covering versions, all product development stages, the whole product life-cycle, and the alternatives that were considered at any one stage.
- *process data*: the rationale behind the product data (arguments, decisions, design intent)
- *process administration data*: planned and applied resources (who, what, when and how).

The fourth stage of the project involved the development of the model of design (the Design Matrix) and a description of the architecture of the support system built around this Design Matrix. This system was called PROSUS, PROcess-based SUpport System. The aim of PROSUS is to improve the design process by using a *process model* to support the capture of the data resulting from design activities and to support the creation of these data throughout the process. Several alternative high-level architectures were generated and evaluated using the main requirements (that encapsulate the way designers design) and the functions that had been formulated. The project did not include the implementation of the system. This stage answered the fifth research question and the first part of the scientific aim.

In the final stage of the project a first evaluation of the concept against the formulated functions and aims took place to determine the applicability and the usefulness of a process-based approach to computer-supported engineering design. This stage contributed to the second part of the scientific aim. The evaluation was based on a comparative analysis of the design processes of designers working with and without the Design Matrices.

C.2.3 Results

The envisaged system PROSUS is shown in Figure C.2.

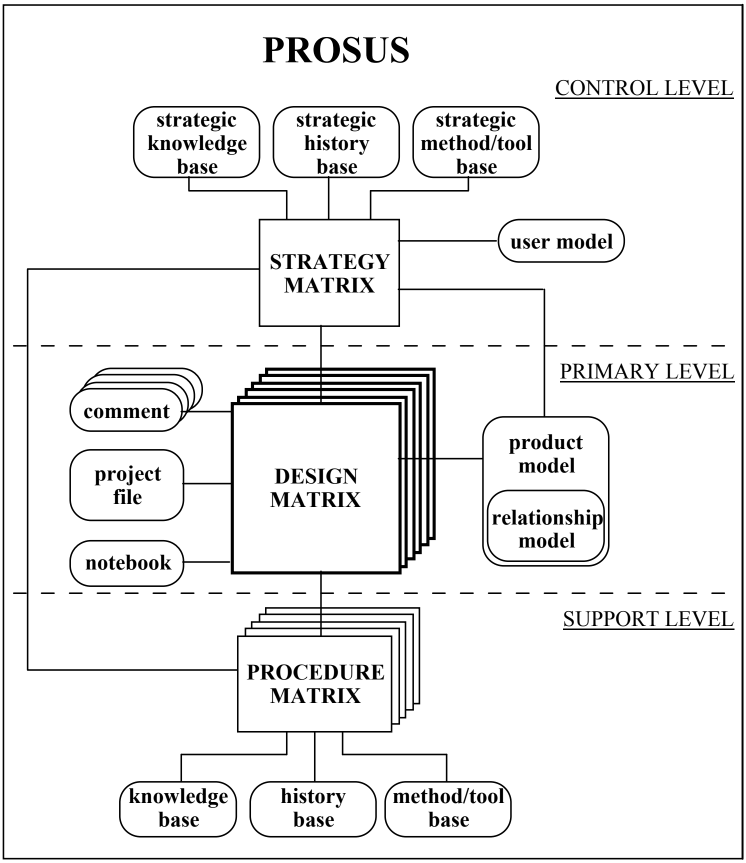


Figure C.2 PROSUS

PROSUS has three levels:

- the primary level, containing the basic building block of PROSUS, the Design Matrix, which represents the design process and is the main working area for the design team;
- a control level to aid in finding the current most promising strategy, that is, the sequence of steps (cells in the Matrix);
- a support level to determine the best resources (people, time, means) for each step.

The chosen model of design, the Design Matrix, is a compilation of models of design proposed in the literature. The granularity of the model was chosen such that it was fine enough to provide the context for the system to be able to act on the input of the designers; and coarse enough not to be a burden to the designers when documenting their process. The model constitutes the system's generic knowledge of the design process. The Design Matrix represents the design process as a structured set of issues and activities. Figure C.3 shows a simplified design matrix.

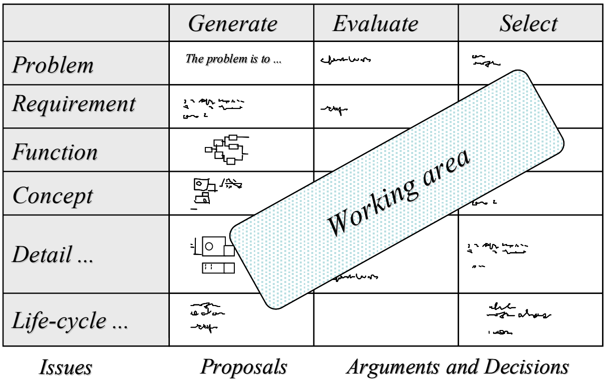


Figure C.3 A simplified design matrix (the issues following Concept design have been grouped together under Detail design)

Each cell can be viewed as a window in which data can be entered, either using the default notebook type interface, or using existing design tools such as requirements capture, drawing or simulation tools. Data can be entered in any order; the sequence of addressing the cells is not prescribed, nor do all cells have to be visited.

The first column contains a generic list of issues for which to generate a proposal. For reasons of clarity, the various issues following 'Concept', including material selection, assembly, installation, use, *etc*. have been grouped in this figure under the heading 'Detail design'. Each issue can be solved in three steps: Generate, Evaluate and Select. A Generate step results in proposals that address a given issue. Evaluation is the comparison of each proposal with the requirements and results in one or more candidate solutions (decisions) and related arguments. In the case of more than one candidate solution, a Selection has to be made to decide upon the solution to pursue. This activity results in one and sometimes more solutions (the decision) together with the arguments for this selection. A distinction has been made between Evaluation and Selection because of the nature of these activities. Evaluation focuses to a large extent on the demands in the requirement list, selection focuses more on the wishes (or soft requirements), and will involve making trade-offs. As a consequence, different methods and tools will be used for each of these activities.

Any one design project consists of a set of matrices, all linked together. New matrices are introduced throughout a project, one for each assembly or component that has been considered. This is based on the fact that the process model repeats itself, in a slightly modified form, for every assembly and component that is being generated (the two types of matrices vary slightly). A design matrix thus contains a description of the design process for a specific product element at a certain point in time. The description is structured around the rationale applied and may not be chronological (automatic date/time/person-stamping the entries allows retrieval of the chronological order).

Each of the formulated functions is realised as follows:

- supporting methodical design activity;
  The matrix does not describe how designers *do* design, nor does it prescribe how they *should* design. It suggests how designers *could* design by providing a framework for their collective activities in a project. The matrix and the use of multiple matrices encourages a more systematic approach on a project as well as on the assembly or component level.
- supporting structured documentation of design data;
  The matrix provides a structure for the design team to document and retrieve design data in all stages of the design process, which is enhanced by the use of multiple matrices. In each cell in a Design Matrix users can document the results, both final and intermediate, of executing that particular step, using a notebook-type interface.
- supporting retrieval of design data for reuse;
  The cells in which the data is entered and the matrix provide the context for the system to 'understand' what the data is about. That is, the process model (the Matrix) enables the system to determine the context in which the data was generated and to use this to index the data for storage and retrieval.
- structuring and supporting retrieval of knowledge, methods, tools and data of past designs;
  Knowledge, methods, tools and data of past designs (matrices used in past projects) can be linked to specific activities in a design process, *i.e.*, linked to one or more cells of a Design Matrix. This supports integration and is expected to encourage application.
- providing context-sensitive advice:
  - assistance: the Design Matrix provides the context for the system to know in which activity a designer is involved by identifying the cell in which he or she works. This enables the system to advise on relevant knowledge, methods, tools and past designs, because these can be linked to specific activities, that is, to one or more cells.

  - guidance: The model as such suggests an overall approach, while allowing different approaches. The Design Matrix in combination with the strategy level of PROSUS will enable the system to suggest which step to take next in a more advanced way.

- supporting communication and teamwork.
  As a workbench for the design team all data becomes available as it is being created so that designers can act upon the latest data (apart from the data in the temporary personal notebook area). The comment facility allows suggestions and remarks to be added to entries of others in the team.

C.2.4 Evaluation of the Results

The PROSUS concept was evaluated by evaluating the Design Matrix concept against the formulated functions and aims, prior to computer implementation. The Design Matrix was chosen because it is the core of the system. The applicability and usefulness of the Matrix is crucial to the applicability and usefulness of the system. Evaluation prior to implementation was considered because implementing a system that would be suitable for evaluation would put high requirements on the quality of the user interface (the interface had to allow writing and drawing as easy as in a notebook). Given the available time and computing skills the quality of the interface was likely to have affected the evaluation in such a way that conclusions about the concept could not have been drawn. More importantly, it was considered possible to evaluate the matrix as a concept before actual implementation, even though not all system functions could be evaluated. The solution was to draw the matrices on large sheets of paper on which designers could work, thus avoiding the interface problem. Given the experimental setup, absolute statements about the applicability and usefulness of the matrix concept are difficult, if not impossible, to obtain. Instead, it was decided to focus on obtaining relative statements by comparing designers using the matrices with those who do not use the matrices.

To focus the evaluation, the system functions and the overall aims were translated into evaluation questions. For each of these questions measurable hypotheses were formulated that described the expected effects of the system on the process and on the resulting design. For example, ‘Compared to designers working without design matrices, designers using design matrices document more arguments and decisions when evaluating and selecting their designs’, or ‘Designers using design matrices do not experience difficulties in using the matrices’.

An experiment and a questionnaire were set up to evaluate these hypotheses. The experiment involved eight designers doing the same design task in a laboratory environment. Four designers applied the design matrices (drawn on paper), four designers worked in the usual way without matrices (on blank sheets of paper*).* The designers were asked to design a wall-mounted swivel mechanism and produce a layout drawing. They were given a one-page problem statement, data about an imaginary workshop and some general handbooks. They worked individually and were asked to think aloud while designing. All processes were recorded on videotape. On average the designers spent 4 hours. The researcher had a list of possible questions and answers in case designers had any questions about the problem or its context, to avoid different answers being given to different designers. Questionnaires before and after the experiments were used to obtain information about the designers' backgrounds and their opinions about the experiment. The designers had between 5 and 40 years of experience. The task and set up were a modification of the experiment described in Dylla (1990); Fricke (1993).

The videotapes were transcribed and broken down into events (usually meaningful parts of sentences). Events lasted on average 8.6 seconds. Using a spreadsheet, each event was categorised for each of the issues that needed analysing in order to evaluate the hypotheses. One column, for instance, was used to categorise the events according to whether the event contained an argument and the nature of this argument, another column was used to categorise events according to the cell in which the designer was working. Once categorised, suitable graphs and tables were generated. Given the type of data and the low number of cases (eight) Fisher's exact probability measure was used to determine significance of the findings.

With respect to the functions of the system, the analysis showed that designers using the design matrices used a variety of approaches. The questionnaires revealed that they did not have the feeling that their ability to work had been strongly restricted because of the design matrices and they felt the design matrices provided a reasonable structure for their design processes. The designers who used the matrices generated more concepts; addressed more issues; documented more arguments and decisions; evaluated more often; and evaluated more continuously throughout the design process. These findings are promising: many typify successful design processes.

With respect to the aims, the designers who used the matrices had slightly more confidence in their solution. Product quality was used as the measure of effectiveness in this experimental setup, which was determined by two independent experienced designers who compared the designs with a set of demands and wishes that were partly derived from the task description and partly generic, such as low wear. No difference could be measured. This result was expected because the subjects were experienced designers and the task chosen was relatively simple because of the time-consuming nature of this type of experiment. Design time was used as the measure of process efficiency in this experiment. Designers who used the matrices took longer than the other designers; but documented more; were on average less experienced (strong correlation between design time and experience was measured); assessed their solutions significantly more often; and addressed more issues, *i.e.*, they covered a larger part of the design process.

The evaluation suggested that the proposed process-based approach to design is applicable. With respect to the functions of PROSUS that could be evaluated in the experiment, the approach was considered useful. The aim of improving the effectiveness and efficiency of the design process could not be proven: possible reasons related to the experimental context were identified. However, the use of design matrices was found to modify the design processes in a way considered likely to contribute to improvement. Based on the evaluation, it seemed worthwhile to implement the concept.

C.2.5 Conclusions About the Research Approach

The research approach developed as the project progressed since no overall approach to design research existed. Research approaches from various non-engineering disciplines had to be consulted and several iterations took place. The use of descriptive studies and the evaluation involving designers were considered extremely worthwhile methods for the development of design tools. The overall approach was found to be sufficiently rigorous to act as the starting point for the design research methodology described in this book.

C.2.6 Continuation

After the project finished, a first implementation of PROSUS was developed. Based on the experience with this implementation a completely revised version is now being implemented to allow a more thorough evaluation of the system.

C.2.7 References

Asimov M (1962) Introduction to design. Prentice Hall, Englewood Cliffs NJ

Blessing LTM (1991) Engineering design and Artificial Intelligence: a promising marriage?. In: Cross N, Dorst K and Roozenburg N (eds) Research in design thinking, Delft University Press, Delft, pp 177–212

Blessing LTM (1994a) A process-based approach to computer-supported engineering design. PhD thesis, University of Twente, Twente

Blessing LTM (1994b) From design characteristics to system requirements. Cambridge University Engineering Department. Technical Report CUED/C-EDC/TR16

Dylla N (1990) Denk- und Handlungsabläufe beim Konstruieren. PhD Thesis, Technische Universität München, Fakultät für Maschinenwesen, München

French M J (1971) Engineering Design: the conceptual stage. Heinemann, London

Fricke G (1993) Konstruieren als flexibler Problemlöseprozeß: empirische Untersuchung über erfolgreiche Strategien und methodische Vorgehensweisen beim Konstruieren. PhD thesis, Technische Universität Darmstadt, VDI Fortschrittberichte 1.227,VDI Düsseldorf

Pahl G, Beitz W (1996), Engineering Design. 2nd edn, Springer Verlag, Berlin

C.2.8 Reflections from the DRM Perspective

The work of Blessing *et al.* has the *social aim* of improving the mechanical engineering design process by making it more conscious, effective and efficient. With the assumptions that focusing on the design process influences its effectiveness and efficiency, that awareness is the first step towards improvement, and that computers can enhance effectiveness and efficiency of the design process, the *scientific aim* of this work is to develop and evaluate a model of mechanical engineering design upon which to build a computer based support tool in which the design process is the core. The central question asked is: What model of design can be used to develop a process-based computer support system for mechanical engineering design that improves the design process?

The work is carried out in four stages. The first stage focused on an extensive study of the prescriptive and descriptive literature that could be important for support development (*e.g.*, what the characteristics are of effective and efficient design processes, successful designs and actual design processes). This was supplemented by detailed observation of a design process in industry (comprehensive DS-I) to identify additional, relevant characteristics of the design process. The chosen Success Criteria are a more effective and efficient design process, the Measurable Success Criteria are increased quantity and quality of documentation (consciousness), high product quality (effectiveness), and reduced design time (efficiency). On the basis of this, a comprehensive PS was undertaken. The Intended Support had the functions of supporting a methodical process, structured documentation, retrieval of data, knowledge methods and tools for reuse, providing context-sensitive advice, and supporting teamwork. A paper-based Actual Support was developed, which was evaluated using a comparative analysis in a laboratory setting of design processes of experienced designers working with and without the Actual Support (comprehensive DS-II). Based on the evaluation, suggestions for improvement of the support were made. The work continued beyond the thesis to develop an initial implementation of the Intended Support (Initial PS).

The work of Blessing *et al.* is an example of a PhD thesis that carried out comprehensive studies in the stages of DSI, PS and DSII. Using the DRM framework, the project can thus be classified as a Type 7 for the PhD work, as well as for the whole project:

Type 7: RC (Rev) → DS-I (Comp) → PS (Comp) → DS-II (Comp) → PS (Init)

C.3 A Program for Computational Synthesis and Conceptual Design Support

Student: Amaresh Chakrabarti (author)
PhD dissertation University of Cambridge, UK, 1992 (Chakrabarti 1991)
Supervisor: Thomas P Bligh

C.3.1 Introduction and Aim of Research

Conceptual design is an early stage of the design process where design concepts are created, evaluated and explored before being selected for further development. The process is characterised by information that is often imprecise and incomplete. However, it is in this early stage, which is the least resource-intensive among design stages, where most of the product cost is committed. This is why improving conceptual design can have a far-reaching and effective impact on the design process. The broad, long-term aim of this project was therefore to develop means for supporting designers at the conceptual stage in order to develop better products.

An initial study was undertaken into existing literature in order to understand the state of work in this area, and to focus the research aim. The literature studied include characteristics of designers, design process and design aids at the conceptual stage. Design research prescribes that in order to develop better products, designers must generate a wide range of concepts and evaluate these appropriately. Evaluation can only be as good as the criteria used, and many criteria are generated from the positive and negative features of proposed concepts. This means that considering a wide range of concepts should not only help generate more innovative ideas, but also help evaluate these using a wider set of relevant criteria.

However, designers seldom generate more than a few concepts. There are several reasons for this: lack of awareness of (partial) solutions in other designs or domains; bias towards or against specific ideas found useful or difficult respectively in the past; and a limited information-handling ability to detail more than a few ideas, especially since information tends to grow rapidly as ideas are further detailed. Computers are potentially useful for enhancing information handling, and they can be unbiased in terms of the ideas presented. Large databases are easy to create, maintain and use in computers. However, computers are hardly used in this early stage of design. One way to fulfil the aim of this project was to computationally support designers to develop a wide range of concepts, and to help them evaluate these using the widest possible range of relevant criteria.

Existing aids to concept generation were either systematic manual approaches, such as in Pahl and Beitz (1996); catalogues of existing components or designs (Roth 1970); or product development principles (French 1985). These rely heavily on designers as to how effectively they are used, and little domain expertise is reused from the past, except in the case of design catalogues. However, much valuable experience is stored in the designs catalogued, and it was hypothesised that the building blocks that these designs share can be more effectively used if they can be combined freely, rather than remaining part of the designs where they were originally used. Each aid has its merits however. Systematic approaches prescribing combination of partial solutions into concepts hold the promise that all combinations can be considered without bias. Catalogues of existing solutions, if seen as combinations of common building blocks, promise their reuse, and therefore consideration of a wide range of concepts. Design principles could be used in guiding evaluation of concepts. The specific aim of this research therefore was:

> *to develop computer methods for supporting conceptual design by supporting (i) formulation of intended design problem, (ii) unbiased generation of a wide range of concepts to solve the problem by combining a set of building blocks, and (iii) evaluation of these concepts.*

However, due to limitations of time, the immediate aim of this research was to focus on the first two of the above sub-aims, with the understanding that evaluation, the third sub-aim, would be left entirely to designers at this stage, without any active support from the computer. The assumption was that fulfilment of the first two aims would support evaluation.

C.3.2 Research Approach

There were four, variously coupled stages in the development of the computer support. The first stage was to sufficiently clarify the required characteristics of the support and its components. Stage II involved developing these components so that these could be used to start developing a flow-chart, and consequently a pseudo-code (stage III) of the eventual program, which could be hand tested by taking example cases. Once this was satisfactory, the pseudo-code could be turned into a real program and tested for these cases (stage IV), before gaining confidence for a more formal evaluation.

In stage I, several alternatives were considered as to how an unbiased generation of a wide range of concepts could be supported on computers. Given that the broad approach was to generate concepts by combining building blocks, generation would require a set of building blocks, and some means for combining these into possible concepts. Several alternative approaches were considered and evaluated. It was decided that a database of a *wide* range of building blocks, separately developed, would be used for synthesis. In order to make unbiased generation of a wide range of concepts possible, an *exhaustive* synthesis of these building blocks would be performed *by* the computer. These solutions would then be offered to designers for consideration, further exploration and evaluation.

This required developing (stage II) a representation of design problems and conceptual solutions, and a program (*i.e.*, a knowledge base of building blocks, and a synthesis procedure) capable of generating an exhaustive set of conceptual solutions to given design problems, both expressed using the representation developed. The research method for developing these was to use as 'data' known design solutions and the design problems they solved. The knowledge base of building blocks was to be such that it could be used to adequately describe at least these solutions. The reasoning procedures should be able to generate back at least these solutions.

The method is shown in Figure C.4. First, a set of design problems, in terms of their functions, and their known design solutions was collected. By analysing these for commonality across their functions and means, a representation for these designs and their functions was developed, and a knowledge base of common building blocks to describe the designs created. A reasoning procedure was then developed for combining these building blocks into concepts for solutions to a given design problem. These concepts were then compared with the known designs. It was important that the concepts generated by the computer include the existing designs as well as new concepts. To be generally useful, it must also generate concepts for new designs for design problems not used in the development of the program.

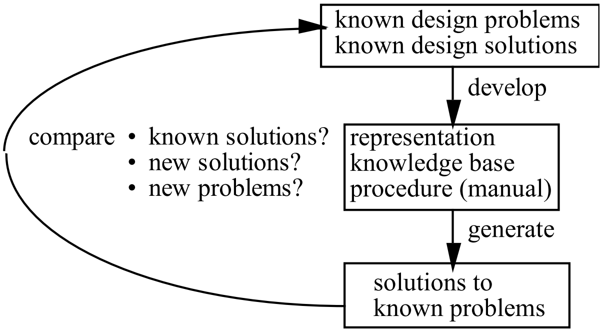


Figure C.4 Research Method in Stage II

Designs were collected from the literature, particularly design catalogues, for a wide variety of mechanical design problems to determine a common representation. Analysis of these designs revealed that mechanical devices had temporal as well as spatial (topological and configurational) features: they function through time, and at each instant in time they can assume potentially different spatial configurations. However, it was also noticed that the designs analysed have, or can be approximated to simple, orthogonal configurations at some instants of time during their operation. In other words, if the design problems were considered as a combination of a set of instantaneous functions, then a design solution could be synthesised to satisfy one of these instantaneous functions, this could then be checked for satisfaction at the other instants for complete satisfaction of the temporal function. The synthesis process then became two-stepped, one of solving for an instantaneous function, and then for checking for the overall temporal function. This simplified generation of solutions from the temporal perspective, although they had to be synthesised at both topological and configurational levels, as many existing designs used spatial variations of the same principle. Developing representation and the knowledge base of building blocks required, therefore, describing concepts and their building blocks at these levels.

The central issue in developing the database of building blocks was that their exhaustive combination had to provide a wide range of concepts. Several alternatives were considered. The process was to iterate through analysing the designs and consulting the relevant literature, especially design catalogues, trying to identify the minimum number of building blocks with which these designs could be represented, without any of these building blocks being rigid combinations of the other building blocks. The building blocks developed were named basic elements.

The first step in developing the synthesis procedure was to clarify the relationship between the functions of the building blocks and the overall function of the designs to which they belonged. The approach then was to consult relevant literature and adapt available combinatorial algorithms. Although AI Search literature provided general guidance, there was nothing that could be used directly. There were two difficulties. One was that potentially an infinite number of topological and spatial combinations were possible among a set of building blocks. This meant that a rational approach had to be provided to limit these possibilities. Investigation revealed that the reason the number of topological combinations was infinite was due to not restricting the number of building blocks used in a single solution. It was tentatively decided to generate an exhaustive set of building block combinations within a pre-specified restriction on the number of building blocks to be used in a single combination. In this was verified by discussions with a number of designers within the department, who all felt that this was justified, as designers would like to explore simple concepts (with a minimum number of elements) first. Regarding the possibility of an infinite number of spatial configurations for a concept, it was a compromise decision to generate only orthogonal spatial configurations for the concepts. This way at least one configuration, and often many, for each concept could be generated, and with a facility for their configuration modification this can be extended by the designers. The second problem was that while usual search procedures had single initial and goal states, a mechanical design problem could have several inputs and outputs, which required substantial modification of the search procedures. The approach was to do this in steps. First, the single input–output procedure was generated, implemented and verified. This was inverted to use as multiple input–single output procedure, and these two were used as building blocks for the development of multiple input–output procedure.

Stage II of the development of the synthesis procedure was the development of its flow chart and pseudo-code. This was tested, prior to computer implementation, by the researcher simply following the procedure by hand, and evaluating the outcomes. The testing was tedious and error prone, and needed doing several times before gaining confidence. However, once this was done, the representation was much closer to stage III, implementation. Implementation was a major step, and required a language in which it was easy to describe the representation and that provided the symbolic manipulation the reasoning procedure demanded. LISP was chosen, because it was available and satisfied the above criteria. The approach was to implement the simplest, lowest-level functionalities first, and bind them using the high-level functions only when confident that the low-level code worked. Little effort was spent on user-interface development, as the aim was to see if concepts were generated exhaustively. The comparison with existing design problems and solutions was to be done by the researcher himself.

C.3.3 Results

The research led to a database of building blocks, a method for representation along with a reasoning procedure implemented as a computer program that can automatically synthesise these blocks into concepts to solve 'instantaneous' versions of a required temporal function (the problem). These are illustrated with three devices: a door latch, a paper punch and a jig-saw mechanism (Figure C.5).

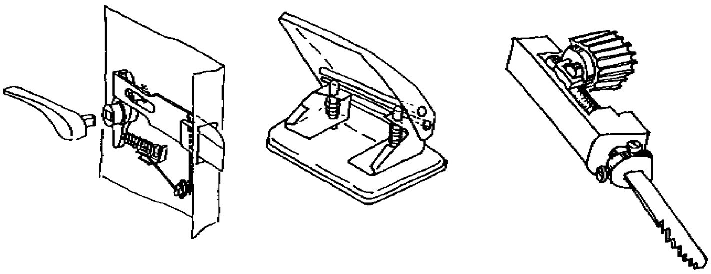


Figure C.5 Three illustrative devices

In terms of their functions, each device requires some input and produces some output, sometimes more than one. Each input (I) or output (O) has characteristics that may change with time. However, if the functions of the devices are observed at a single instant of time, their functions can be adequately described in terms of the following I/O characteristics: (i) their *kind* such as force, torque, linear or angular motion; (ii) their *direction* of action (*e.g.*, up, down or sideways); and their (iii) *magnitude* and (iv) *position*. For details, see Chakrabarti (1992) and Chakrabarti and Bligh (1994, 1996a, 1996b). This is shown in Figure C.6.

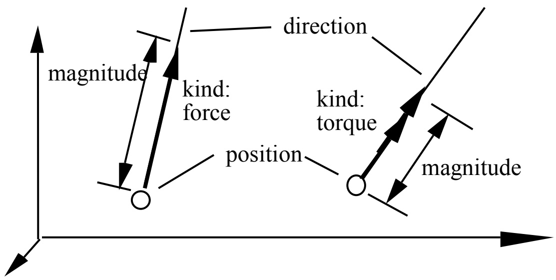


Figure C.6 Problem representation (I/O characteristics)

A design problem is represented by its intended function, represented by a number of inputs and outputs, each having a kind, direction, magnitude and position. Using this representation, for example, one could present the door latch function as a vertical downwards force input (I) to be transformed into a horizontal leftwards translational output (O), see Figure C.7.

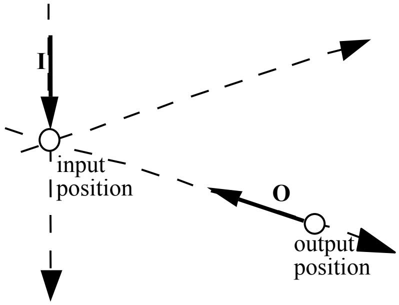


Figure C.7 Representation of the door-latch problem

How are these overall, intended functions realised by the devices? Each device is composed of a number of elements (building blocks) connected in a certain way. Some of these elements function in a similar way, although their embodiments are different. For instance, when the door latch handle is pushed on the input end it rotates at the other. Although visually different, functionally this element is similar to the top part of the paper punch that, when pushed down on its input edge produces a rotation at its pivot end. The crank of the jig-saw is similar, where the kinds of input and output are simply reversed. The spatial configurations of the input and output of these elements have a definite relationship: they are orthogonal and non-intersecting. The direction of their inputs and outputs depend on the direction in which the element is laid out in space. Solution concepts, therefore, are described as combinations of functional elements such as a lever, of which the door latch handle is an example embodiment. Each element is defined as one of five basic types or their combinations, see Figure C.8, and is distinguished by the spatial relationships between its inputs, outputs and distance between their positions (*i.e.*, the length of the element).

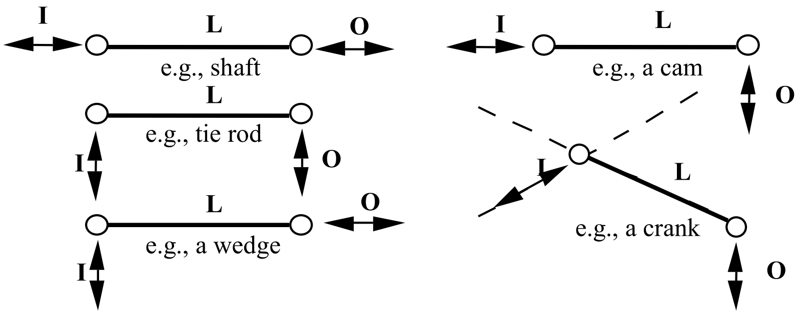


Figure C.8 The five basic elements

The five basic elements were arrived at by finding that an input and output of a mechanical element, both of which are (pseudo-) vectors, can have only five spatial relationships: both input and output can be parallel to each other and coaxial to the length of the element (L); both can be parallel to each other but perpendicular to L;

the input can be perpendicular to L but output parallel; the input is parallel to L but output perpendicular; or both input and output are perpendicular to L.

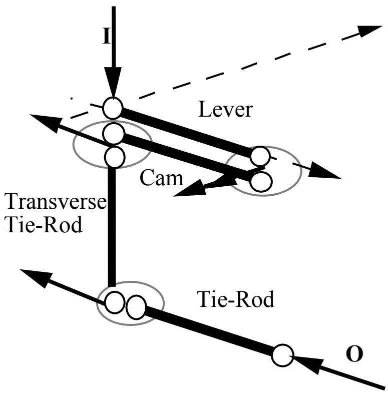


Figure C.9 Representation of a door-latch solution

C.3.4 Evaluation of Results

The evaluation was to test whether the program generated an unbiased and wide range of concepts. Two evaluations were carried out. The first included a test of exhaustiveness so as to test unbiased generation of solutions, *i.e.*, whether all valid combinations of given building blocks were produced. This was tested by comparing the number of solutions generated with those calculated mathematically. They matched in all cases (over fifty) tested. As exhaustive combinatorial synthesis is computationally expensive, it was critical to measure the computational efficiency of the program and how this related to the size of the database of elements and the specified maximum number of elements allowed in a single solution. As the time to generate a set of solutions was proportional to the number of solutions generated, computational performance was measured by measuring the number of solutions generated by the computer with various maximum numbers of elements allowed per solution for databases of various sizes. It was found to be exponential with respect to both database size and the maximum allowed number of elements per solution. This was further verified by developing independent, theoretical models. This confirmed the earlier hypotheses: limiting the number of elements in the database by using basic elements to ensure efficient generation without compromising range, and using the maximum allowed number of elements per solution as a means to control the number of solutions generated.
External evaluation was to test whether a wide range of design concepts were generated, which was taken as generating existing as well as new concepts. The program was developed and tested using several devices, including door latches, toilet door locks, bicycle transmissions, lawn-mower drives, nutcrackers, paper punches, power-saws, holding devices and window-opening mechanisms. Some of these were used in the development of the representation and the database. In all

cases, the program generated solutions that included the existing designs along with 'new' solution proposals. This indicates the potential of the program for generating a wide variety of solutions in a wide variety of cases. However, all evaluations were done by the researcher, and no studies were undertaken to test the usability and usefulness of the program to practising designers, or to identify side-effects produced by the theory or implementation.

C.3.5 Conclusions About the Research Approach

The program was evaluated by the researcher by comparing its outcome with existing designs for a variety of problems. However, the set of existing designs used for comparison could not be guaranteed to be exhaustive. Comparison can be extended by considering more problems, and more existing designs for each problem. Comparison can also be made with concepts generated during the design process. As the next step, the potential of such a program can be tested by a systematic evaluation by designers using it, and by comparing designers' concept generation performance with and without the use of the program, preferably in real design tasks. The generality of the approach needs to be tested in other domains.

Two issues are crucial for evaluating computer programs that synthesise using building blocks. The variety of solutions produced by such programs depends on the variety of building blocks in the database. Therefore, the more basic and wide-ranging the collection of building blocks in its database, the more likely it is that the concepts synthesised using these will compare well with existing designs. However, a database will always be limited, so the question is how to decide when a program works? It was felt that if the concept of an existing design cannot be generated using the database–procedure combination developed, the test should still be taken as satisfactory if the reason lies in missing building blocks in the database rather than in problems with representation or procedure. The second issue is that of abstraction. Comparison of existing designs with conceptual solutions generated by the program requires abstracting these designs to the level of the program, which requires elimination of detail and can be subjective. One way to resolve this is to get the abstraction done by more than one person and compare these.

C.3.6 Continuation

Since the completion of this thesis, several evaluations have been carried out. The potential of the program for generating a wider range of concepts than designers during the design process has been evaluated by comparing its concepts with those generated by designers during a real design project (Chakrabarti and Bligh 1996c). Hands-on experiments by experienced designers, where they tried to solve a common problem, indicated that the program generated all solutions expected by the designers and more. The experiments, however, highlighted two side-effects that had to be solved before the program could be used by designers. One was the number of concepts generated that was too large to be considered manually in-depth. The other was the abstract nature of the representation that was difficult for designers to understand. Research was continued to eliminate these effects

(Chakrabarti and Tang 1996). The issue of large number has been tackled by developing clustering methods and heuristics for grouping similar solutions (Langdon and Chakrabarti 1999), and for eliminating infeasible solutions (Liu *et al.* 1999a). Visualisation has been improved by developing methods for generating possible physical embodiments of the abstract solutions and for qualitatively animating their behaviour (Liu *et al.* 1999b). The generality of the approach has been further evaluated by extending it to synthesis of solution principles in the physical systems domain (Chakrabarti *et al.* 1997).

C.3.7 References

Chakrabarti A (1991) Designing by functions. PhD Thesis, Department of Engineering, University of Cambridge, Cambridge, UK

Chakrabarti A, Bligh TP (1994) Functional synthesis of solution-concepts in mechanical conceptual design. Part I: Knowledge representation. Res Eng Design 6(3):127–141

Chakrabarti A, Bligh TP (1996a) Functional synthesis of solution-concepts in mechanical conceptual design. Part II: Kind synthesis. Res Eng Design 8(1):52–62

Chakrabarti A, Bligh TP (1996b) Functional synthesis of solution-concepts in mechanical conceptual design. Part III: Spatial configuration. Res Eng Design 8(2):116–124

Chakrabarti A, Bligh TP (1996c) An approach to functional synthesis of design concepts: Theory, application, and emerging research issues. AI EDAM 10(4):313–331

Chakrabarti A, Tang MX (1996) Generating conceptual solutions on FuncSION: Evolution of a functional synthesiser. In: Gero JS and Sudweeks F (eds) Artificial Intelligence in Design ’96, pp 603–622, Kluwer Academic, Doredrecht

Chakrabarti A, Johnson AL, Kiriyama T (1997) An approach to automated synthesis of solution principles for micro-sensor designs. In: International Conference on Engineering Design (ICED97), Tampere, 2:125–128

French MJ (1985) Conceptual design for engineers. Design Council, London

Langdon P, Chakrabarti A (1999) Browsing a large solution space in breadth and depth. In: International Conference on Engineering Design (ICED99), Munich

Liu YC, Chakrabarti A, Bligh TP (1999a) A strategy for functional design. In: International Conference on Engineering Design (ICED99), Munich

Liu YC, Chakrabarti A, Bligh TP (1999b) Transforming functional solutions into physical solutions. In: ASME DTM Conference, Las Vegas

Pahl G, Beitz W (1996) Engineering Design. Design Council, London

Roth K (1970) Systematik der Maschinen und ihrer Mechanischen Elementaren Functionen. Feinwerktechnik 74:453–460

C.3.8 Reflections from the DRM Perspective

The work of Chakrabarti and Bligh had the aim of supporting designers at the conceptual stage in order to help develop better products.

A study of the literature on the characteristics of designers, design processes, and design aids at the conceptual stage was used to clarify the state of progress in this area, and to decide on the characteristics of the support (review-based DS-I). From the literature, it was found that, in order to develop better products, designers must generate a wide range of concepts and evaluate them appropriately. The literature indicated that evaluation is enriched by exploring the features of the concepts considered; thus it was argued that considering a wide range of ideas should help both generation and evaluation of concepts. It was also found that designers often consider only a few ideas, and reasons include lack of awareness, bias, and limited information processing ability.

The literature study was used to formulate the specific aim of the support – to support the conceptual stage of design by supporting formulation of the intended design problem and unbiased generation of a wide range of concepts to solve this problem, for consideration, inspiration and evaluation by the designer.

The overall Success Criterion was product quality (better products), and Measurable Success Criteria were adequacy of support for formulation of intended problem, and wide range of ideas generated. The support development stage (comprehensive PS), consisted of four related steps. The first was to clarify the required characteristics of the computer support and its components; the second to specify the components using flowcharts; the third to develop these into pseudo-code and hand-evaluate with example cases (Support Evaluation), and the fourth was to turn the pseudo-code into a computer program and evaluate using example cases and mathematical models (Support Evaluation). A further Support Evaluations was carried out once the support was realised to check for internal consistency and efficiency (whether the synthesis approach was exhaustive in a reasonable time). This was followed by an Application Evaluation, with the researcher as user, to evaluate the adequacy of problem formulation, and to evaluate the range of concepts generated by the support compared to known concepts for a list of test problems (initial DS-II).

The work of Chakrabarti and Bligh is an example of research with a primary focus on computer-based support development (comprehensive PS), along with a review-based DS-I and an initial DS-II carried out by the researcher. DS-I used study of the literature along with argumentation in order to ascertain the needs for, and requirements of the support system. The result of PS was an Actual Support that was close to the Intended Support except for its user interface, which was still inadequate for use by potential users. After the PhD, work was continued to carry out in the following way. A user interface was developed to create the first Intended Support (Comprehensive PS), the system was evaluated with designers and additional retrospective real-design process cases (comprehensive DS-II). This led to identification of areas of improvement and further development using new PhD and post-doctoral researchers (Comprehensive PS). The latter illustrates the efficacy of DRM for explaining and planning larger research projects.

Using the DRM framework, the project can thus be classified as a Type 5 for the PhD work, and Type 7 for the whole project:

Type 5/7: RC (Review) → DS-I (Review) → PS (Comp) → DS-II (Init) → PS (Comp) → DS-II(Comp) → PS (Comp)

C.4 Teamwork in Engineering Design

Student: Eckart Frankenberger (author)
PhD dissertation Technical University of Darmstadt, Germany, 1997
(Frankenberger 1997)
Supervisor: Herbert Birkhofer

C.4.1 Introduction and Aim of Research

In the research project described in this section, engineers and psychologists investigated engineering design processes of teams in industry. The overall aim was to identify the main factors influencing design work and their interdependencies, and to build up a model of collaborative design work in practice. This model should describe the interaction of the different influencing factors on the design process and provide the basis for further development of systematic design with special emphasis on teamwork. Figure C.10 shows the general model of influences from the *individual*, *group*, *external and organisational* conditions on the design process, from with which the project started.

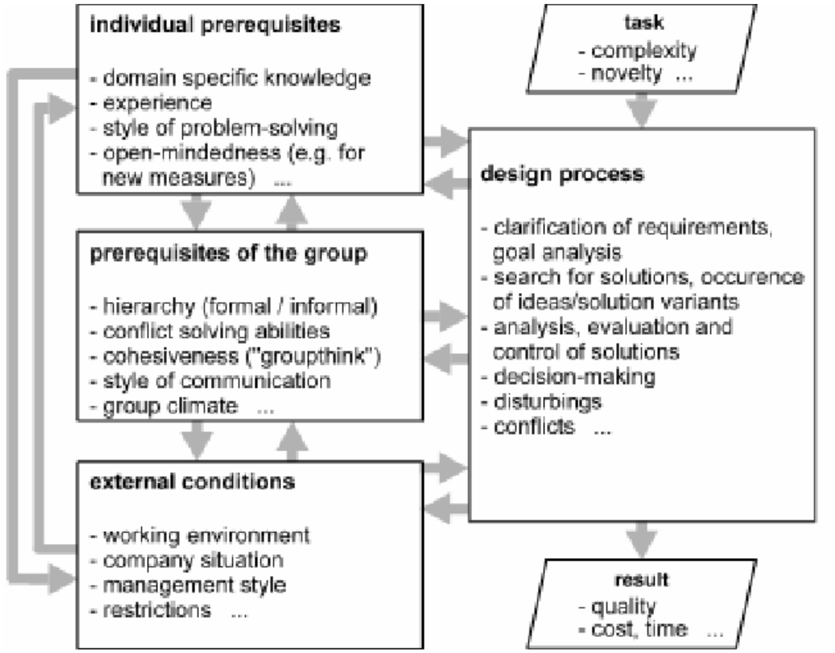


Figure C.10 Initial model of possible factors influencing the design process and its result

C.4.2 Research Approach

Although a variety of empirical studies has been undertaken in engineering design, interdisciplinary investigation of individual- and group-related factors in an industrial environment involves a new field of research, requiring a further development of existing research approaches to deal with the specific conditions of engineering design work in industry (Frankenberger and Badke-Schaub 1997).

The complexity of dealing with the large number of variables in an industrial situation makes it impossible to investigate so-called ‘comparable groups’ in several companies. The investigation therefore focused on a detailed study of two cases over an extended period of time. The researchers did not participate in the process to minimise the effect of the observer on the observed process and to allow enough time to concentrate on relevant aspects. The *first* case involved a company producing agricultural machinery. Over the course of four weeks, the design process of a group of four designers redesigning a fruit press was observed. The observed interrelations between the different influencing factors were integrated into the initial model. This model was then validated by means of a second case study, in which the important variables were reviewed. The *second* case study involved a company making capital goods. Three projects were observed of a design team developing and redesigning several components of a particle board production plant over a period of eight weeks.

The large number of influencing factors requires a variety of investigation methods to be used. The following sections describe the methods used to capture the ‘external conditions’, the ‘design process’, the ‘individual pre-requisites’ and the ‘pre-requisites of the group’.

External Conditions and Design Process

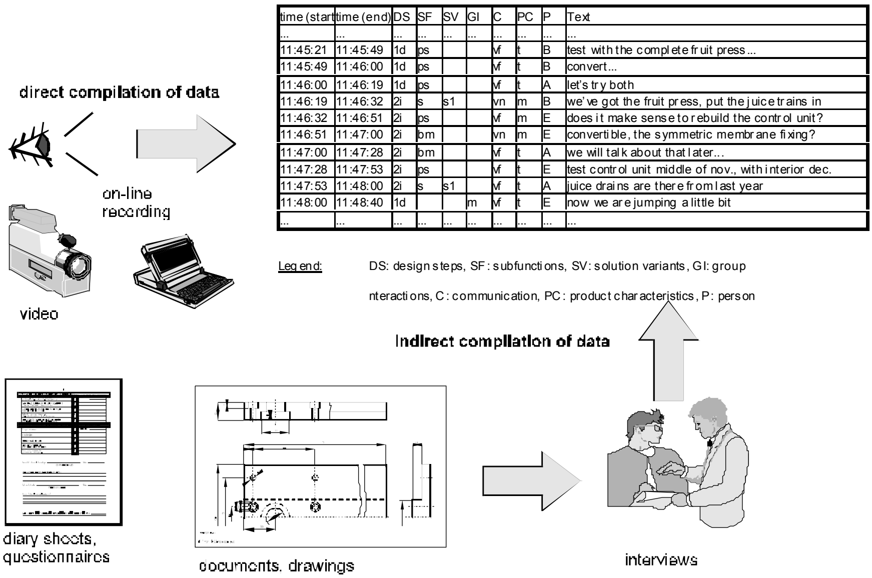
Several external conditions were captured, such as ‘branch’, ‘economic situation of the company’, its ‘culture’ and ‘organisation’, the ‘flow of information’ and the ‘communication’ within the organisation, and last but not least, the ‘direct working environment with its restrictions’.

To capture the dynamic course of the design process a detailed description at short time intervals is needed. The duration of the interval was determined by the process characteristics (the categories) used to describe the process. For this a standardised approach for investigating co-operative design in industry was developed, combining direct and indirect methods.

The primary direct method was continuous non-participant observation, involving two observers – a mechanical engineer and a psychologist - sitting in the same room as the designers. The mechanical engineer observed the activities of the designers in terms of, *e.g.*, working steps in accordance with those used in systematic design approaches such as Pahl and Beitz (1996), and the development of the technical solution in terms of sub-functions/components, ideas and solution variants. The psychologist focused on the social aspects such as decision making and group interactions. A standardised laptop-based ‘online’ protocol was used to document the observations real-time. This provided a first description of the design work as a problem-solving process. Video recordings of all team work and relevant phases of individual design work were used to review and obtain a detailed account of specific interesting phases (Frankenberger and Auer 1997).

The final protocol consisted of a word-by-word transcription of the entire process with an average duration of 30 seconds per protocol line. These protocols formed the material for a qualitative and quantitative analysis of the process, using special software, that allows easy analysis by presenting graphical representations of each process characteristic against time. These graphs, *e.g.*, represent the development of the solution by showing the moves between the subproblems and solution variants.

In addition, indirect methods were used, such as diary sheets on which designers could write down the sub-problems on which they worked, how they solved problems and when they contacted their colleagues. The sheets were designed to be completed with a minimum of effort in order to avoid a loss of motivation. Moreover, the documents produced by the designers were collected and they were asked about their process and the results. These interviews, based on the diary-sheets and the documents, provided important information about the design process and helped to understand the development of the solution and the technical decisions. Figure C.11 shows the procedure of compiling data of the design process and an excerpt of a revised online protocol.



time (start	time (end)	DS	SF	SV	GI	C	PC	P	Text
...	...	...	...	...	...	...	...	...	...
11:45:21	11:45:49	1d	ps			vf	t	B	test with the complete fruit press...
11:45:49	11:46:00	1d	ps			vf	t	B	convert...
11:46:00	11:46:19	1d	ps			vf	t	A	let's try both
11:46:19	11:46:32	2i	s	s1		vn	m	B	we've got the fruit press, put the juice trains in
11:46:32	11:46:51	2i	ps			vf	m	E	does it make sense to rebuild the control unit?
11:46:51	11:47:00	2i	bm			vn	m	E	convertible, the symmetric membrane fixing?
11:47:00	11:47:28	2i	bm			vf	t	A	we will talk about that later...
11:47:28	11:47:53	2i	ps			vf	t	E	test control unit middle of nov., with interior dec.
11:47:53	11:48:00	2i	s	s1		vf	t	A	juice drains are there from last year
11:48:00	11:48:40	1d			m	vf	t	E	now we are jumping a little bit
...	...	...	...	...	...	...	...	...	...

Legend: DS: design steps, SF: subfunctions, SV: solution variants, GI: group nteractions, C: communication, PC: product characteristics, P: person

Figure C.11 Compiling the design process using direct and indirect investigation methods

Individual Pre-requisites

Individual behaviour is influenced by several factors. A reduction of the complex cognitive, emotional and behavioural processes to one or two ‘important’ characteristics seems almost impossible. People usually behave according to the situation at hand: no paradigms can be considered universally valid for any situation or any person's behaviour, with the exception of a few psychological theories (*e.g.*, Müller–Lyer Deception (see Dorsch (1982)). For example, a person confronted by a novel, complex problem will take longer for analysis if there is enough time, if the problem is important, or if there seems to be a good chance of solving the problem. To obtain the individual pre-requisites, the following methods were chosen (Table C.2). Biographical data and personal opinions of the working environment were mainly collected by means of semi-structured interviews.

Table C.2 Variables and methods for compiling individual pre-requisites

Field of data	Variables	Methods
biographical data	- age - professional education, career - qualification and experience	- semi-structured interview - questionnaire
work environment	- motivation; job satisfaction - evaluation of the organisation - evaluation of the actual project - relationship to colleagues and to superiors	- semi-structured interview - questionnaire
ability to deal with complex problems	- analysis and information-gathering - action planning - dealing with time pressure - dealing with stress	computer-simulated micro-worlds - fire (individual) - machine (individual) - Manutex (group)
special competencies	- heuristic competence - social competence	- questionnaire (Stäudel 1987) - observing and analysing the interactions of the group
abilities concerning the design process	- clarification of the task - search for conceptual solutions - selection and control	- diary sheets/marks-on-paper - online protocol of the design process (video and tapes)

Assuming that design processes are fairly realistic examples of complex problem-solving processes, it is important to look at the strategies of engineers in complex and novel situations. The ability of the designers to deal with complex problems was assessed by analysing the thinking and action-regulation behaviour of each designer while solving computer-simulated problems (see Dörner and Wearing 1995). Each designer had to solve two problems, which were both novel, complex and dynamic. These simulations were selected because they require different manners and strategies of action regulation. Contrary to design tasks, the computer-simulated problems can be solved without any specific previous knowledge. The focus is on the action-regulation styles of individuals when being confronted with the specific requirements of different complex situations. Thus, in using these standardised problems individual heuristics and strategies can be investigated (Badke-Schaub and Tisdale 1995).

The behaviour of the subject is not measured as a single, numerical variable (*e.g.*, the 'quality' of problem solving). The planning and actual processes of the subject were investigated as strategies containing sequences of different variables such as evaluations of questions, decisions, *etc*. Other studies showed that the strategic behaviour in these simulated problems is similar to that in design work. These similarities can be interpreted as individual action styles (Eisentraut 1997).

The assessment of the special competencies (*heuristic* and *social* competence) of the designers was based on their design process (captured in the final protocols and the diary sheets) and on a self-assessment questionnaire developed by Stäudel (1987). Several studies on *heuristic* competence indicate that a positive self-assessment of problem-solving abilities supports successful problem solving in complex situations (see Stäudel (1987)). The *social* competence of the individual designers was assessed using the observations of group activities, both in the case studies and in the simulations. The aim was to capture the individuals' abilities to guide a group or to integrate into the group.

Group Pre-requisites

The aim of collecting group pre-requisites was to investigate the structure and the organisation of the group during the problem solving process and to investigate how the group approaches the problem in terms of behavioural patterns that may be responsible for producing the observed results. It was decided to focus on group interaction processes and describe these in terms of individual and group behaviour patterns. Consequently, we chose group interaction processes during the design processes and described them in terms of individual and group behaviour patterns. Another important diagnostic situation was a third computer-simulated problem that was given to the designers as a group.

Whereas the problem-solving activities demand a high extent of goal-analysis and emphasising priorities, the group situation causes the necessity for each individual to express their own ideas and strategies of proceeding. Getting his or her own suggestions accepted is linked to different characteristics of the individual, mainly the concept of social competence, which includes several abilities of acting in groups (*e.g.*, the ability to co-operate, the ability to communicate, *etc*).

The results of the computer simulation were compared with the results of specific periods in the observed design process. The same encoding system was used in both cases, based on the phases of action regulation developed by Dörner (1996). Additionally, socio-emotional behaviour and organisational aspects were categorised.

Methods and Initial Model

A summary of the data of the different elements of the initial model – the domains of influencing factors – is given in Table C.3.

Evaluation and Modelling

Distinguishing between critical situations and routine work

The preparation and evaluation of the extensive data called for a new approach that connected the data from the different fields (design process, external conditions, the individual and the group) and allowed for both, the proof of the relations and the generalisation of the findings. The basic idea of our method is the reduction of the documented design process to phases of routine work on the one hand and critical situations on the other hand where the design process takes a new direction on a conceptual or embodiment design level.

Table C.3 Methods for compiling data on the elements of the initial model

methods	**domains of influencing factors**			
	design process and result	external conditions	pre-requisites of the individual	pre-requisites of the group
interviews	•	•	•	•
online protocols	•	•	•	•
diary sheets	•	•	•	•
marks-on-paper	•		•	
questionnaires			•	•
computer-simulated problems			•	•

Types of critical situations

This method of critical situations sounds familiar with reference to the 'critical incidents' by Flanagan (1954) or the 'critical moves' by Goldschmidt (1996) but the identification of critical situations follows defined rules fitting the action-requirements of general problem solving processes (see Dörner 1996, Ehrlenspiel 1995) and of the social context. There are five types of different critical situations that can be classified into situations of goal analysis, solution analysis, solution search and additionally disturbing or conflict management as is shown in Figure C.12.

Establishing the model

Critical situations, although relatively short, determine the course of the process and are therefore of special interest in isolating the main influences on the design process. In order to extract these influences and to explain the effect of a critical situation, a sub-model was developed of the interdependencies between the influencing factors and the process characteristics for each critical situation (see Figure C.13). Evidence for each identified relation was gathered separately. Special interviews with the designers combined with video-feedback of selected 'critical situations', helped to revise the sub-models.

The sum of the different interrelations in the individual sub-models led to a model of relations between influencing factors and process characteristics in all critical situations of the design process. Altogether, 265 critical situations were identified in the four analysed projects of the two case studies. These explained the course of work by more than 2200 single interrelations between factors, process characteristics and the result. The reduction to 34 different influencing factors illustrates the suitability of the model.

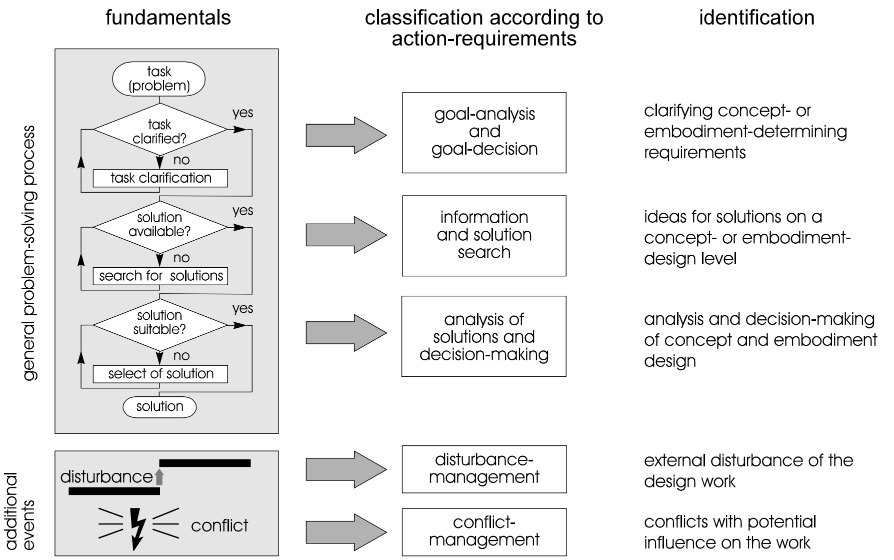


Figure C.12 Division of 'critical situations' according to the general problem-solving process and additional events (social context).

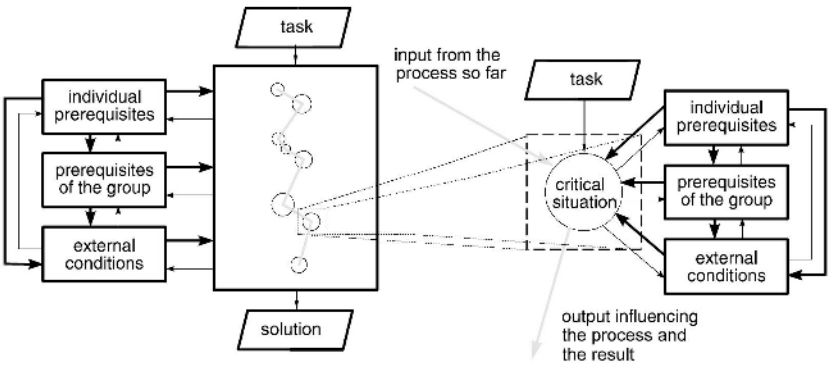


Figure C.13 Influences on the design process as influences in 'critical situations'

C.4.3 Results

As a first step, the importance of the influencing factors and their interrelations has been determined by the frequency of occurrence in all critical situations in the four design projects. However, each type of critical situation has a specific role in the design process. In order to make more specific statements on the core mechanisms leading to success or failure in the design processes, the models of each type of critical situation have been analysed separately. On the basis of this analysis we can answer questions such as, ‘which are the main factors responsible for a deficient analysis of goals?’ or ‘which are the mechanisms leading to low quality’?

In the following example the factors and relations that are important for successful or deficient decisions of solutions are analysed. Figure C.14 describes the main mechanisms responsible for *deficient decisions of solutions* found in the four projects. The thickness of the arrows depicts the frequency (in per cent) of the relations occurring in this type of critical situation. The thickness of the frames depicts the frequency (in per cent) of the factors identified in all critical situations of ‘deficient decision’.

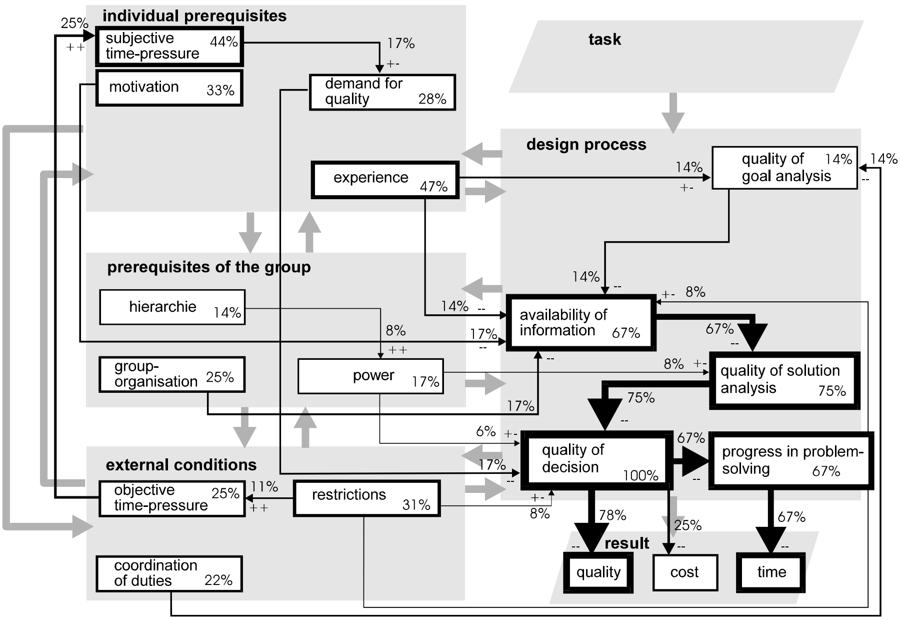


Figure C.14 Factors and relations responsible for deficient decision of solutions (in% of all 36 critical situations of this type; - - less X less Y, + - more X less Y)

From Figure C.14 we can conclude *that non-availability of information* is the crucial factor responsible for deficient analysis of solutions, which is the major reason for deficient decisions. The main reasons for lack of information are loss of *individual motivation* (*e.g.*, for transferring knowledge to inexperienced colleagues), *lack of experience*, insufficient *goal analysis* (caused by the routine of highly *experienced* engineers) and limitations in *group organisation*. Based on detailed knowledge of the situations, measures can now be proposed to avoid the identified negative mechanisms. For instance, putting emphasis on goal analysis especially if designers are highly experienced, or clearly organising the responsibilities for the clarification of requirements.

In the same way the mechanisms supporting and hindering good solution decisions and any of the other types of critical situations have been analysed. Moreover, analysis of the situations that directly influence the result reveal the interrelations leading to a high or low 'design productivity' in terms of the quality, time and cost.

C.4.4 Evaluation of Results

The central mechanisms for successful or deficient critical situations provided insights into the conditions and consequences of factors in the different decisive situations of the design process. After evaluating the results an important question remains: In spite of the high number of analysed single cases (265 critical situations with more than 2200 single detected effects of factors), is it possible to quantify generally accepted determinants and relations on the basis of the data of two companies and four projects?

Figure C.15 depicts the number of influencing factors that had to be added as each project was analysed in order to explain the critical situations in that particular project. The percentage of newly added factors as well as the relationships decreased from project to project. Furthermore, the most important relations occurred in all four projects and did so very often. This result leads us to the assumption that by analysing the four projects we captured the most important influencing factors and relations.

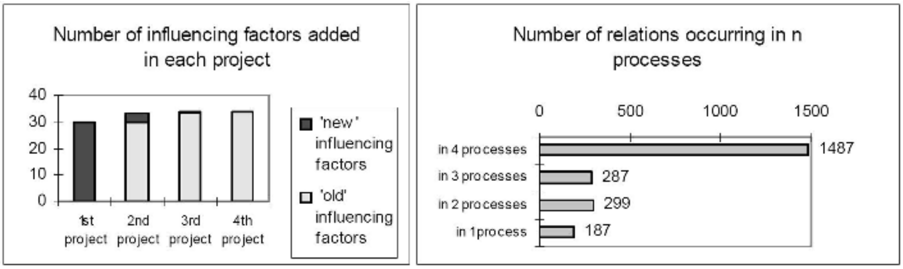


Figure C.15 Number of infuencing factors added in the four projects and number of relations occurring in *n* processes

Many of the influencing factors related to the individual and group pre-requisites. The importance of the human factor on co-operative design work becomes evident by the fact that in spite of the fact that designers worked individually 85% of their time, 88% of the critical situations occurred in group situations! Structured interviews with engineering designers and department leaders in 10 companies supported the results of the observations.

C.4.5 Conclusions About the Research Approach

The detailed compilation of data by various methods was necessary to detect critical situations in design and to explain the course of design work by means of influencing factors. The new concept of 'critical situations' allows quantification of the importance of central mechanisms determining design in industry. The knowledge gained on important positive and negative mechanisms acting in particular design situations can be used to develop suitable measures in industry and to make design education at universities more practically relevant.

However, to compile data with the methods used requires enormous effort. Consequently, one aim of further research is to develop investigation methods that require less effort but are still accurate enough to allow more focused research questions.

C.4.6 Continuation

In order to generalise the model based on a larger variety of design situations, additional investigations in the field of teamwork in engineering design practice are planned. After detailed interviews in 10 companies, a training course to enable engineering designers to detect and analyse their own critical situations was developed. So far, two investigations with a combined approach of observation and self-assessment of critical situations were carried out. Essentially the same central mechanisms of success and failure in product development were detected.

C.4.7 References

Badke-Schaub P, Tisdale T (1995) Die Erforschung menschlichen Handelns in komplexen Situationen. In: Strauß B and Kleinmann M (eds): Computersimulierte Szenarien in der Personalarbeit. Verlag für angewandte Psychologie, Göttingen

Dörner, D (1996) The logic of failure. Metropolitan Books, New York

Dörner D, Wearing AJ (1995) Complex problem solving: Toward a (computer simulated) theory. In: Frensch PA and Funke J (eds.): Complex problem solving: the European perspective. Lawrence Erlbaum Associates, Hillsdale NJ

Dorsch F (1982) Psychologisches Wörterbuch. Huber, Wien

Ehrlenspiel K (1995) Integrierte Produktentwicklung: Methoden für Prozeßorganisation, Produkterstellung und Konstruktion. Hanser, München

Eisentraut R (1997) Problem solving styles in mechanical engineering design. In: Riitahuhta A (ed.) International Conference on Engineering Design (ICED'97). Schriftenreihe WDK 25, Edition Heurista, Zürich

Flanagan JC (1954) The Critical Incident Technique. Psychological Bulletin 51(4): 327-358

Frankenberger E (1997) Arbeitsteilige Produktentwicklung – Empirische Untersuchungen und Empfehlungen zur Gruppenarbeit in der Konstruktion. PhD thesis, Technische Universität Darmstadt, Germany Frankenberger E, Auer P (1997) Standardized observation of teamwork in design. Research in Engineering Design 9(1):1–9

Frankenberger E, Badke-Schaub P (1997) Influences on design productivity – empirical investigations of group design processes in industry. In: Duffy A (ed.) The design productivity debate. Springer, London

Goldschmidt G (1996) The designer as a team of one. In: Cross N, Christiaans H and Dorst K (eds) Analysing design activity. John Wiley and Sons, New York

Pahl G. Beitz W (1996) Engineering design. Springer, London

Stäudel T (1987) Problemlösen, Emotionen und Kompetenz: die Überprüfung eines integrativen Konstrukts. Roderer, Regensburg

C.4.8 Reflections from the DRM Perspective

The work of Frankenberger and Birkhofer is primarily an investigation into the engineering design process in industry to identify the main factors influencing design work and their interdependencies, so that a model of collaborative work in practice can be developed (comprehensive DS-I). This would form the basis for further directions in supporting systematic design with emphasis on teamwork in an industrial environment.

The study led to a model of group design processes consisting of two kinds of situations – critical and routine situations – and the factors that influence these. Furthermore, the model described how these factors, especially the critical ones, influenced these situations. Critical situations are those where the design process takes a new direction on a conceptual or embodiment level; routine situations are those where it does not. Four domains of influence were identified: individual influences, group influences, external conditions and organisational conditions.

Measurable Success Criteria were taken to be product quality, cost and development time. Since the complexity of dealing with the large number of variables in an industrial situation makes it impossible to investigate the so-called 'comparable groups', the investigation focused on a detailed study of two cases over an extended period of time, one to formulate the model, and the other to validate this model. This illustrates the importance of validation of the Reference Model in an empirical study and the inherent iterations involved in the DS-I stage.

The investigation combined a variety of research methods: continuous non-participant observation by an engineer and psychologist analysing the technical and social aspects of the design respectively, online protocols, diary sheets, questionnaires, document analysis, and interviews of participants.

This project is an example of research where the primary focus is on a comprehensive understanding of the current situation (Comprehensive DS-I), with the eventual aim of support based on this understanding. It is an example of DS-I carried out in industry, where successive industry observations are used for formulating and validating the understanding. It is also an example of the results being described in what we define as 'Reference Model'.

This project continued beyond the PhD, and led to the development of a support in the form of a training programme for industry (Comprehensive PS), which has been introduced and evaluated (Comprehensive DS-II). This also illustrates a different type of support, namely training, and the possibility of using DRM to explain research programmes.

Using the DRM framework, the project can thus be classified as a Type 1 for the PhD work, and Type 7 for the whole project.

Type 1/7: RC (Review) → DS-I (Comp) → PS (Comp) → DS-II (Comp)

C.5 Measuring Conceptual Design Process Performance in Mechanical Engineering: A Question-based Approach

Student: Ade Mabogunje (author)
PhD dissertation, Stanford University, Palo Alto, USA, 1997 (Mabogunje, 1997)
Supervisors: Larry Leifer, Rolf Faste, Sheri Sheppard, Ilan Kroo

C.5.1 Introduction and Aim of Research

The investigation reported in this dissertation began with a casual observation of the pervasiveness of pronominal questions (*i.e.*, what, why, how, how much) in several of the structured design methodologies (*e.g.*, value analysis, quality function deployment, design for assembly) used by engineers in several Japanese companies, particularly those in the automobile industry. Knowing the important gains in the market share by these companies towards the end of the last decade and the widespread adoption of these methodologies by other companies (Barkan 1992), it seemed possible that there could be a link between the product development performance of companies and the question-asking behaviour of their engineers. This was the basic premise of the dissertation.

An important motivation was that establishing such a link could enable the development of other means (tools and methods) to augment the question-asking behaviour of engineers and possibly lead to further improvements in product development performance. For example, as the use of computers becomes more prevalent in the engineering profession, engineers conceivably could be given design process feedback based on their question-asking behaviour, see Figure C.16.

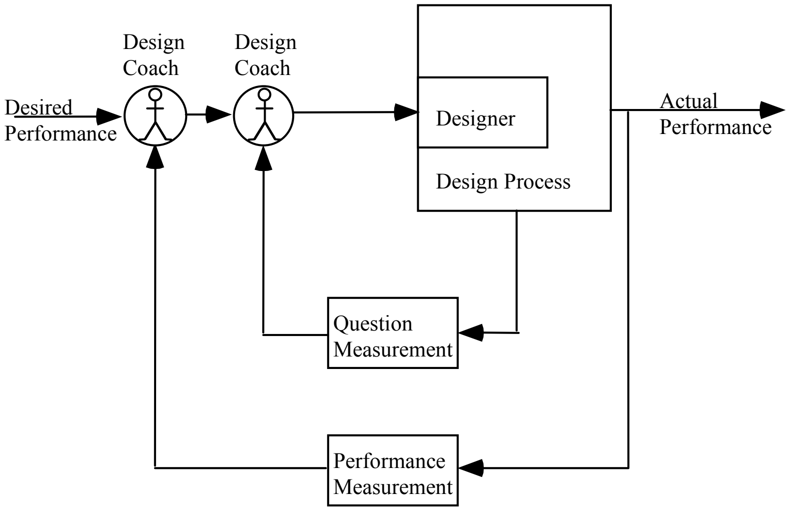


Figure C.16 Schematic of a process monitoring system for designers during product development where feedback is based on the questions being posed

If a link could be established between product development performance and the question-asking behaviour of designers, it is conceivable that questions could be used as a means of monitoring the design process. Given the increased use of computers in engineering this sort of monitoring could be used as a basis for giving process feedback to designers. A designer predisposed to asking questions of a particular nature could be encouraged to explore asking other types of questions or to pair up with another designer with a different question-asking style. It may even be possible to distinguish between a variety of question-asking styles that have led to creative solutions and those that have not.

C.5.2 Research Approach

As I thought more about the possible link between product development performance and question-asking behaviour, I realised that even if the link existed, it would be difficult to establish because the product development process as a whole is very complex. This difficulty of establishing the link greatly influenced the research methodology and cannot be overstated.

Engineering design today takes place in a complex internal (company) and external (market) environment. There are several components involved and these are related to each other in multiple ways. This sort of complexity makes the process nonlinear, highly iterative, and difficult to study. This difficulty manifests itself in two distinct research environments today – those conducted in a laboratory and those conducted in industry (Stomph-Blessing 1991). Table C.4 gives an overview of the difficulties associated with either environment. It points to the fact that the dynamic nature of real-world design and the complexity of the organisational context tend to make laboratory results irrelevant to industry, and industry results difficult to generalise.

Table C.4 Effect of the environment on types of tasks studied and on the research process

	Industry Environment	**Laboratory Environment**
Design Task	Task has a history and future	Task is self-contained
	Problem definition and requirements may change	Frozen assignment
	Internal and external factors emerge from data	Mainly functional problem solving
Research Process	Difficult to plan	Time and location can be set in advance
	Difficult to control	Control easy
	Observation interferes with project	Observation does not interfere with project
	Not repeatable	Can be repeated
	Existing environment	Environment can be optimised for observation
	Design time in months	Design time in hours

It was therefore obvious very early in this investigation that an important issue to be addressed was the level of complexity of the environment in which a product was designed and developed. As I delved into the literature, I found strong evidence supporting the idea *that the practice of design was in reality a set of strategies for dealing with complex situations* (Vincenti 1990; Lawson 1990). This in turn suggested to me that design was a way to do design research. In other words, design could be turned on itself. I will quote at some length from two of the sources I came across because they provide concrete illustration of the research methodology adopted in this thesis. The first source is historical and is from a book titled "What Engineers Know and How they Know it", (Vincenti 1990). It reviews the history of development of the "control-volume analysis", an important method of analysis for problems in the field of thermodynamics that is used by engineers and considered inadequate by physicists.

> *"The undoubted value of theorems (i.e., of control-volume analysis) lies in the fact that their application enables one to obtain results in physical problems from just a knowledge of the boundary conditions. .. engineers frequently must deal with flow problems so complex that the underlying physics is not completely understood or the differential equations that describe the phenomenon point by point cannot practically be solved throughout the flow. In such situations control-volume analysis, by working with information only on boundaries and ignoring the interior physics, can often supply limited but highly useful results of an overall nature"*

An important point to note in the above excerpt is the interplay of problem and results in the engineer's work. The results desired help to determine the appropriate formulation of the problem. The second source is empirical. It was based on a study of final-year architecture students and final-year physics students solving a contrived laboratory problem (Lawson 1990).

> *"The scientists tended to use a strategy of systematically exploring the problem, in order to look for underlying rules which would enable them to generate the correct, or optimum, solution. In contrast, the designers tended to suggest a variety of possible solutions until they found one that was good or satisfactory. .. The problem-solving strategies used by designers probably reflect the nature of the problems that they normally tackle. These problems cannot be stated sufficiently explicitly that solutions can be derived directly from them. The designer has to take the initiative in finding a starting point and suggesting tentative solution areas. 'Solution' and 'problem' are then both developed in parallel."*

As in the previous excerpt, the interplay between problem and solution should be noted. The excerpt also highlights the manner in which designers solved problems through an iterative generation of solutions. From the foregoing, four design strategies that have been used to deal with complex problems can be enumerated. These are: (1) the reduction in the size of the problem, (2) the focus on results, (3) the generation of multiple solutions and (4) the parallel elaboration of problems and solutions. As a whole, these strategies provide the key constituents of the research methodology used in this investigation.

Strategy 1: Reduction in the Size of the Problem

The necessity of carrying out design research in the real world and the difficulty associated with such studies, led to a need to reduce the size of the problem. There was the opportunity to use a design class in the university as a test bed. This class, ME210, uses industry-sponsored engineering projects as a framework for teaching graduate students the product development process. Every year, during the fall quarter, the students are supported in forming three-person teams that then bid on 12 to 18 projects submitted by a variety of companies. Over the next seven months the students propose alternative designs, investigate these alternatives, build, and test a functional prototype of the preferred design.

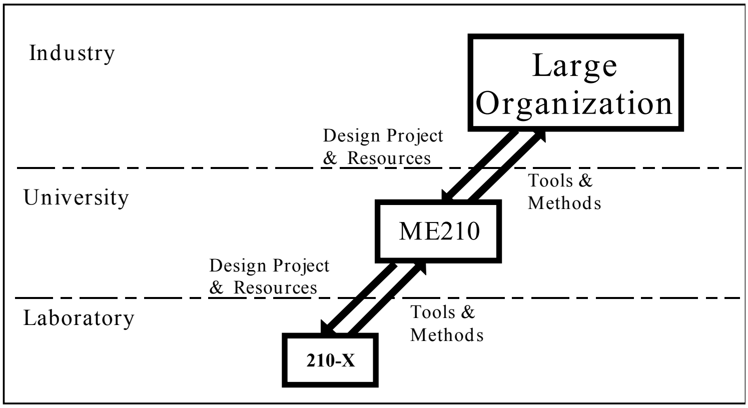


Figure C.17 Carrying out design research in less complex environments than large organisations provides a way of overcoming some of the difficulties usually associated with research in large organisations

The change in environment from industry to the classroom represents an important reduction in the complexity associated with the environment. A further reduction is possible if specific projects in the class are further isolated and studied, so to speak, in a laboratory. This strategy is illustrated in Figure C.17, where design projects and resources are taken from the larger organisation to the smaller one and lessons learned in terms of tools and method are transferred from the smaller organisations to the larger one.

Strategy 2: Focus on Results

The strategy calls for an early identification of desired outcomes as an aid to determining the aspects of the phenomenon that should be given attention. It was therefore important to have a tentative criterion for measuring performance. Earlier, I alluded to the increase in the market share of Japanese automobile companies as a measure of product development performance. In the context of this investigation I chose to begin by looking at development time. As will be seen later, as different solutions were elaborated, the criteria for performance changed. Nevertheless, a distinctive feature of this investigation was that there was always a measure of performance.

This feature marked a departure in approach from some earlier empirical studies that were particularly informative of more salient features of the engineering design process (Tang 1989; Minneman 1991). According to their approach, design study begins with observation. There is then an attempt to understand design activity and to provide design support on the basis of this understanding. If the support proves ineffective, what is proposed is to go back to better understand through observation before attempting further design support.

In the approach used in this dissertation, the research activities revolve around design performance rather than the more general notion of design activity. Starting with an assumed model of the design process, the proposition is that understanding can be developed through trial-and-error attempts to improve design performance. The error in such attempts is defined as the difference between the assumed model and the real model. Each trial inevitably seeks to minimise the error by changing the assumed model. Perfect understanding ideally occurs when the assumed model mimics the real model. From this perspective the problem involves finding appropriate parameters of performance and building models of the process that can be used to predict the performance. The two approaches are shown in Figure C.18. If an analogy is made with the case of the control-volume analysis cited earlier, it should be clear that while this approach may miss some of the detail and scope of the design process, it should make it possible to obtain operationally useful results.

Strategy 3: Generation of Multiple Solutions

Having reduced the complexity of the environment from industry to the classroom, three areas for improving performance were explored. The first, named 210-X, explored the area of information retrieval in design. The second, named Virtual 210, explored the area of computer-based organisational simulation. The third, named 210-NP, explored the area of noun-phrase analysis of the quarterly reports submitted by project teams in ME210. These three solutions were connected by two underlying themes: (1) all were based on the premise of a link between design performance and the question-asking behaviour of designers and (2) all the explorations were clustered or localised around the ME210 class.

Strategy 4: Parallel Elaboration of Problems and Solutions

In the course of elaborating the problems and solutions (*i.e.*, 210-X, Virtual 210, and 210-NP), several hypothesis were formulated and tested. This process led to the final hypothesis of this research. To better understand the premise of the final hypothesis, it is important to clarify the relationship between the three solutions, 210-X, Virtual 210, and 210-NP. To do this I will summarise each area explored under the following four sub-headings: problem, solution, results, and need. Where it is appropriate, a brief background will be included.

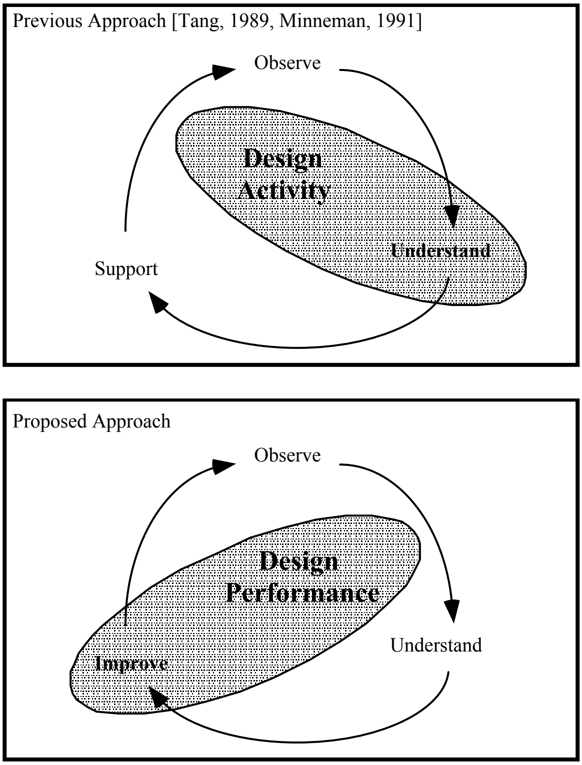


Figure C.18 A basic difference exists in the approach of this study when compared to the work of Tang and Minneman. While both approaches go through the cycle of observation, understanding, and support, the focus is different, as shown by the way that the problem at the centre is formulated. In the previous approach, there is an activity to be understood, In the proposed approach, there is a performance to be improved and as such there is a greater emphasis on the relationship between the support and the performance

C.5.3 Results and Evaluation of Results

Solution 1: 210-X

a) Problem

Given the focus on development time as a parameter of performance, there was a need to think of ways in which questions could have an effect on development time. A scenario was imagined in which the questions asked by a designer during the development of an *n*th generation design would be made available to another designer during the development of an (*n*+1)th generation design in such a way as to facilitate the access to answers.

Hypothesis: The hypothesis was that facilitating the access to answers would reduce the development time.

b) Solution
To meet the objective, two designers were separately asked to redesign a rotary friction damper to meet a new set of specifications. They were provided with a comprehensive report of the previous design and also had access to a member of the previous design team. They were given roughly six hours to come up with a design that met the new specifications. In addition, they were instructed to verbalise their questions so that these could be recorded for later analysis.

Based on the analysis of the questions, a system for indexing and retrieving answers from the comprehensive report was developed. The indices were composed of a fixed list of descriptors and a variable list of subject words that were taken from the report of Baya *et al.* (1992). The subject words were in turn used in building hierarchical question-based models of the previous requirements and the previous artifact. Based on a set of heuristics about different relationships between the models and the indices, the system was able to reformulate questions and retrieve possible answers (Baudin *et al.* 1992; 1993).

c) Results
Preliminary empirical tests using the system developed, DEDAL, showed that it reduced the development time in the (n+1)th generation design. However, this gain was largely offset by the additional effort and time needed to index the report and build the question-based models that increased the development time of the nth generation design and hence the overall development time.

d) Need
Based on these initial tests, it was clear that the key to improving the overall performance was to minimise the effort in indexing the report and building the question-based models.

Solution 2: Virtual 210

a) Background
While 210-X concentrated on a possible symbiosis between the laboratory and the classroom, Virtual 210 focused on a possible symbiosis between industry and the classroom. In 1991 a program named Virtual Design Team (VDT) had been developed in the civil engineering department at Stanford University to simulate the impact of new communication technologies on the performance of construction projects in industry (Cohen 1992; Christensen 1993). The independent variables of the model consisted of a set of communication tools and the organisation structure, and the dependent variables were the project duration and quality. The intent was to model DEDAL as a communication tool and explore the impact of its use on project duration and project quality in ME210. In other words could we predict the effect of a question-asking support program like DEDAL on the project performance of designers in ME210?

b) Problem
The reality was that the version of VDT at the time could not adequately represent information technologies as complex as DEDAL.

c) Solution
I went ahead by using a simpler and easier to model information technology than DEDAL. While this deviated from the original intent, it allowed me to explore the limits and opportunities of VDT and to gain a better sense of how it could be used to develop an organisational or multi-team perspective on the link between question-asking and performance.

Hypothesis: Since a project-based engineering class like ME210 is similar to a construction organisation in terms of the team organisation and relationship to the client, the duration, and the end product, the hypothesis here was that project-based engineering classes could be simulated with at least the same degree of realism as current computer simulations of engineering organisations.

d) Results
In VDT, there is a procedure for calculating the complexity of both the requirements and the solutions of a project. I soon realised that the question-based models developed for DEDAL could be reused in VDT for calculating these complexities. This was an unexpected synergy between design information retrieval and organisational simulation. In addition, the results of the simulation showed that much more than having the same degree of realism as current computer simulations of engineering organisations, programs like VDT could be used as a curriculum planning tool for project-based engineering classes like ME210. This was not the sort of outcome I anticipated but it clarified the need and provided additional motivation to continue the research along the lines of 210-X as will now be explained.

e) Need
The most time-consuming aspect in the VDT modelling procedure was calculating the requirements and solution complexities. This was because, compared to other inputs, they required a deeper knowledge of the requirements and the artifact. Given the benefits of the simulation, it was obvious that finding an easier way to calculate these complexities will increase the usability of VDT. This situation was similar to that encountered in 210-X where the key to improving the overall performance of DEDAL was to minimise the effort in indexing the report and building the question-based models. Therefore, at the very least, it seemed that the effort to model ME210 in VDT helped to reinforce the need and motivate the search for ways to minimise the effort involved in model building.

Solution 3: 210-NP

a) Problem
Revealed by experience from 210-X and Virtual 210, a need became apparent for a quick way of building the hierarchical question-based models. Satisfying this need would benefit designers, instructors of project-based classes, and design researchers. In the case of designers, indexing and model building will be easier to do in real time, and this will make their reports more reusable in future. In the case of instructors, simulation studies could become an additional tool for planning the curriculum and designing the organisation. In the case of researchers, the reduced

overhead would mean that a larger number and variety of projects could be used in experimental studies like 210-X. This increase in sample size in turn would improve the reliability of the research results. Thus, finding a quick way of building the hierarchical question-based models had a potentially high payoff in the sense that a single solution would be beneficial to three groups of potential "customers" simultaneously.

b) Solution

Since the hierarchical question-based models, and the requirements and solutions complexity matrices were built from the subject words in the design reports, and since the subject-words often turned out to be the noun phrases, a quick way was needed to extract noun phrases from the report.

Definition: A noun phrase is a word or group of words with a noun as its head and functioning as the subject, object or complement of a sentence. An example of a noun phrase is: “the steel bolt”.

c) Need

From a different perspective, the hierarchical question-based models and complexity matrices functioned as substitutes or surrogates for the questions whose answers were documented in the report. Since the reports were graded at the end of the quarter, it was felt that if indeed a link existed between the question-asking behaviour of designers and the development performance, a correspondence must exist between some properties of these surrogates and the grade assigned to the report. This led to the third and final hypothesis of this research.

Hypothesis: It is hypothesised that certain properties of the noun phrases extracted from the report will be positively related to the project grade in ME210.

d) Results

A quick way for extracting the noun phrases from design reports was developed using a parts-of-speech tagger and a suite of Microsoft Excel macros. The tagger takes as its input, a string such as:

The inner hub holds the steel friction disks and causes them to rotate.

and produces a tagged output consisting of two strings.

The inner hub holds the steel friction disks and causes them to rotate
at jj nn vbz/2 at nn nn nns cc vbz/2 ppo/2 to/2 vb

The lower string consists of parts of speech tags corresponding to words in the upper string. E.g., the following sequences were used to identify the noun phrases in the tagged output: “jj nn” – “inner hub”; “nn nn nns” – “steel friction disks”

Thus, it was possible to extract noun phrases from all the reports for each quarter of an entire academic year and test the hypothesis. Thirty reports (averaging 40 pages each) and representing ten projects were analysed. The results showed that the number of distinct noun phrases in a report was strongly associated with the project grade (for the winter quarter gamma = 1; for the spring quarter, gamma = 0.7). That is, projects with grades A+ or A, had a higher number of distinct noun phrases than projects with grades A– or B+, which in turn had a higher number of distinct noun phrases that projects with a B. The degree of association was not as strong as this when the project grade was compared with measures of the readability of the documents, the number of pages, or the number of words in the document. In Figure C.19, the project grades have been superimposed on a graph of the quarterly variation of the number of distinct noun phrases.

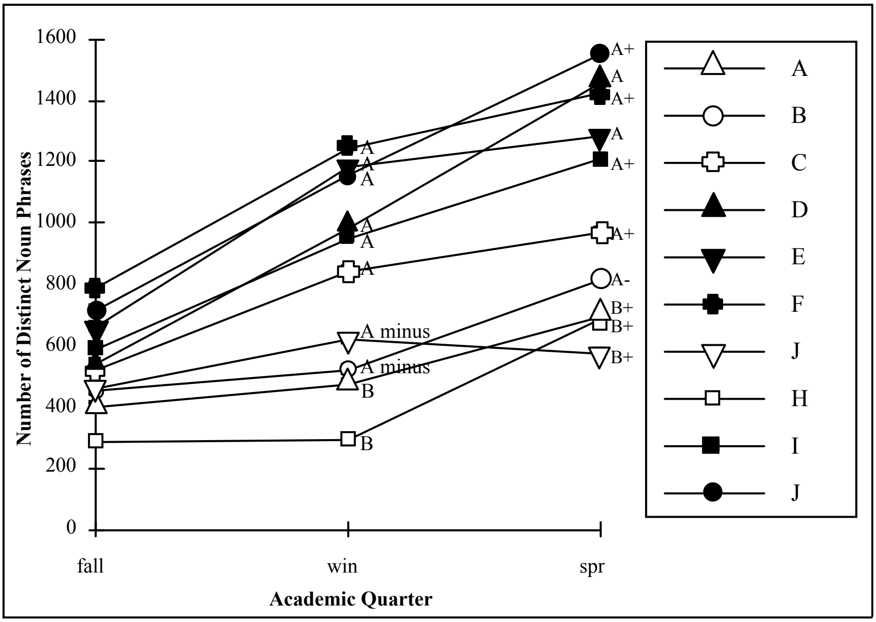


Figure C.19 The amount and change in number of distinct noun phrases per quarter versus the academic grade

On closer observation, it was also seen that between the start of the project in the fall quarter and the end in the spring quarter, several new noun phrases were formed by the students to express their ideas. In other words, new knowledge was literally being created. This phenomenon had not been observed during the experiments in 210-X, where the process lasted less that six hours. These results were quite unexpected and had several important implications. One of these being that there is a part of the process, distinct from question asking, that was involved in the creation of new knowledge. The fact that it can be readily observed and that its effect on design performance can be traced opened up a new and exciting line of research that unfortunately is beyond the scope of the dissertation.

C.5.4 Conclusions About the Research Approach

An important trade-off in this research was one between a synthetic approach and an analytic approach to research. Alternating between these two approaches was critical to obtaining operationally useful results. At the same time it had the disadvantage of introducing discontinuities in the logical lines of thought. I have had to be explicit about these discontinuities when they occur for they provide fertile ground for further research. Needless to say, it is obvious that as advances are made in our knowledge about the design process, the synthetic approach to research will become better codified and easier to explain.

C.5.5 Continuation

Since the end of this dissertation, two lines of work have been pursued. The first line has to deal with the capture of tacit knowledge during design. Yen (1999) has extracted noun phrases from several other media besides the requirements documents. Using presentation documents created by designers, email records, and video transcripts of design meetings, he found that the incidence of noun phrases is in fact much higher in informal than in formal documents. He has designed a computer-based tool, RECALL, to improve access to this informal information and is currently investigating the impact of such access on design performance.

The second line of work is aimed at understanding how the learning that occurs during the design process impacts product development performance. In traditional learning situations, the concepts to be taught and the new vocabulary are known in advance by at least one party in the transaction, the instructor. By contrast, in design-based learning situations, the concepts and in particular the new vocabulary are not known in advance. Given the increase in incidence of new noun phrases during design it is obvious that learning is occurring. What is not so obvious is how it can be improved to have a positive impact on development time or product quality.

C.5.6 References

Barkan P (1992) Productivity in the product development process. In: Susman GI (ed.) Integrating design and manufacturing for competitive advantage. Oxford University Press, New York, pp 56–68

Baudin C, Gevins J, Baya V, Mabogunje A (1992) Dedal: Using domain concepts to index engineering design information. In: 14th Annual Conference of the Cognitive Science Society, Bloomington, Indiana, pp 702–707

Baudin C, Gevins J, Baya V, (1993) Using device models to facilitate the retrieval of multimedia design information. In: 13th Int. Joint Conference on Artificial Intelligence. Chambery, France, pp 1237–1243

Baya V, Gevins J, Baudin,C, Mabogunje A. Toye G, Leifer LJ (1992) An experimental study of design information. In: 4th Int. Conference on Design Theory and Methodology, Scottsdale, Arizona, pp 141–147

Cohen GP (1992) The virtual design team: an information processing model of design team management. Department of Civil Engineering, Stanford University, Stanford, California, PhD thesis

Christiansen TR (1993) Modeling efficiency and effectiveness of coordination in engineering design teams. Department of Civil Engineering, Stanford University, Stanford, California, PhD thesis

Lawson B (1990) How designers think. 3rd edn., Butterworth, London

Mabogunje A (1997) Measuring conceptual design process performance in mechanical engineering: A question based approach. Department of Mechanical Engineering, Stanford University, Stanford, California, PhD thesis

Minneman SL (1991) The social construction of a technical reality: empirical studies of group engineering design practice. Department of Mechanical Engineering, Stanford University, Stanford, California, PhD thesis
Stomph-Blessing LTM (1991) Analysing design: a comparison of approaches. Unpublished report presented at the International Conference on Engineering Design (ICED'91), Heurista, Zurich
Tang JC (1989) Listing, drawing, and gesturing in design: A study of the use of shared workspaces by design teams. Department of Mechanical Engineering, Stanford University, Stanford, California, PhD thesis
Vincenti WG (1990) What engineers know and how they know it. The John Hopkins University Press, Baltimore
Yen S, Fruchter R, Leifer LJ (1999) Facilitating tacit knowledge capture and reuse in conceptual design activities. In: Proceedings of the 1999 ASME Design Engineering Technical Conferences, Las Vegas

C.5.7 Reflections from the DRM Perspective

The work of Mabogunje *et al.* is on measuring conceptual design performance using a question-based approach. The investigation started with the observation of the pervasiveness of pronomial questions in several structured design methods used by a number of Japanese companies, which led to an initial hypothesis that question-asking behaviour of designers may have a strong influence on the product development performance of companies, and therefore could be developed into an indicator of its performance. Several criteria for success were used, product development time was one of these. A central assumption was that student design cases could be taken as a first indication of professional performance, and hence the study could involve students rather than professional designers.

Three linked student case studies were used to progressively develop the understanding of the relationship between question-asking behaviour and performance (comprehensive DS-I), and develop and validate computer programs to enhance this behaviour (comprehensive PS and comprehensive DS-II). The evaluation showed various side-effects that offset the principal, anticipated effect, which led to the modification of the programs (revisiting PS and DS-II). An interesting feature of this project was that the solution developed for offsetting the side-effect turned out to correlate very well to product development performance and shifted the focus from question-asking behaviour to processes leading to noun-phrase creation as a principal indicator of product performance. This is an example of the opportunistic nature of scientific inquiry.

The work of Mabogunje *et al.* is an example of research where a comprehensive understanding of the current situation is developed (Comprehensive DS-I) in order to feed into a Comprehensive PS, developing a support that is then evaluated comprehensively (Comprehensive DS-II). The results led to further Comprehensive PS and Comprehensive DS-II.

Using the DRM framework, the project can thus be classified as:

Type 7: RC (Review) → DS-I (Comp) → PS (Comp) → DS-II (Comp →
PS (Comp) → DS-II(Comp)

C.6 Design for Quality

Student: Mikkel Mørup
PhD dissertation, Technical University of Denmark, Denmark, 1993 (Mørup 1993)
Supervisor: Mogens Myrup Andreasen (author)

C.6.1 Introduction and Aim of Research

In the modern, technological society products have become an inextricable part of our private and professional lives. In turn, the ability of individuals, companies and even society to operate and prosper depends heavily on the *quality* built into products. Without such quality we have at the least annoyance; at worst, poor quality can cause accidents and even disasters. The idea of quality improvement has received enormous interest from industry during the last two decades. The reason is that quality, in terms of both high quality on the market and low cost of quality within the company, is a major competitive factor on the international market. Therefore many companies spend large resources on implementing and using various approaches to the management of quality; some companies do it because 'it pays', to others it has become a matter of survival.

The subject of this thesis is the creation of quality, as perceived by the customer and the manufacturing company, during product design and development, in short, *Design for Quality*.

Through the literature and industrial studies, the research shows that despite considerable efforts in Total Quality Management and advances in technology, many companies still have serious problems in developing, manufacturing and marketing profitable products that live up to customer expectations of quality.

Quality is believed to be a strategic competitive factor, and the mastering of the quality and costs are seen as crucial in the competition with Japanese quality results. Quality is built in during mainly conceptual design, but there are many dispositional effects related to quality (Olesen 1992). It is our basic belief that quality cannot be created by control, but needs a sound synthesis approach.

At the Department of Control and Engineering Design we have worked for a decade with different Design for X-tools. It was naturally to assume that DFQ should follow the same pattern of qualitative and semi-databased principles, governed by some kind of procedure. We see product development as being the single most important contributor to product quality. Ideally seen DFQ should be based on 'best practice' organisation and support of the product development function, including

- strong links to the company's strategies;
- a supporting organisational structure;
- technology implementation linked to business objectives;
- a system for measuring quality performance.

This led to the following objectives for the research project:

- To define the concept(s) of product quality in relation to the customer, the manufacturing system and other life-phase systems, such as distribution, sales, and service.
- To establish a theoretical basis for Design for Quality including terminology and descriptive models. Major elements in this theoretical basis must be:
  - a description of where quality is placed in the technical system, *i.e.*, an identification of how different product characteristics influence product quality;
  - an identification of how these characteristics can be manipulated during design.
- To formulate a general framework for Design for Quality, containing the major means for supporting and performing Design for Quality. Obvious means already known could be: teamwork, quality tools and techniques, design procedures, organisation, management, *etc*. The DFQ framework should be exemplified and presented in a pedagogic and operational form.

The work is aimed at the following company types:

- belonging to the mechatronic industry, where the product is a mixture of mechanical, electronic and software components;
- having medium to high volume production;
- covering all activities necessary to manufacture products, including: stakeholder analysis, product development, marketing, production, distribution and sales.

The research project hereby has an interesting problem in identifying the research object. On the one hand, the object is the product developer's skill and knowledge in building in quality. On the other hand, the object is the proper understanding of the relations of his work to the business and operations of the company and to the product life phases and the quality results here. This proper understanding should be 'implemented as a mind-set' at the product developer, ensuring integrating efforts. The four central objects of research are: the product development project, the core product, production processes, and the customer/user.

In our efforts to establish a scientific basis for the research, the following eight separate and distinct theory areas or established research areas were drawn upon: Theory of Technical Systems, Design Theory, Manufacturing process theories, Statistical theory, Theory of consumer behaviour, Man Machine Interface theory, Organisational (management) theory and Cognition theory. It is interesting, that Total Quality Management did not fit into our mapping, being based upon fragments hold together by an ideal model and 'guruficated' believes. In spite of this, the TQM approach became central in the research.

C.6.2 Research Approach

The first hypothesis[29] of this research is that the opportunities for quality improvement can only be exploited if we acknowledge the central role that product development has in creating quality in all product life phases. If Design for Quality is to be developed accordingly, a first step of this research would therefore *not* be to look at high-level organisational or managerial issues, but to start with investigating the *core* of the design process – synthesis – and then expand from that. The design process is ultimately where the genesis of product quality takes place.

A second hypothesis is that it is possible to identify those product characteristics that 'carry' quality; both quality with respect to the customer (*e.g.*, ease of use, reliability, robustness) and with respect to the company (*e.g.*, robustness towards variability, low quality control costs).

The third hypothesis is that the successful pursuit of the two first hypotheses will create an insight into core processes of designing for quality, which is instrumental in analysing and enhancing already existing approaches and methods for DFQ and for developing new ones.

Finally, this research has operated with hypotheses regarding the choice of scientific viewpoint. It is argued that the Theory of Technical Systems and Design Theory should constitute the theoretical outlook of this research.

The hypothesis and scientific questions were a natural consequence of the identified industrial needs and problems, but also showing where we believe to find insight in the phenomena of quality creation. Finally, they mirror the Department's tradition as 'tool makers'.

The sequence of activities of this research is not so easy to identify. Through a final-year master thesis the area of DFQ was scrutinised, so a certain pattern of necessary insight element was established. The scientific procedure (see Figure C.20) is based on the ideas behind critical rationalism and fallibilism, in which existing models and methods are improved to provide a better description of the empirical reality. This is done through the literature studies, logical structuring, empirical observations, experiments, *etc*. The research has its starting point in a practical problem base where real phenomena in industry and in the literature are analysed. In parallel the discovered problem areas are analysed in the context of the theoretical basis, in which new hypothetical statements on the nature of quality and Design for Quality are also formulated. In order to check their validity and their applicability the solutions and hypotheses are applied to product examples and design cases and presented to designers and researchers.

The main sources of information, inspiration, and experience in this research will briefly be commented upon.

Literature

Because of the large interest in quality in industry and research many publications on this topic exist. It is characteristic that the publications are mainly about

[29] Today we would call these basic assumptions

management aspects, case studies, or isolated methods (*e.g.*, reliability techniques or design of experiments). Disagreement between authors on how quality issues are handled during design still prevails, and coherent theoretical work on the concepts of product quality and Design for Quality is missing. The literature on Design Theory has had a greater degree of clarification to offer, but in the discussion of quality issues focus is still on *technical* functionality and manufacturability and not so much on quality in relation to the customer. Design for Manufacture and Assembly literature has provided inspiring models for this research, especially because of their successful application in practice.

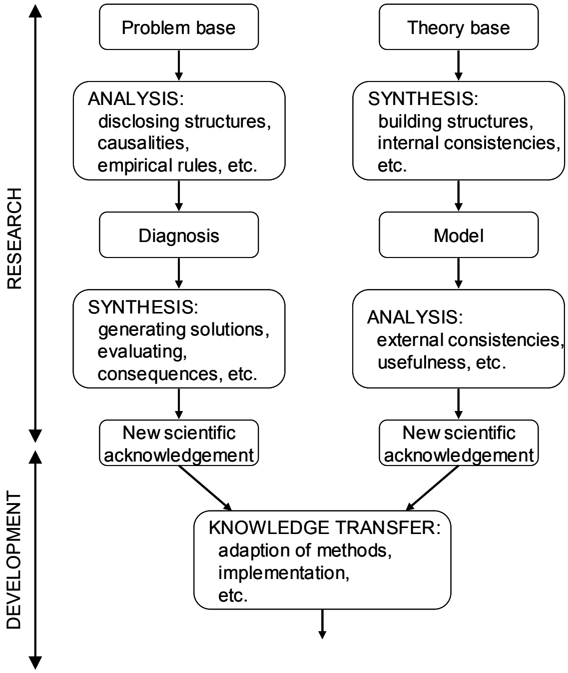


Figure C.20 A method for applied research in which attention is focused on the interplay between theory and practice, after Jørgensen (1992)

Interviews and Case Studies in Industry

In order to obtain an understanding of practical design problems and problem-solving patterns, the research comprised a three-month stay in a Danish manufacturing company as active project participant. Together with shorter case studies in Danish and American industry this stay has been used as a means to detect problems of methodological nature in a designer's working practice. In addition, interviews with quality and product development managers in six American companies have been made. The interviews concentrated on general procedures for product development and specific methods and techniques, especially the Quality Function Deployment method.

Discussions with Colleagues and Professors

Education and research in so-called quality engineering is still limited to single topics such as reliability techniques or problem solving techniques in quality management. However, the project has benefited greatly from discussions with colleagues and professors in a variety of countries on the scope and future perspectives of Design for Quality.

Research in the United States of America

The project included a 10-month stay at the University of Wisconsin-Madison as a doctorate student and guest researcher at the CQPI (Center for Quality and Productivity Improvement). The stay at the CQPI led to valuable insights into the statistical aspects of quality and the methods in this field.

The research process was, on the one hand, arranging the above mentioned activities opportunistically based upon expected insight, on the other hand the 'gradual filling up' of files of results, which became the chapters of the thesis.

C.6.3 Results

The discussion of the inadequacy of current approaches to the management of quality led to the conclusion that incremental improvements of these approaches are not enough; a change of paradigm is needed. This should be based on the fact that:

- Quality is created in product development, not only product quality during use but to a great extent also quality in other product life phases.
- Product quality is actually *built in* during the synthesis process, whereas activities where quality is specified, analysed or verified are 'only' means of indirectly controlling synthesis.

Thus, product design and development hold the key to solving many of the quality-related problems and, perhaps more importantly, to exploit the opportunities for achieving better quality in company activities and on the market.

The main result is a framework (see Figure C.21) that describes the aspects of product development that should be emphasised in order to conduct DFQ. In the process of defining this DFQ framework several other results were obtained. Each of these and the framework will be described in the following sections.

Defining Quality

The discussion of traditional definitions of product quality led to the conclusion that the quality terminology belonging to product development is confusing and poor, which can have serious consequences for how quality is handled during design. Consequently, a new set of quality concepts was introduced, which make a clear distinction between customer perceived quality, termed 'Q'-quality, and quality perceived by stakeholders within the company, termed 'q'-quality. The concepts also distinguish 'true' quality from the quantifications of quality that can be expressed in, say, product properties or manufacturing variability.

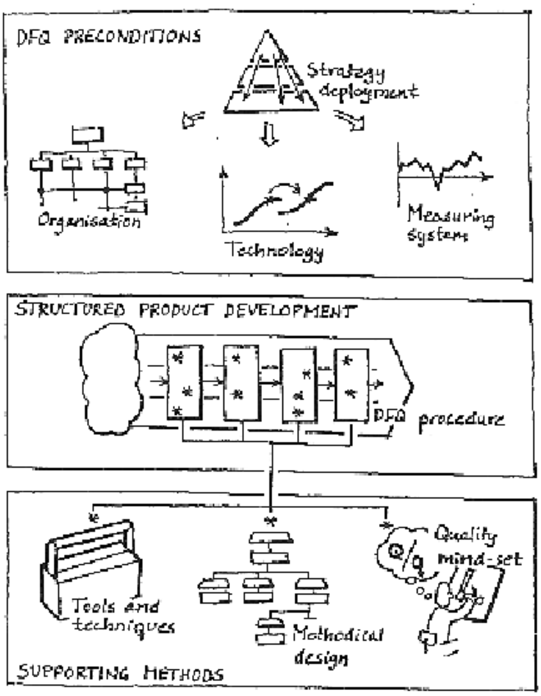


Figure C.21 Preconditions for and main elements of Design for Quality

Defining the Objects of Quality

After clarifying the quality concept, those objects within the product that carry quality were identified and classified. A particular role is played by the organs (also called function carriers), because they carry the product functionality and quality properties, which in turn are perceived by customers as quality. During design, parameter models are the primary means for modelling the relationships between quality properties and the organ solutions.

Clarifying the Relationships Between Quality and the Realisation of the Product

Due to the fact that manufacturing variation influences quality properties, manufacturing processes are often subject to meticulous quality control. However, it was stated that variability evolves in the meeting between the process and the product, and that by making the product robust to variation the need for quality control can be lessened.

Apart from influencing variability in manufacture, the product, and therefore DFQ, also has significant influence on the Q quality and q quality in all other product life phases. Hence, it was suggested that DFQ should be conducted in close contact with important stakeholders, and that their immediate statements about quality (Q and q) should be acknowledged.

Defining the Quality Mind-set of Product Developers

The new quality concepts (Q- and q quality, quality properties, robustness, *etc.*) were incorporated in the context of a 'quality mind-set', which encompasses those parts of the product developers' understanding of quality-related aspects and objectives, which cannot be formulated in traditional specifications.

Developing a DFQ Framework

A new framework for DFQ was presented, which is derived from the analyses of industry's needs, and that is based on the theoretical findings of the research.

The framework has eight 'elements', all of which should be incorporated in a comprehensive implementation of DFQ in practice. The elements cover major activities, from strategic work to the detailed design process, in product development, where quality is specified, built into the product (synthesised), and verified.

The originality of this research lies in:

- The division of the quality concept into Q quality and q quality reflecting the two main categories of stakeholders, external and internal.
- A description of how quality is perceived throughout the product life phases, and how the product itself, due to dispositional mechanisms, can play a prominent role in what is normally regarded as 'quality of sales', 'quality of service', *etc*.
- The establishment of a complete chain of quality concepts, namely Q and q and their relation to product characteristics, their variability and subsequent breakdown of quality specifications.
- A comprehensive framework for Design for Quality which covers the major elements of DFQ, including quality-related activities at all levels of product development from the formulation of product strategies to the everyday activities in design work. The framework makes it possible to map existing elements of DFQ, including procedures, tools and techniques, and to identify voids and inconsistencies in a company's total approach to Design for Quality.
- The identification and description of the quality mind-set and the role it plays in product development. Among other things, the mind-set enables the product developer to constantly evaluate the means he chooses with regard to quality. Hence, the mind-set is very dynamic and flexible in its use.
- The characterisation of robustness and robust design in relation to the new theoretical insight. For instance, robustness can be analysed by means of parameter models derived from knowledge about the product's organs.
- The explanation of the basic significance of technology on product quality, and how technologies can be utilised in DFQ. For instance, mechanical technologies can be used for creating Q quality in the product, and process technologies can improve q quality in manufacture.

C.6.4 Evaluation of Results

Direct verification of design tools and methods is only achieved through successful application to practical design problem. Buur (1990) argues that this method is unrealistic due to the stochastic nature of the design process and the large number of influencing factors that make repetition of experiments virtually impossible. Buur suggests two methods for verifying the validity of design theory:

Logical verification:

- Consistency: there are no internal conflicts between individual elements of the theory.
- Completeness: all relevant phenomena observed previously can be explained or rejected by the theory.
- Well-established and successful methods are in agreement with the theory.
- Cases and specific design problems can be explained by means of the theory.

Verification by acceptance:

- The theory is accepted by a relevant scientific community.
- Models and methods derived from the theory are acceptable to experienced designers.

Logical verification has the draw-back that design theory is basically confirmed by *analysis* of cases and observations. This does not automatically provide a guarantee for the validity of the *synthesis* activities. Verification by acceptance implies a pedagogic problem: the willingness of a designer to accept a statement or method (his need, knowledge, experience), the complexity of the information (how much training is required?), and the pedagogic presentation (Buur 1990). Both types of verification method were applied in this research.

The research marks the first attempt to describe the fundamentals of Design for Quality in relation to all aspects of product development, in particular by way of new concepts and descriptive models. Thus, the primary objectives of the research are fulfilled. The approach to evaluate the results is mirrored in the selection of industrial activities and confrontations, see above, and the described verification concerns in this section.

The verification of the theoretical results constitutes a major obstacle to this work. Verification has been limited mostly to logical reasoning, in that the thesis represents a line of argumentation to show that the proposed concepts and models conform to the theoretical basis (theory of technical systems and properties theory) and are internally consistent.

As for verification by acceptance, the concepts of Q quality, q quality, and positioning and obligatory properties have been implemented by at least two major Danish manufacturing companies in their quality manuals/specifications, quality training programmes, *etc*. This has led to changes in mindset, language and written routines (ISO 9000). The pedagogic formulation of the other scientific results has not yet been pursued to a level where product developers in industry would feel confident about 'experimenting' with their use.

C.6.5 Conclusions About the Research Approach

The research has sought a delicate balance between practical studies in industry and pure academic research, where the latter has been predominant due to the novelty of the research topic, and in order to maintain an ‘unbiased mind’ regarding what is supposed to be so-called ‘best practice’.

It has been the goal to give a comprehensive review of the literature on Design for Quality, to ensure the best scientific basis for the research. However, due to the large number of publications on Total Quality Management, specific quality tools and techniques, and other issues related to DFQ, only a limited selection of the quality tools and techniques has been covered.

Part of the research that did not reach a satisfactory level of results was the discussion of the three last elements of the DFQ framework (tools and techniques, methodical design, and quality mind-set). This may lead to the incorrect assumption that these elements have a low priority in the framework. On the contrary, methodical design and the quality mind-set should play a crucial role in DFQ practice and be subject to further research.

The thesis core research approach problems were

- How can we add to a theory? The expansion of the Theory of Technical Systems and the Theory of Properties is made based upon the perceived need for better concepts and the ‘invention’ of new explaining models. From this point of view it does not matter how formal the empirical studies were.
- What type of phenomenological understanding or insight does the designer need for coping with DFQ? This mind-set and its proper implementation is an open question beyond proof or making probable in this thesis.
- What is formally regarded to be a framework? The DFQ framework seems powerful for communication and similar frameworks have been worked out based upon the idea in this thesis. But actually we do not know the formal identity of such a framework.

C.6.6 Continuation

The proposed quality concepts have been taken up by many important Danish companies as an important step forwards in their DFQ operations and dialogue between design and production. The thesis has led to an articulation of quality as an explicitly treated dimension of design in our teaching at the university and in industrial courses. The research has continued with a focus on ‘soft quality elements’ and the handling of Man–Machine Interfaces and Industrial-Design-related qualities.

C.6.7 References

Mørup M (1993) Design for Quality. Institute of Engineering Design, TU Denmark, PhD thesis

Olesen J (1992) Concurrent development in manufacturing – based on dispositional mechanisms. Institute for Engineering Design, TU of Denmark, PhD thesis
Buur J (1990) Theoretical Approach to Mechatronics Design. Institute for Engineering Design, TU of Denmark, PhD thesis
Jørgensen KA (1992) Paradigms for research work. Instituttet for Produktion, AUC Denmark (lecture notes in Danish)

C.6.8 Reflections from the DRM Perspective

The work of Mørup and Andreasen is on identifying the influencing factors in the design process on product quality, with the intention of building these into a framework for companies to benchmark their design process for quality. They establish, from the literature, product quality as a strategic competitive factor, and the Success Criterion for this project. The central assumptions in their work are that product development plays a central role in creating quality, that it is possible to identify those product characteristics that 'carry' quality, and that successful pursuit of the above two should lead to an insight into the core process for designing in product quality.

The objectives are (i) to define the key concepts in quality in relation to the customer, the manufacturing and other life-cycle systems, (ii) to establish what aspects of a technical system influence quality and how these can be manipulated, (iii) to formulate a general framework for supporting and performing design for quality. The focus is on a Comprehensive DS-I using a literature survey, interviews and case studies in industry. The results are a new definition of quality that differentiates between quality perceived by stakeholders within the company (q) and that perceived by customers (Q), a list of product characteristics that influence these, and an initial framework (Initial PS) that can be used for benchmarking companies' design processes for quality. Logical verification (DS-I) and (after completion of the PhD) verification by acceptance in industry (initial DS-II) are used to evaluate the results.

Using the DRM framework, the project can thus be classified as a Type 2 for the PhD work and Type 5 for the whole project.

Type 2/5: RC (Review) → DS-I (Comp) → PS (Initial) → DS-II (Initial

C.7 Multi-disciplinary Design Problems

Student: Cirrus Shakeri,
PhD dissertation, Worcester Polytechnic Institute, USA, 1998 (Shakeri 1998)
Supervisors: David C. Brown (author) and Mohammad N. Noori

C.7.1 Introduction and Aims of the Research

Introduction

The scope of this research is the multi-disciplinary design of engineered systems. The research aims to develop and explore a new approach *to discovering design methodologies for multi-disciplinary design problems.* The objective is to demonstrate the generation of design methodologies using that approach. A design methodology is a scheme for organising reasoning steps and domain knowledge to construct a solution. It provides both a conceptual framework for organising design knowledge and a strategy for applying that knowledge (Sobolewski 1996). A design methodology can provide the knowledge for decomposing the problem into sub-problems, synthesising partial designs, evaluating and then combining them into more complete partial designs, ordering design tasks by considering proposals from all participants, and discovering and resolving conflicts.

Current multi-disciplinary design methodologies are based on *ad-hoc* strategies for handling the complexities that multiple points-of-view bring to the design process. *Ad-hoc* techniques reduce the complexity but give up the potential advantages of diversity. The common methodologies for multi-disciplinary design are based on compromising between different disciplines rather than collaborating between them. These methodologies do not use a systematic approach and, consequently, are not as efficient and effective as they could be.

In particular, the most common methodologies use sequential design to overcome the complexities of multi-disciplinary design. In sequential design, different disciplines take part in the design process sequentially. Hence, information sharing between different disciplines is limited to the interfaces between disciplines (Levitt *et al.* 1991). As a result, conflicts between disciplines are not discovered until they are very expensive to resolve, because their resolution may need to destroy the partial designs generated by the previous discipline.

Often in sequential design there is a lead discipline, perhaps a designer that makes some of the key decisions and tries to anticipate and remove some of the conflicts. The other disciplines conform to those decisions. However, that may prevent them from producing their best solutions. In a lead-discipline approach a single point-of-view dominates, and therefore constraints from that discipline are favored. This produces a lower quality design product and increases the number of iterations required to reach an answer. Hence, a key goal for better methodologies that was followed in this research, is to ensure *well-integrated* reasoning that provides equal opportunity to all the disciplines involved.

The approach to generating better design methodologies for multi-disciplinary problems was computer *simulation* of the design process. Analysing the behaviour of physical systems in engineering applications by computer simulation using mathematical models has been a powerful tool in engineering. It is particularly useful in situations where real experiments with physical prototypes are not viable. This work extends the idea of using simulation-driven analysis to the area of engineering design research, by applying the concept to the design *process* instead of the design product. In this situation too, 'real experiments' would be extremely costly.

A *computational model* was developed and implemented that simulates the multi-disciplinary design process. By running that simulation system under many different conditions, and analysing the performance, detailed understanding of the design process is gained (Shakeri *et al.* 1998). As for simulations of physical systems, the computational model used is a *simplified* one in which the design activities that are usually carried out by humans are performed by software agents in a slightly simplified manner. We have developed these ideas using the multi-disciplinary domain of robot-arm design (Rivin 1988).

Importance

Using system-developed methodologies allows effective and efficient practices to be used from the *start* of a project instead of being learned from experience. These new methodologies are radically different from the sequential, discipline-based ones. Integration reduces the number of failures and backtracking by facilitating information sharing, thus saving resources and reducing design time. Integration also provides collaboration between different participants that, as a result, enhances the quality of the design.

The agent-based approach that we have adopted for building the computational model allows the incorporation of new technologies systematically and quickly through the addition or deletion of agents (Brown *et al.* 1996; Wooldridge 1997). Thus, new knowledge can be added, and old knowledge removed rapidly. Running a modified system will result in new designing behaviours being simulated, allowing production of new methodologies in response to a change in knowledge. In addition, design processes can be biased toward more environmentally friendly products, by altering the preferences for the alternative design methods that are built into each agent.

In industry, the number of specialists is increasing, while the number of generalists, capable of doing system integration, is decreasing. An increasingly specialised technological environment tends to force designers to concentrate on some disciplines more than others. Also, the knowledge burden on the designer keeps increasing due to more materials and more options (National Science Foundation 1996). Thus it is becoming harder to develop methodologies for the integration of multiple disciplines in design. This research directly attacks this problem.

Computers have mostly been used to support the manipulation and analysis of design product information. This work focuses on the design process, an aspect that has not benefited from computers very much. Simulation of design processes based on a multi-agent paradigm is a new area of research that has a high potential for practical as well as theoretical impact on the design of products. The use of multi-agent systems technology is growing rapidly with the development of Java-based systems and agent access across the world-wide web.

The research is also important because it recognises the importance of incorporating knowledge, judgement and experience. “System integration, many consider, is an ill-structured problem.. No specific rules have to be followed when doing integration.. Experienced designers deal with system integration using judgement and experience. Knowledge-based programming technology offers a methodology to tackle these ill-structured integration and design problems” (Sobolewski 1996).

According to NSF’s report on Research Opportunities in Engineering Design (National Science Foundation 1996), “research areas that will have greatest impact on engineering design over the next 10 years are: Collaborative Design Tools and Techniques, Prescriptive Models/Methods, System Integration Infrastructure/Tools, and Design Information Support Systems”. This work covers all of these areas of research and hence is expected to have a strong impact.

C.7.2 Research Approach

What Was Developed

Part of the goal of the research was to produce an ‘approach’ (*i.e.*, to producing methodologies). First we will describe in more detail the approach that was developed during the research. Then we will describe the way the research progressed: *i.e.*, the approach to the approach!

This work proposed a new approach to the problem of producing better design methodologies for multi-disciplinary design based on the tight integration of different disciplines. The discipline-sequential approach, while poor, is quite simple. Its flaws are well known and have been part of the motivation for concurrent engineering (Brown *et al.* 1996).

However, integration tends to make the design process more complicated. To overcome this complexity, a computer system was developed based on a multi-agent systems paradigm in order to automate the simulation of the design process. The system also allows multiple design problems to be simulated in a small amount of time.

The system simulates examples of multi-disciplinary design processes while applying integration principles to the problem. The principles were developed from an examination of the literature. These principles include common design knowledge representation schemes and common communication mechanisms; design knowledge sharing among participants; cooperative problem-solving strategies among participants; simultaneous design process where possible; and mechanisms for conflict discovery and resolution. The principles are embodied in the system both in its architecture and at run-time.

The large chunks of discipline-specific knowledge are broken into small pieces, each being represented in the design system by an agent. Agent activation is triggered in an opportunistic manner and is unaffected by discipline boundaries. Agents might participate sequentially or in parallel. This leads to well-mixed use of knowledge from different disciplines, and the possibility of parallel design activity for tighter integration and better efficiency.

The multi-agent design system is run with a very large number of different design problems. This is done by systematically varying the individual design requirements across their ranges in order to cover the space of requirements. Hundreds of design problems are presented to the system. Some problems do not lead to a successful design.

For each problem the traces of the agent activations (*i.e.*, knowledge use) during the course of the design process are recorded. The many recorded traces consist of orderly patterns of different design actions that have led to a design solution.

Candidate design methodologies are extracted by generalising the patterns in the recorded design traces using clustering techniques. This both groups and identifies common aspects of related traces. The best clusters are the most 'convincing' methodologies. For each cluster identified, the commonalities in the Requirements are identified. This allows combinations of Requirements to be recognised as being most appropriately handled by a particular methodology.

Research Questions and Hypotheses

The *main hypothesis* for this research was that methodologies could be generated by using a computer to build up 'experience' by simulating design activity.

The question was 'how?'. Clearly doing it with real people was impractical. This led to the idea of simulation. The need for integration led to the notion that any knowledge should have the potential to be applied opportunistically at any time, and that the knowledge should be split into pieces. These ideas, and the analogy with the design teams used to support Concurrent Engineering, made us decide to use a multi-agent design system.

The multi-agent approach intuitively captures the concept of deep, modular expertise that is at the heart of knowledge-based design (Lander 1997). A multi-agent system is composed of multiple interacting agents, where each agent is a coarse-grained computational system. Agents are used as an abstraction tool for conceptualising, designing, and implementing the knowledge-based design approach. An agent is a self-contained problem-solving system capable of autonomous, reactive, pro-active, social behaviour. It is a powerful abstraction tool for managing the complexity of software systems (Wooldridge 1997). Thus, the multi-agent paradigm not only matched the problem, but also provided some Software Engineering advantages.

We started by investigating 'methods', 'methodologies' and 'integration', as well as studying the literature on multi-disciplinary design. The latter confirmed the belief that there was a need for a systematic way of building good methodologies, and that many of those currently in place were *ad hoc*.

Early in the research we picked a domain in which to work: one that was well known to the student, had a clear multi-disciplinary flavour, was of a manageable size without being trivial, and appeared to have no strong, existing methodology of the type we were seeking. Robot-arm design seemed to be perfect. It demonstrated well the tendency for researchers from each discipline to write about the design problem as if their discipline's contribution were dominant.

We started by implementing a 'base-level' working system that used a non-integrated approach to robot-arm design. This tested our understanding of the relevant knowledge and methods used. Early prototype implementation is an important research technique that enhances domain understanding, acts as a catalyst for learning programming and system development techniques, and also forces precise definitions of concepts.

The base-level system also provided some feedback about where errors occurred during designing and where the knowledge might be decomposed into pieces. The choice of Java for the implementation allowed portability between systems and provided the ability to effectively handle agents in parallel.

Next the framework for a multi-agent design system, to be called Robot Designer (RD), was developed. The question of how to split the knowledge from each discipline into pieces was addressed, and the resulting pieces were encoded as agents, and added to the framework. Decisions were made as to what needed to be stored as a record of every agent's action, such that these traces might be able to form suitable methodologies. The traces were accumulated, but not the designs.

A great deal of time was spent on developing a failure-handling system for RD so that the constraint failures could be recovered from, while allowing parallel paths through the agents to be recorded correctly.

An important issue at this time was the relationship between design quality and methodology quality. Clearly one would like the methodologies produced to lead to high-quality designs. The traces that actually lead to designs must include no failing constraints, and hence at least possess a certain level of quality. But they may not all be of equal quality. In addition, our simulation of the design doesn't include *all* of the design knowledge available, and hence may not lead to the best possible designs – in fact, it would be foolish to pretend that one could make a perfect simulation of the design activity in a complex multidisciplinary situation. However, as we expected these methodologies to be followed principally by people, they only need to act as guidance, and the human designer should be able to ensure the quality of the result.

The compound hypothesis was that less precise knowledge might lead to adequate designs, and that adequate designs were associated with traces that might form a methodology capable of guiding the production of high-quality designs. This hypothesis was never *fully* explored.

A key question to be addressed at this point was how to exercise RD such that the whole design space was explored. This was important because we wanted to generate methodologies for all types of designs in the space. The approach taken was to drive the system with as many different sets of requirements as possible, such that the whole design space was explored. The hypothesis was that all reachable regions of the design space could be found by systematically varying the requirements in order to adequately scan the requirements space. By experimenting with the degree of change between requirements we were able to convince ourselves that this approach was successful.

Another issue that this raised, and that we spent a lot of time investigating, was the relationship between the requirements space, the design space and the trace space. These relationships were explored and revealed using graphical representations.

The next question to be addressed was how to form traces into methodologies. Several techniques were considered until a clustering algorithm was tried that gave appropriate results. While other methods might be appropriate we focused on just one. Once clusters were identified they were re-expressed as rules.

A final issue that was examined was how to best express the requirements that correspond to a particular methodology.

The main hypothesis, that methodologies could be generated using the computer by simulating enough design experience, was demonstrated by this research.

C.7.3 Results and their Evaluation

A knowledge-based model of design was adopted in order to implement the proposed strategies for integration. To implement the proposed model a knowledge-based multi-agent design system, RD, was developed that simulates the design process.

Both the general multi-agent design system architecture, and the RD system developed are results of the research. The approach to breaking knowledge up into pieces such that they can be incorporated into agents is also a result of this research.

The Java-based computer program called RD (Robot Designer) was implemented for parametric design of two degrees of freedom (2-DOF) planar robot arm. We used RD to solve a set of 960 design projects. Figure C.22 shows how many projects followed a specific trace. The promising result is that many projects followed similar traces. The total number of possible traces is the product of the number of design approaches of all the designer agents. For the experiments shown in Figure C.22 (Shakeri and Brown 2004), the total number of possible traces is 2304. However, despite all those possible traces only 84 were followed to generate successful designs, *i.e.*, less than 4%.

The low percentage of successful, relative to 'possible', traces indicates that for each group of projects that followed a particular trace there is a unique combination of approaches leading to successful designs. Hence, there is a high chance that if similar projects follow the same trace they will succeed in generating a successful design. As a result, the path followed by those projects can lead us to formulating a design methodology for the projects that followed that trace as well as projects that are similar.

The traces in the set of successful traces that are close enough can be clustered together to form a generalised trace. A generalised trace covers all the projects that followed each of the traces incorporated in the generalised trace. Design methodologies are formulated by extracting the correlation between a generalised trace and the design projects (*i.e.*, the sets of requirements) that produced that trace. The sample design methodology that is shown below is the English translation of the correlation between design projects and the corresponding traces.

Methodology:

- choose the location of the base of the robot: *'left or below midway of the workspace length'*
- choose the material: *'steel stainless AISI 302 annealed'*
- select the shape of the cross section of the link: *'hollow round'*
- choose the structural safety factor: *'3'*
- **do** the design and proceed to the next step
- choose the link 2 to link 1 length ratio: *'0.5'*
- **do** the design and proceed to the next step
- pick the configuration of the arm: *'left-handed'*
- select the ratio of the cross section dimension of the link to minimum required by stress analysis: *'4'—if it fails select '3'*
- **do** the design and proceed to the next step
- find the accessible region: *use Equation 2-4*
- find the deflection of the tip: *use Equation 2-14*
- choose the type of controller: *'PD'*
- **do** the design and finish the process.

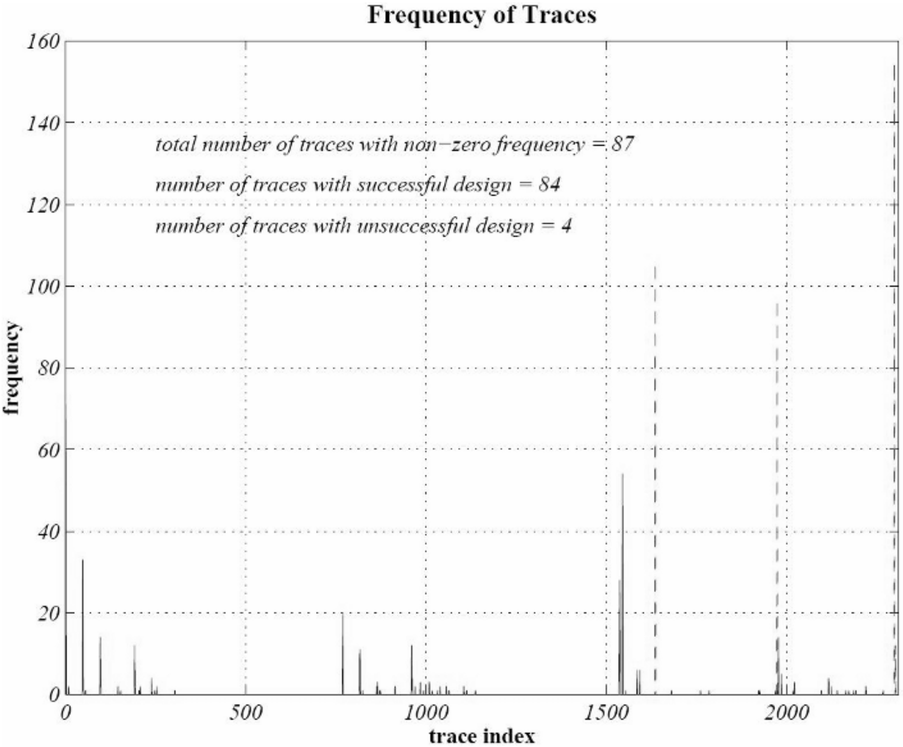


Figure C.22 The plot shows how many projects followed a specific trace (This figure appeared in Shakeri and Brown 2004)

C.7.4 Conclusions About the Research Approach

The main characteristic of this research approach was the use of a computational model to simulate design activity, that, due to its size and complexity, would otherwise be impossible due to the time and cooperation required. The main positive feature of this approach is therefore that it allows the normally ‘impossible’ to be handled. Other positive features should be self-evident from what has been presented so far.

As with all research, there are a variety of trade-offs and assumptions incorporated. Some of the assumptions need to be verified more rigorously.

The largest issue is that for every domain the approach requires collection of a very large body of knowledge – recognised as a very difficult task – and the construction and testing of a large and complex software system. Once that’s done the methodologies should follow fairly easily, assuming that other domains behave in a similar way to our test domain.

However, there is a very large investment in time and effort required in order to start getting results. This approach will only be viable if the methodologies generated are extensively used and if they provide large gains in quality or cost/time savings. Of course, we have limited evidence that such gains are in fact provided.

Also, a big assumption is that both the simplified model of designing (*e.g.*, with respect to failures) and the simplified knowledge that are embodied in the system, are not so simplified that the traces generated are atypical. Also assumed is that all traces that lead to successful designs should be included in the process that leads to methodologies. This is without regard to the quality of the designs that these traces produced: all we know is that they are ‘correct’. This may affect the quality of designs that can be achieved by following the methodologies.

Clearly this leads to a trade-off between the effort required to build the system (more complexity and authenticity leads to more system building effort) and the quality of designs that can be achieved by following the methodologies. This trade-off needs to be explored.

The interesting issue of the relationship between the clustering of designs, the clustering of requirements, and the clustering of traces (to form methodologies) still needs more exploration. The approach of systematically scanning across the requirements space needs to be investigated further to test for sensitivity to different domains and problem areas. Perhaps an adaptive scanning approach could be tried?

The final weakness is that it is very hard to test the methodologies that are generated under realistic situations, as this would require extensive use by multi-disciplinary design teams with many design problems. As a consequence, it is hard to establish whether the designs produced by using the methodologies are indeed of high quality.

C.7.5 Continuation of Project

The potential applications of this research are in multi-disciplinary design situations, such as those that occur throughout the automotive industries, where large gains can be achieved with integrated methodologies. In addition, current methodologies can be analysed for flaws and bottlenecks, and necessary refinements made. New methodologies can be customised so that they are biased toward specific objectives such as manufacturability or being environmentally friendly. By applying this approach the response time for the incorporation of new technologies in design processes should be reduced. Methodologies can be refined as soon as a change occurs in the market or in the organisation of the company.

At this time no additional work has been carried out on this project since the Ph.D. (Shakeri 1998) was completed. As with any large, student-driven software project, transferability is an issue. There would be a very steep and significant learning curve associated with taking over the complex Java code that was designed. Another serious issue is that finding a student with the right combination of CS and ME skills is very rare, and would take some time to develop. We would very much like to experiment with confirming the quality and utility of the methodologies generated, especially for new multi-disciplinary problems. Automobile, Aircraft, Computer and Mechatronics design should be fruitful areas to investigate. The approach that we have proposed has been developed based on parametric design problems. Applicability of the approach to other types of problems needs to be investigated. This research has proven that the following hypothesis is true: Computers can provide us with better ways of doing design by discovering design methodologies that integrate multiple disciplines into the design process. It has been shown that it is possible to use computers to simulate the design process, and then analyse the results of the simulation to synthesise design methodologies that have superior features. This forms the basis of a new approach to the study of multi-disciplinary design processes.

C.7.6 References

Brown DC, Lander SE, Petrie CJ (1996) The application of multi-agent systems to concurrent engineering. In: Editorial Concurrent Engineering: Research and Applications 4(1): 2–5

Lander S (1997) Issues in multi-agent design systems. IEEE Expert: Intelligent Systems and their Applications, 12(2):18–26

Levitt RE, Jin Y, Dym CL (1991) Knowledge-based support for management of concurrent, multidisciplinary design. AI EDAM 5(2):77–95

National Science Foundation (1996) Research opportunities in engineering design. NSF Workshop Final Report, April 1996

Rivin EI (1988) Mechanical design of robots. McGraw-Hill Book Company, TX

Shakeri C (1998) Discovery of design methodologies for the integration of multi-disciplinary design problems. Mechanical Engineering Dept., WPI, Worcester, MA, PhD thesis,

Shakeri C, Brown DC (2004) Constructing design methodologies using multi-agent systems. AI EDAM 18(2): 115–135

Shakeri C, Brown DC, Noori MN (1998) Discovering methodologies for integrated product design. In: 1998 AI and Manufacturing Workshop: State of the Art and State of Practice, Albuquerque, NM,: pp 169–176

Sobolewski M (1996) Multiagent knowledge-based environment for concurrent engineering applications. Concurrent Engineering: Research and Applications 4(1):89–97

Wooldridge M (1997) Agent-based software engineering. IEEE Transactions on Software Engineering, 144(1):26–37.

C.7.7 Reflections from the DRM Perspective

The work of Shakeri *et al.* is to develop a computational approach to the generation of successful design methodologies for multi-disciplinary design, with the intention of reusing these system-developed methodologies in appropriate situations, *i.e.*, for situations in which the methodologies were found to have been particularly successful. The assumptions are that using systematic approaches will lead to effective and efficient design processes, and that common methodologies for multi-disciplinary design do not use a systematic approach and are not as effective and efficient as they could be. The most common methodology used today is sequential design that, the authors argue, is simple, but does little information sharing, which leads to a large amount of backtracking in the design process, a large number of failures and little collaboration. These lead to long time to design, over-use of resources and poor design quality, respectively. Their overall Success Criteria, we interpret therefore as high design quality, low resource use and low design time, and the Measurable Success Criteria as high level of information sharing, low number of failures, low number of backtrackings. Furthermore, the methodology should be simple.

The authors start with a review-based DS-I using the literature search and analysis. They then developed a computer simulation of design processes to identify sets of activities that lead to successful designs (that satisfy all constraints), *i.e.*, successful design methodologies. The software uses software agents as 'designers' who have specialist domain knowledge and work together using integration principles taken from the literature. The activities of the agents are traced, and patterns identified using clustering techniques in order to extract successful methodologies (Comprehensive PS). A central assumption in this work was that the simulated processes represent actual design processes. From an AI point of view, an interesting feature of this work was to enable the computer to 'discover' design methodologies from test cases.

The work therefore involved a review-based DS-I, which was followed by a comprehensive support development. Using the DRM framework, the project can thus be classified as a Type 3.

Type 3: RC (Review) → DS-I (Review) → PS (Comp)

C.8 Design for Reliability in Mechanical Systems

Student: John Stephenson
PhD dissertation, Cambridge University, UK, 1995 (Stephenson 1995)
Supervisor: Ken Wallace (author)

C.8.1 Introduction and Aim of Research

In order to maximise their competitive positions and ensure their future profitability industrial organisations aim to develop the best products in the shortest time and at the lowest cost. Defining the 'best' product is not easy as a complex blend of many characteristics must be achieved and delivered at the right time. Every new product must fulfil its intended functions safely, continue to do so over its intended life, and at a cost that customers are prepared to pay. These external characteristics can be grouped under the broad headings of performance, reliability and economy.

Human processes are controlled by the decisions taken, and all decisions depend on forecasts and the evaluation criteria used. The decisions made during the design stage influence all the subsequent stages in the product development process. During the design process numerous forecasts are needed to answer questions such as: How easy will the proposed product be to manufacture and assemble? How well will it perform in operation? How reliable will it be? What will be the selling price? What will happen to it at the end of its useful life? A great deal of research effort goes into developing methods to improve our ability to forecast.

When starting a design research project one of the challenges is to isolate one aspect on which to focus the study. This is more difficult than in the natural sciences as the influences are so complex and interlinked. It is important to define an area of interest that is sufficiently self-contained that it can be studied in a reasonable period of time and useful conclusions drawn.

The research project described here was located in the product area and addressed the main question: For a mechanical product, how does one 'design in reliability' as early as possible in the design process? It did not address performance or economy, though, of course, these are linked. It concentrated on the technological aspects of the product with the intention of developing an understanding of the issues and hence a theory on which a specific design for reliability (DFR) method could be based and tested.

The research was based on the theory of the properties of technical systems developed by Aguirre during his PhD research project at Cambridge in the 1980s (Aguirre 1990). He identified basic properties after creating and analysing a database of around 3500 guidelines extracted from the literature. The properties can be classed as being either 'external' or 'internal'. The external properties, which have already been mentioned, are what the customer or user sees: performance, reliability and economy. The internal properties were identified as simplicity, clarity and unity. Design engineers do not manipulate these properties directly. Aguirre argued that designers should aim to maximise the internal properties of simplicity, clarity and unity. This raises a number of questions: What are the links between the internal and external properties? How does one measure these internal properties? What are the guidelines for achieving high levels of these properties?

The objectives of this research were to study specific cases of good and bad reliability, develop a theory of mechanical reliability, test this theory, and produce a method to support mechanical designers. The crux of the research was to establish the relationship between the internal properties and the external property reliability.

How can these research objectives be achieved and tested in a rigorous way in a PhD project? This was achieved by undertaking this research in collaboration with a manufacturer of earth-moving equipment. Extensive empirical data were gathered from case studies of mechanical systems on the company's backhoe loader (Figure C.23). The company operates in a very competitive market and had identified that increasing the reliability of its products would give them a stronger marketing position.

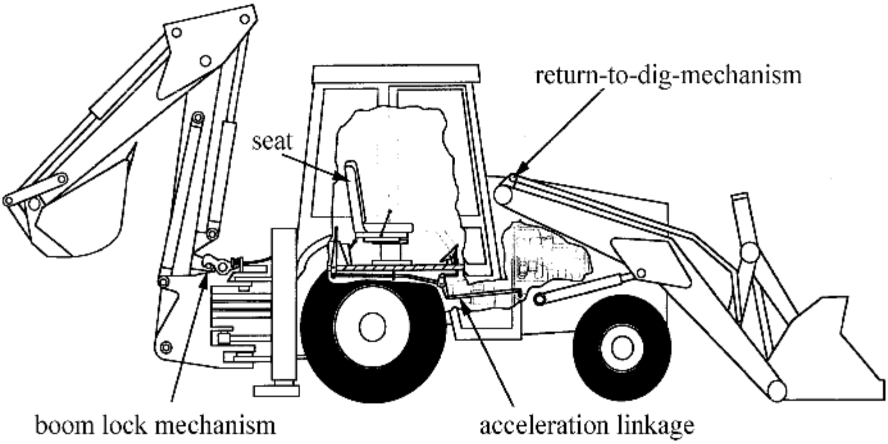


Figure C.23 Backhoe loader (case study mechanisms are identified)

The company operates a warranty scheme whereby faults are rectified at the company's expense within the warranty period. If a particular piece of equipment is returned regularly for a particular fault, the company responds by modifying the design to overcome the fault. The company held excellent records of all its warranty claims and all its design changes — but no attempt had been made to relate the two, that is, identify in a systematic manner the effect that specific design changes had on reliability and, more importantly, whether there were design guidelines that could be extracted to help designers 'design in reliability' early in the design process and prevent the necessity for corrective action.

The need to improve reliability, the availability of a large quantity of data and a clearly defined problem area provided a clear specification for this research project.

C.8.2 Research Approach

One of the challenges in design research is validating a new method or technique, particularly within the time frame of a PhD research project. The true test of a theory or method is its ability to predict, *i.e.*, make accurate forecasts of future results. This is referred to as *predictive testing*. The difficulty arises because of the long timescales involved in designing, manufacturing and testing new mechanical engineering products. By the time that a method has been developed, there is seldom time to undertake predictive testing on a real design project.

To overcome this problem the idea of a *historical testing* was introduced. A series of four case studies, described in more detail later, was planned. The first was to investigated the practicability of gathering, storing and analysing the data with a view to identifying the issues involved. The second investigated whether the results and ideas obtained from the first case study could be repeated and was used to formulate the preliminary DFR method. The third and fourth were used to test and develop the DFR method using historical testing. By chance, two opportunities arose during the project to test the method in a predictive mode.

The idea behind historical testing was to go back to the time to when a particular system was being designed, to pick up the *design data* at that point and then apply the new DFR method (which was not available when the product was being designed) to the data to predict areas where reliability problems would be expected to occur. For this to be a “true” experiment, it was essential not to look at the reliability records of the system being studied until after the predictions had been made using the method.

The research followed three main stages:

Stage 1: A literature review and survey of design principles

Stage 2: Developing the theory and method (Case histories 1 and 2)

Stage 3: Testing the method (Case histories 3 and 4)

Stage 1: Literature Review and Survey of Design Principles

A study of the literature was undertaken to determine the extent of the existing work on DFR and to identify relevant theories, principles, guidelines and methods.

The literature review showed that reliability is determined by three phases of the product life: design, production and exploitation. Design has the potential for the greatest effect on reliability. The following reliability approaches were identified:

- reliability apportionment;
- design layout decisions;
- load strength theory;
- robust design;
- reliability prediction;
- failure analysis;
- reliability testing.

Within each of these approaches a number of specific methods are available. For example, nine methods are available for failure analysis including: Failure Modes and Effects and Criticality Analysis (FMECA), see BS 5760; and Hazard and Operability (HAZOP) analysis. Despite the many techniques available, it was concluded that there was no coherent reliability approach available that can support decisions made during conceptual and embodiment design. Although some design layout principles are available, in their current form they do not aid the development of component shapes to improve reliability at the embodiment design stage. It was also concluded that the most relevant concepts for conceptual and embodiment design are generally applicable design principles. Design principles encapsulate 'best practice' and so a theory of reliability based on them suggests that a 'good' design will be a reliable design. Considerable effort therefore went into identifying design principles, developing a theory for reliability based on them, and producing a method based on this theory.

The most important contributions in this area were identified as Aguirre (1990); Pahl and Beitz (1996); French (1994); Suh (1990); Taguchi (1993); Hubka and Eder (1988) and Dimarogonas, Redtenbacher, and Reuleaux (Dimarogonas 1993). Of these, only Aguirre and Suh adopt approaches that are truly based on principles or axioms and thus provide a sound basis for a theory of reliability. French, along with others, emphasises the importance of interfaces stating: "It is important, and often a great help to pay full attention to the design of interfaces, particularly in the early design stages, and also because the constraints associated with them point the way to go" (French 1992).

The principles identified by Aguirre, based on internal and external properties, and the importance of interfaces underpinned the theoretical aspects of this research project. Empirical data was gathered from four case studies and in each case study, around 2000 machine records were consulted.

Stage 2: Developing the Theory and Method (Case Histories 1 and 2)

Case study 1 — Return-to-Dig Mechanism (See Figure C.23)

The return-to-dig mechanism is part of the sub-system that controls the positioning of the front bucket. Under the control of the positioning system, the mechanism performs two functions: bucket levelling and return-to-dig. The bucket-levelling function maintains a constant angle of the bucket at the loader arms rise and fall; the return-to-dig function uses some of the same mechanism as the bucket levelling function and speeds the digging cycle by automating the positioning of the bucket between dropping off a load and picking up the next.

Five separate configurations were studied and these provided examples of the differences in reliability between configurations with shared and separate function carriers. Part of a typical configuration history is shown in Figure C.24.

This first case study was used to help identify those design factors that influence reliability. The evolution of the configuration showed that clarity was particularly important, with simplicity having a secondary influence.

Failure probabilities were calculated from the warranty data and presented in a number of forms. A typical representation is shown for three return-to-dig configurations in Figure C.25.

Case study 2 — Boom Lock Mechanism (See Figure C.23)

The boom lock mechanism locks the boom and prevents it swinging around while the backhoe is, for example, being driven on the road. Eleven boom lock configurations were studied.

This case history was used to confirm the relevance and relative importance of clarity and simplicity in reliable configuration. It was during this case study that a preliminary DFR method for assessing clarity and simplicity was formulated based on a component/interface model. This model will be described later. During the research project a new boom lock configuration was released and this provided a valuable opportunity to test the DFR method as a predictive tool.

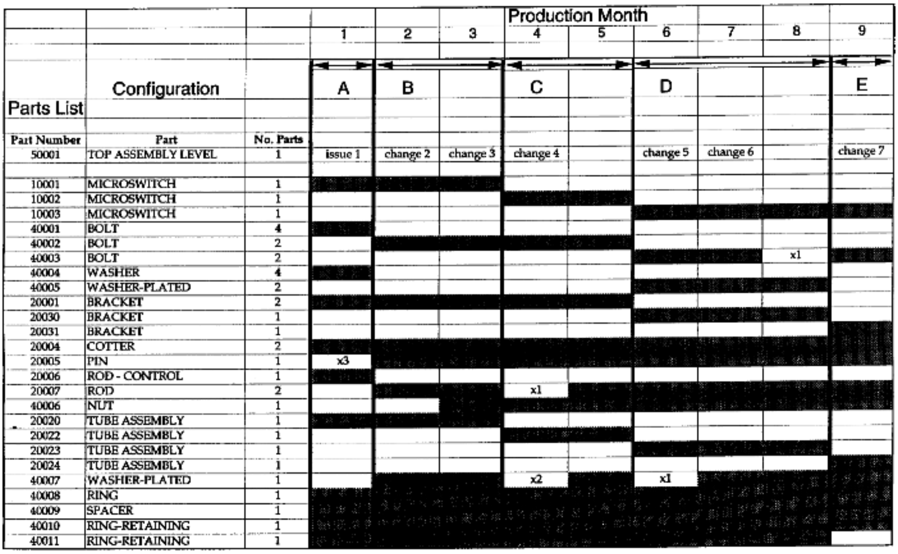


Figure C.24 Part of a typical configuration history

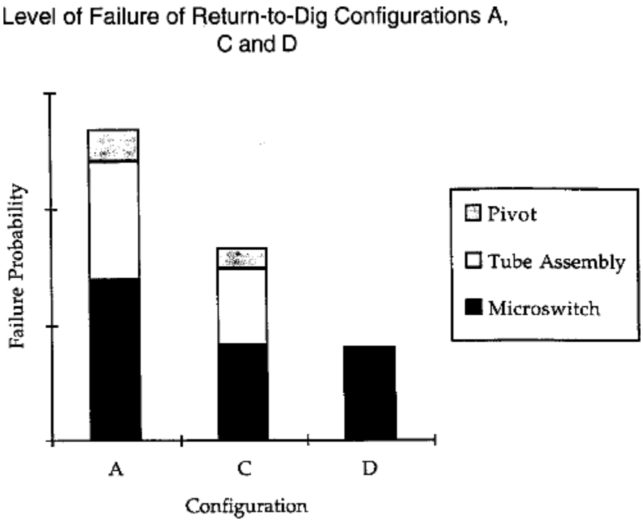


Figure C.25 Level of failure of return-to-dig configurations A, C and D

Stage 3: Testing the Method (Case Histories 3 and 4)

Case Study 3 — Seat Mechanism

This mechanism provides forwards and backwards adjustment of the driver's seat as well as allowing the seat to rotate through 180 degrees between the driving position and the backhoe operating position. These functions are performed by a complex slide and turn mechanism that, for obvious reasons, must be extremely robust and reliable. Seven seat configurations were analysed.

This case history was used to test the validity of the DFR method as well as deepening the understanding of the influence the internal properties had on reliability. Special attention was given to the third internal property, unity.

For each configuration, a component/interface model was produced and each assessed for potential areas of unreliability using the DFR method. Only *after* these analyses had been completed, were the data gathered on the actual reliability of the different seat configurations. This represents a way of validating the DFR method using historical testing. Again, during this investigation a new seat mechanism configuration went into production, again allowing the DFR method to be used as a predictive tool.

The different seat mechanism configurations show different levels of function sharing, particularly within the slide rails — which were the components to fail most frequently. The case history therefore provided excellent examples of successful and unsuccessful function sharing. In addition, the importance of unity was established in the sense of providing sufficient strength for the components.

Case Study 4 — Accelerator Linkage Mechanism

The accelerator linkage mechanism has two inputs: one is the accelerator pedal for determining the speed of the backhoe when being driven; the other is the lever used when operating the loader arm. Four configurations were studied.

As with the third case study, each configuration was analysed using the latest version of the DFR method and potential failure areas identified *before* the actual failure data was obtained. Again this provided a way of using historical data to validate the DFR method.

C.8.3 Results

One of the key results is the DFR method developed. It will be described by applying it to part of a cable and lever mechanism from a boom lock configuration shown in Figure C.26. To begin with the component/interface diagram is created as shown in Figure C.27. Then, four steps are undertaken as described below.

Step 1: Simplicity

Simplicity is assessed by counting the number of components and interfaces in the model. This approach is not complete as simplicity relates to other concepts such as ease of manufacture and maintenance. However, this assessment is easy to perform and provides a rough comparison of configurations. In most cases only significant differences in the overall numbers of components and interfaces between configurations impact on reliability, provided the levels of clarity and unity are maintained. For the model shown in Figure C.27, there are 6 components and 7 interfaces. As the cable housing and the handle base are both fixed to the frame, these are considered to be one 'component' and are thus joined together in the model.

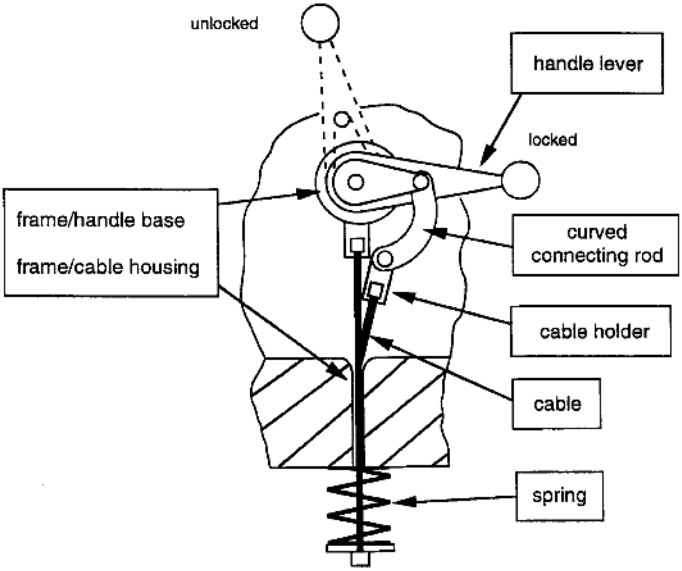


Figure C.26 Simplified lever and cable from one of the boom lock mechanisms

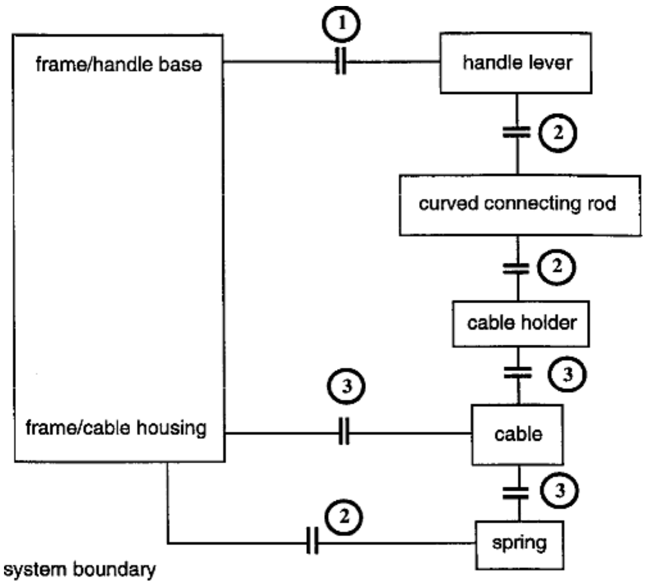


Figure C.27 Component/interface model

Step 2: Clarity

Clarity is assessed by using a ranking system to indicate interfaces with different levels of clarity. The ranking system is equivalent to a 'penalty' point system, where a '1' is the clearest type of interface and a '3' is the least clear. The assessment of clarity is based on how the forces performing each function are transferred at an 'active' interface. This is a crucial and difficult aspect of the method and is described in detail in Stephenson (1995). Of the 7 interfaces shown in the model in Figure C.27, 1 interface is ranked as Clarity 1, 3 as Clarity 2, and 3 as Clarity 3.

Step 3: Unity

Unity in this case is assessed by estimating the peak loads and checking to see if the components are strong enough to fulfil their functions. If there are components that might fail under load, then the appropriate component box in the model is shaded. In the lever and cable mechanism shown in Figure C.26, quick calculations show that each component is strong enough.

Step 4: Use of Results

Clarity is the main issue in this example as simplicity is only relevant when several design solutions are being compared. Unity is not an issue as no weak components were identified. From the assessment of clarity, three of the interfaces were ranked Clarity 3 and these have the greatest potential to fail and need to be reviewed carefully. After reviewing the design configuration carefully, the clarity at these interfaces can be improved by:

- Increasing the force of the positioning function (cable tension) by providing a suitable combination of a longer lever, a stiffer spring and a pre-loaded spring.
- Reducing the resistance force by pulling the cable coaxially with its housing and changing the cable forces to reduce friction.

C.8.4 Evaluation of the Results

The new DFR method described above requires further research and development. During the project, a number of instances were identified where the method did *not* work and tentative reasons were suggested for these instances. However, overall the results are encouraging.

The DFR method has been validated in two ways; historical and predictive testing. A total of 27 configurations was analysed. Out of these, 6 were used to test the method historically and 2 predictively. For the historical cases, 14 failure-mode predictions were made and 9 turned out to be correct; and for the predictive cases, 4 predictions were made and 2 were correct. So out of a total of 18 predictions, 11 (61%) were correct suggesting the value of the method. It is important to note that in each of the 8 configurations analysed, the *major* failure mode in service was predicted.

Failure can occur at an interface or in a component in a mechanism. Simplicity thus relates to the number of *potential* failure areas (components and interfaces). However, simplicity does not relate to how likely the potential failures are to occur. This is because it does not explain how the failure mode will occur. Therefore simplicity has an indirect relationship with reliability.

Clarity relates to how functions are performed, and so how they can fail. Thus clarity has a direct, causal relationship with reliability and can be most easily assessed at active interfaces.

Unity relates to the strength of components that includes the effects of stress and materials, *e.g.*, rupture, wear, fatigue and corrosion. Like clarity, unity has a direct relationship with failure by being able to explain failures.

It is always interesting when research produces a counter-intuitive conclusion. This occurred in this research. When designers were asked what characteristic they thought had the greatest influence on mechanical reliability, most said it was simplicity. In the case histories analysed, the majority of failure modes occurred due to a lack of clarity. This has resulted in the importance of simplicity in achieving good reliability being challenged. Simplicity, as explained above, is seen as having an indirect relationship with reliability, whereas clarity is seen as having a direct relationship. Complex configurations that have clarity will be more reliable than simple configurations that lack clarity.

The clear and important conclusion from this research is that whilst reliability can be achieved by a product that is simple, it cannot be reliable without being clear.

C.8.5 Conclusions About the Research Approach

A research project sets out to answer a number of questions. It is therefore reasonable to ask at the end of a project whether or not the starting questions have been answered. In the case of this project the main starting question was: For a mechanical product, how does one ‘design in reliability’ as early as possible in the design process? This depended on answering many subsidiary questions such as: What are the links between the internal properties and reliability? How does one measure these internal properties? From what has been described above, it can be seen that substantial progress has been made in answering these questions, though many questions remain unanswered.

It is also important to ask if the research method was rigorous. In this case it was based on a valid, if tentative, theory of the properties of mechanical systems. The research was underpinned by the gathering of large quantities of empirical data from four case studies - each case study involving reference to around 2000 machine records. During these case studies, a total of 27 different configurations were studied in detail. A DFR method was developed and tested both historically and predictively. Of particular importance was the development and use of historical testing as a valuable design research method.

Research always raises more questions than it answers. That was the case in this research project and there is much work to be done. The method is in a preliminary form. A number of instances were found where the method did not work. These provided valuable insights and highlight the need for further testing and development.

C.8.6 Continuation

There are five areas where further research is needed:

- Developing further the understanding of clarity.
- Broadening the DFR theory beyond its current area of mechanisms.
- Developing a complete theory of the properties of mechanical systems to include the links between the internal properties of simplicity, clarity and unity and the external properties of performance, reliability and economy.
- Extracting guidelines that will help designers achieve high levels of simplicity, clarity and unity.
- Integrating the DFR method into a CAD system.

C.8.7 References

Aguirre GJ (1990) Evaluation of technical systems at the design stage. University of Cambridge, Cambridge, UK, PhD thesis,

Dimarogonas AD (1993) On the axiomatic foundation of design. In: Knight TK and Stauffer LA (eds) Design Theory and Methodology (DTM'93,) Albuquerque, New Mexico, ASME

French MJ (1992) Design principles applied to structural functions of machine components. J Eng. Des. 3(3):229-241

French MJ (1994) An annotated list of design principles. In: Proceedings of the Institution of Mechanical Engineers 208:229-234

Hubka V, Eder WE (1988) Theory of technical systems. Springer-Verlag, Berlin

Pahl G, Beitz W (1996) Engineering design - A systematic approach. 2nd edn, Springer-Verlag, London

Stephenson J (1995) Design for reliability in mechanical systems. PhD thesis, University of Cambridge, Cambridge, UK

Suh NP (1990) Principles of design. Oxford University Press, Oxford

Taguchi G (1993) Robust technology development. Mechanical Engineering 115(3):60–62

C.8.8 Reflections from the DRM Perspective

The work of Stephenson and Wallace is on design for reliability in mechanical systems, with the specific objective of supporting analysis of reliability potential of solutions during the early embodiment stages of design. Analysis of work from the literature as well as data from industry and its immediate problems with product reliability led to the decision of taking product reliability as the Success Criterion, and reduction of frequency of failure of products, as its operationalisation. A central assumption in this work was that being able to predict potential reliability of a design would support development of more reliable products.

The primary focus of this work is a Comprehensive DS-I using historical testing, which uses data gathered earlier in the company for developing and testing hypotheses. The research led to a method for estimating potential reliability by analysing the product characteristics (comprehensive PS). While part of the historical data was used in developing the understanding in DS-I, another part of the data was used to test the method developed (initial DS-II).

The work therefore involved a Comprehensive DS-I, which was followed by a comprehensive support development, and subsequent Initial DS-II. Using the DRM framework, the project can thus be classified as Type 5:

Type 5: RC (Review) → DS-I (Comp) → PS (Comp) → DS-II (Init)

References

Ackroyd S, Hughes JA (1981) Data collection in context. Longman, New York

Adams JL (1993) Conceptual blockbusting: a guide to better ideas, 3rd edn. Addison-Wesley

Ahmed S (2001) Understanding the use and reuse of experience in engineering design. Engineering Department. University of Cambridge, Cambridge UK, PhD Thesis

Ahmed S, Wallace K, Blessing LTM (2003) Understanding the differences between how novice and experienced designers approach design tasks. Research in Engineering Design 14:1–11

AIAA (1998) Guide for the verification and validation of computational fluid dynamics simulations. American Institute of Aeronautics and Astronautics AIAA G-077-1998

Almefeldt L (2005) Requirements-driven product innovation: industrial needs and supportive methods and tools. Chalmers University, Gothenburg, Sweden, PhD Thesis

Altschuller G (1984) Erfinden: Wege zur Lösung technischer Probleme. VEB Verlag, Berlin

Anderson B (1996) Work, ethnography and system design. Rank Xerox Research Centre, Technical Report EPC 1996–103

Andreasen MM (1980) Machine design methods based on a systematic approach - contribution to a design theory. Department of Machine Design. Lund Institute of Technology, Lund, Sweden, PhD Thesis (in Danish)

Antonsson EK (1987) Development and testing of hypotheses in engineering design research. Journal of Mechanisms, Transmissions and Automation in Design 109:153–154

Ashby M (2005) How to write a paper. www-mech.eng.cam.ac.uk/mmd/ashby-paper-V6.pdf. Accessed 14 December 2008

Beitz W (1979) Anforderungen der Konstruktionspraxis und die CAD-Technologie. Zeitschrift für wirtschaftliche Fertigung 74:459–464

Bender B (2004) Erfolgreiche individuelle Vorgehensstrategien in frühen Phasen der Produktentwicklung. University of Technology Berlin, Berlin, Germany, PhD Thesis

Bessant JR (1979) Preparing for design studies: ways of watching. Design Studies 1:77-83

Bessant JR, McMahon BJ (1979) Participant observation of a major design decision in industry. Design Studies 1:21–26

Björk E (2003) A contribution to insider action research - applied on the development of assistive products. Otto-von-Guericke Universität Magdeburg, Germany, PhD thesis

Blalock AB (1999) Evaluation research and the performance management movement. Evaluation 5:117–149

Blessing L, Wicht A, Meyer-Eschenbach A, Alraheb O (2006) Untersuchung von Pressverbindungen mit gerändelter Welle. Fachgebiet Konstruktionstechnik und Entwicklungsmethodik, TU Berlin, Berlin, Research proposal, unpublished
Blessing LTM (1994) A process-based approach to computer-supported engineering design. University of Twente, PhD thesis
Blessing LTM (1995) Comparison of design models proposed in prescriptive literature. In: Perrin J (ed.) The role of design in the shaping of technology (Workshop COST A3/A4), Lyon
Blessing LTM (1997) A method for identifying scope and assumptions in developing and selecting methods and tools. Cambridge University Engineering Department, Technical Report CUED/C-EDC/TR37
Blessing LTM (2002) What is this thing called 'design research'? In: Annals of the 2002 International CIRP Design Seminar. CIRP, Hong Kong
Blessing LTM (2003) Future issues in design research. In: Lindemann U (ed.) Human behaviour in design: Individuals, Teams, Tools. Springer Verlag, Heidelberg, pp 298–303
Blessing LTM, Andreasen MM (2005) Teaching engineering design research. In: Clarkson PJ, Huhtala M (eds) Engineering Design - theory and practice. A symposium in honour of Ken Wallace. Engineering Design Centre, University of Cambridge, Cambridge, pp 32–41
Blessing LTM, Chakrabarti A (2002) DRM: A design research methodology. In: Les Sciences de la Conception: l'enjeu scientifique du 21e siècle en hommage à Herbert Simon, Ed. J. Perrin, INSA, Lyon, France
Blessing LTM, Chakrabarti A, Wallace K (1995) A design research methodology. In: Hubka V (ed.) International Conference on Engineering Design (ICED'95). Heurista, Zürich, Prague, pp 502–507
Blessing LTM, Chakrabarti A, Wallace KM (1992) Some issues in engineering design research. In: Cross N (ed.) OU/SERC Design Methods Workshop, The Open University, Milton Keynes
Blessing LTM, Chakrabarti A, Wallace KM (1998) An overview of descriptive studies in relation to a general design research methodology. In: Frankenberger E, Badke-Schaub P (eds) Designers: The Key to Successful Product Development. Springer-Verlag, pp 56–70
Bligh TP (1990) Personal communiation
Boehm BW (1979) Software Engineering: R & D trends and defense needs. MIT Press, Cambridge MA
Boehm BW (1988) A spiral model of software development and enhancement. IEEE Compurer 61–72
Booch G, Bryan D, Peterson CG (1994) Software engineering with Ada, 3rd edn. Addison-Wesley
Bracewell RH, Shea K (2001) CaeDre: A product platform to support creation and evaluation of advanced computer aided engineering tools. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). IMecheE, London, Glasgow, pp 539–546
Bracewell RH, Shea K, Langdon P, Blessing LTM, Clarkson PJ (2001) A methodology for computational design tool research. In: Culley S, *et al.* (eds) 13th International Conference on Engineering Design (ICED'01). IMecheE, London, Glasgow, pp 181–188
Braha D, Maimon OZ (1998) A mathematical theory of design: foundations, algorithms and applications. Kluwer Academic Publishers Norwell, MA, USA
Brown D (2008) AI in Design Webliography. http://web.cs.wpi.edu/Research/aidg/AIinD-hotlist.html. Accessed 14 December 2008

Buchanan BG, Duda RO (1983) Principles of rule-based expert systems. Advances in Computers 22:163–216

Buchanan R (2004) Design as inquiry: the common, future and current ground of design. In: Futureground Design Research Society International conference, Melbourne

Bullock A, Stallybrass O, Trombley S (1988) The Fontana dictionary of modern thought. Fontana Press

Burke S (1998) Missing values, outliers, robust statistics and non-parametric methods. http://www.lcgceurope.com/lcgceurope/data/articlestandard/lcgceurope/502001/4509/article.pdf. Accessed 13 December 2008

Button G, Dourish P (1996) Technomethodology: paradoxes and possibilities. In: Tauber MJ, *et al.* (eds) Human Factors in Computing Systems, CHI96. ACM, Vancouver, BC, Canada, pp 19–26

Cantamessa M (2001) Design research in perspective - a meta-research on ICED'97 and ICED'99. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). IMechE, Glasgow, pp 29–36

Carr W, Kemmis S (1981) Becoming critical: education knowledge and action research. Falmer Press, London

Catterall M, Maclaran P (1997) Focus group data and qualitative analysis programs: coding the moving picture as well as the snapshots. Social Research Online 2:12

Chakrabarti A (2006) Defining and supporting design creativity. In: International Design Conference (Design2006), Dubrovnik, Croatia

Chakrabarti A (2008) A post graduate course on DRM – A methodology for design research. In: Chakrabarti A, Subrahmanian E (eds) Indo-US Workshop on Design Engineering 2006. Allied Publishing, New Delhi

Chakrabarti, A, Morgenstern, S, and Knaab, H (2004) Identification and application of requirements and their Impact on the design process: a protocol study, Research in Engineering Design, 15:22–39

Checkland P (1981) Systems thinking, systems practice. Wiley, Chichester

Checkland P (1999) Soft systems methodology: 30-year retrospective. Wiley, Chichester

Clausing D (1994) Total quality development: a step-by-step guide to world class concurrent engineering. American Society of Mechanical Engineers

Coad P, Yourdon E (1990) Object-oriented analysis. Prentice Hall, Englewood Cliffs, NJ

Coffey A, Holbrook B, Atkinson P (1996) Qualitative data analysis: technologies and representations. Social Research Online 1:15

Cook TD, Campbell DT (1979) Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin Company, Boston

Cross N (1994) Engineering Design Methods: strategies for product design, 2th edn. John Wiley and Sons Ltd, Chichester

Cross N, K.Dorst, N.Roozenburg (eds) (1991) Research in design thinking workshop. Delft, The Netherlands

Cuba L (1993) A short Guide to Writing about Social Science, 2nd edn. Harper Collings College Publishers, New York, U.S.A.

Culley SJ (1999) Future Issues in Design Research. In: Culley SJ (ed.) Future Issues in Design Research Workshop. EPSRC/DTI, Bath (UK)

Dagman A (2005) Increased visual quality through robust split-line design. Chalmers University, Gothenburg, Sweden, PhD Thesis

Dasgupta S (1994) Creativity in invention and design: computational and cognitive explorations of technological originality. Cambridge University Press

Davis AM (1990) Software requirements: objects, functions and states. Prentice-Hall, Englewood Cliffs NJ

Denzin NK, Lincoln YS (eds) (2005) The Sage Handbook of Qualitative Research, 3rd edn. Sage Publications, Thousand Oaks

Denzin NK (1978) The research act: a theoretical introduction to sociological methods, 2nd edn. McGraw-Hill, New York
Dick B (1997) A beginner's guide to action research. http://www.scu.edu.au/schools/gcm/ar/arp/actlearn.html. Accessed 13 December 2008
Dillman DA, Smyth JD, Christian LM (2008) Internet, mail, and mixed-mode surveys: The tailored design method. John Wiley and Sons
Dilnot C (2004) The promise and actuality of design research. In: DRS Futureground, Melbourne
Donaldson I (1991) Evaluation of design research demonstator systems. Notes of a joint workshop. CAD Centre, University of Strathclyde, Project Report CONSYN / X / 91–06
Duffy AHB (ed.) (1998) The productivity debate. Springer Verlag, London
Duffy AHB, Andreasen MM (1995) Enhancing the evolution of design science. In: Hubka V (ed.) Proceedings of the 10th International Conference on Engineering Design (ICED'95). Heurista, Zürich, Prague, pp 29–35
Duffy AHB, O'Donnell FJO (1999) A design research approach. In: Mortensen NH, Sigurjonsson J (eds) Critical enthusiasm - contributions to design science. Tapir, Trondheim, N., pp 33–40
Dwarakanath S, Blessing LTM, Wallace KM (1995) Descriptive studies: Starting point for research in engineering design. In: International Conference on Advances in Mechanical Engineering (ICAME), Indian Institute of Science, Bangalore, India, December 20–22, pp 341–361
Eckersley M (1988) The form of design processes: a protocol analysis study. Design Studies 9:86–94
Eckert CM (1997) Intelligent support for knitwear design. Department of Design and Innovation. Open University, Milton Keynes, UK, PhD thesis
Eckert CM, Clarkson PJ, Stacey MK (2003) The spiral of applied research: a methodological view on integrated design research. In: Norell M (ed.) International Conference on Engineering Design (ICED'03), Stockholm
Eilmus S (2007) Methodische Entwicklung eines Prüfstands zur Sitzkomfortmessung. Faculty of Mechanical Engineering and Transportation Systems. Technische Universität Berlin, Berlin, MSc thesis
Erichson B, Plinke W, Weiber R, Backhaus K (2000) Multivariate Analysemethoden. Springer, Heidelberg
Ericsson KA, Simon HA (1996) Protocol analysis: verbal reports as data, 2nd edn. The MIT Press, Cambridge MA
Eriksson J (2007) High performance decision-making in distributed product development. In: Unpublished Poster for the 2007 Summer School on Engineering Design Research
Finger S, Dixon JR (1989a) A review of resaerch in mechanical engineering design. Part II: Representations, analysis and design for the life cycle. Research in Engineering Design 1:121–137
Finger S, Dixon JR (1989b) A review of research in mechanical engineering design, Part 1: descriptive, prescriptive and computer based models of design processes. Research in Engineering Design 1:51–67
Flanagan T, Jänsch J (2004) Models, methodology and finding focus: dealing with the challenges of PhD research. In: International Design Conference Design 2004, Dubrovnik, Croatia
Foraker Design (2005) Online guide to usability resources, Applications: groupware. http://www.usabilityfirst.com/groupware/cscw.txl#introductions. Accessed 14 December 2008
Frankenberger E (1997) Arbeitsteilige Produktentwicklung. VDI Fortschrittberichte. Technische Universität Darmstadt, Darmstadt, PhD Thesis

Frankenberger E, Badke-Schaub P, Birkhofer H (eds) (1998) Designers - the key to successful product development. Springer Verlag, London

Frankfort-Nachmias C, Nachmias D (1996) Reseach methods in the social sciences, 5th edn. St. Martin Press, Inc., London

French M (1985) Conceptual design for engineers. The Design Council, London

Frey DD, Dym CL (2006) Validation of design methods and theories: lessons from medicine. Research in Engineering Design 17:45–57

Fricke G (1993a) Konstruieren als flexibler Problemlöseprozeß: empirische Untersuchung über erfolgreiche Strategien und methodische Vorgehensweisen beim Konstruieren. VDI-Verlag, Düsseldorf

Fricke G (1993b) Empirical investigation of successful approaches when dealing with differently precised design problems. In: International Conference on Engineering Design (ICED'93). Heurista, Zürich, The Hague, pp 359–367

Fricke G, Pahl G (1991) Zusammenhang zwischen personenbedingtem Vorgehen und Lösungsgüte. In: Hubka V (ed.) International Conference on Engineering Design (ICED '91). Heurista, Zürich, Zürich, pp 331–341

Friedrichs J, Luedthe H (1975) Participant observation: theory and practice. Saxon House-Lexington Books, UK

Fuggetta A (1993) A classification of CASE technology. IEEE Computer 26:25–38

Fulcher AJ, Hills P (1996) Towards a strategic framework for design research. Journal of Engineering Design 7:183–193

Galle P (2006) Worldviews for design theory. In: Wondergrounds, Design Research Society International Conference, Lisbon

Galle P (ed.) (2002) Design Studies: special issue on the philosphy of design. 23(3)

Gero J (1990) Design prototypes: a knowledge representation schema for design. AI Magazine 11:26–36

Gero J, Kannengiesser U (2002) The situated function-behaviour-structure framework. In: Gero J (ed.) Artificial Intelligence in Design '02. Kluwer, Dordrecht, pp 89–102

Gibbons M, Johnson C (1982) Science, technology and the development of the transistor. In: Barnes B, Edge D (eds) Science in context, Cambridge MA, pp 177–185

Glaser BG, Strauss A (1967) The discovery of grounded theory. Aldine, Chicago

Grabowski H, Rude S, Grein G (eds) (1998) Universal Design Theory, Proceedings of the Workshop Universal Design Theory, Karlsruhe, Germany. Shaker Verlag, Aachen

Gram C, Cockton (eds) (1996) Design principles for interactive software. Chapman and Hall, London

Griffiths CPA (1995) On being a scientist - responsible conduct in research, 2nd edn. Committee on Science, Engineering and Public Policy, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, National Academies Press, Washington

Grosso WE, Fergerson RW, Musen MA (1999) Knowledge models and axiom languages: The formal underpinnings of Protege. Stanford Medical Informatics, Technical Report

Guba EG, Lincoln YS (1989) Fourth generation evaluation. Sage, Newbury Park, CA

Guindon R (1990) Designing the design process: exploiting opportunistic thoughts. Human Computer Interaction 5:305–344

Hales C (1987) Analysis of the engineering design process in an industrial context. Gant Hill Publications, Hampshire U.K.

Hami-Nobari S (2007) WIBER - Methode zur wirkungsorientierten Beschreibung variantenreicher Erzeugnisse. Berlin University of Technology, Berlin, PhD Thesis

Hammersley M, Gomm R (1997) Bias in social research. Sociological Research Online 2 http://www.socresonline.org.uk/socresonline/2/1/2.html.

Hatchuel A, Weil B (2003) A new approach of innovative design: an introduction to C-K theory. In: Norell M (ed.) International Conference on Engineering Design (ICED'03), Stockholm, Sweden
Hatchuel A, Weil B (2009) C-K design theory: an advanced formulation. Research in Engineering Design 19: 181-192
Heymann M (2005) Kunst und Wissenschaft in der Technik des 20. Jahrhunderts: zur Geschichte der Konstruktionswissenschaft. Chronos Verlag, Zürich
Horváth I (2001) A contemporary survey of scientific research into engineering design. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). ImechE, Glasgow, pp 13–20
Horváth I (2004) A treatise on order in engineering design research. Research in Engineering Design 15:155–181
Hubka V, Eder E (1988) Theory of technical systems: a total concept theory for engineering design. Springer Verlag, Berlin
Hurtienne J, Weber K, Blessing L (2008) Prior experience and intuitive use: image schemas in user centred design. In: Langdon P, *et al.* (eds) Designing inclusive futures. Springer, London, pp 107–116
IEEE (1998) IEEE Std 1012-1998: Standard for software verification and validation. Institute of Electrical and Electronics Engineers
ISO (1998) ISO 9241-Ergonomic requirements for office work with visual display terminals (VDTs) – Part II: Guidance on usability. International Organization for Standardization
ISO (1999) ISO 13407 Human centred design processes for interactive systems. International Organization for Standardization
ISO (2001) ISO 9126-Software engineering – Product quality. International Organization for Standardization
Jacobsen I, Christerson M, Jonsson P, Overgaard G (1993) Object oriented software engineering. Addison-Wesley
Jones JC (1970) Design methods; seeds of human futures. Wiley, New York
Kelle U (1997) Theory building in qualitative research and computer programs for the management of textual data. Social Research Online 2:17
King G, Keohane RO, Verba S (1994) Designing social inquiry: scientific inference in qualitative research. Princeton University Press, Princeton NJ
Kline PJ, Dollins SB (1989) Designing expert systems. John Wiley and Sons, New York
Kuhn TS (1970) The structure of scientific revolutions, 2nd edn. University of Chicago Press, Chicago
Langdon PM, Bracewell RH, Blessing LTM, K.M. W, Clarkson PJ (2001) A practical methodology for integrating sofware and empirical techniques in design. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). IMecheE, London, Glasgow, pp 557–564
Lee RA, Fielding N (1996) Qualitative data analysis: technologies and representations: A comment on Coffey, Holbrook and Atkinson. Social Research Online 1:7
Lenox RS (1985) Educating for the serendipitous discovery. Journal of Chemical Education 62:282
López-Mesa B (2004) The use and suitability of design methods in practice: considerations of problem-solving characteristics and the context of design. Luleå University of Technology, Luleå, Sweden, PhD Thesis
Lossack R-S (2006) Wissenschaftstheoretische Grundlagen für die rechnerunterstützte Konstruktion. Springer, Berlin Heidelberg
Love T (2000) Philosophy of design: a meta-theoretical structure for design theory. Design Studies 21:293–313
Love T (2002) Constructing a coherent cross-disciplinary body of theory about designing and designs: some philosophical issues. Design Studies 23:345–361

Lowe A, McMahon CA, Shah T, Culley SJ (2001) The application of an automatic document classification system to aid the organisers of ICED2001. In: Culley S, *et al.* (eds) International Conference on Engineering Design. IMechE, Glasgow, pp 437–444

Mahlke S (2008) User experience of interaction with technical systems. Technische Universität Berlin, Berlin, PhD thesis

Marsh JR (1997) The capture and utilisation of experience in engineering design. Engineering Department. University of Cambridge, Cambridge, PhD Thesis

Miles MB, Huberman AM (1984) Qualitative data analysis: a sourcebook of new methods. Sage, Beverly Hills

Möhringer S (2005) Entwicklungsmethodik für mechatronische Systeme. Faculty of Mechanical Engineering. University of Paderborn, Paderborn, Germany, Habilitation

Monod J (1969) From biology to ethics. Salk Institute, San Diego

Muller MG, Wildman DM, White EA (1993) Taxonomy of PD practices: a brief practitioners guide. Communications of ACM 36:26–28

N.N. (1977) Linus Pauling: crusading scientist, transcript of broadcast of NOVA, nr. 417, J. Angier, executive producer. In. WGBH-TV, Boston

N.N. (1989) Special issue on rapid prototyping. IEEE Computer

N.N. (2006) Quoting and paraphrasing sources. www.wisc.edu/writing/Handbook/QuotingSources.html. Accessed 19 December 2008

N.N. (2007) What is plagiarism? http://www.lib.berkeley.edu/instruct/guides/citations.html. Accessed 19 December 2008

N.N. (2008) Types of plagiarism. http://www.plagiarism.org/learning_center/types_ of_ plagiarism.html. Accessed 19 December 2008

Nidamarthi S (1999) Understanding and supporting requirement satisfaction in the design process. Department of Engineering, University of Cambridge, Cambridge, PhD thesis

Nidamarthi S, Chakrabarti A, Bligh TP (1997) The significance of co-evolving requirements and solutions in the design process. In: International Conference on Engineering Design (ICED'97). Heurista, Zürich, Tampere, pp 227–230

Nielsen J (1992) Finding usability problems through heuristic evaluation. In: CHI'92. ACM, Monterey, CA, pp 373–348

Norman DA, Draper SW (1986) User-centred system design. Lawrence Erlbaum, Hills Dale NJ

Oliver JE (1991) The incomplete guide to the art of discovery. Columbia University Press, New York

Orne MT (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist 17:776–783

Pahl G, Beitz W (2007) Engineering design: A systematic approach, 3rd edn. Springer

Patton MQ (1987) How to use qualitative methods in evaluation. Sage Publications, London

Patton MQ (1990) Qualitative evaluation and research methods, 2nd edn. Sage Publications, Newbury Park CA

Patton MQ (2002) Qualitative research and Evaluation methods, 3rd edn. Sage Publications Inc., Thousand Oaks, CA

Polson P, Lewis C, Rieman J, Wharton C (1992) Cognitive walkthroughs: a method for theory-based evaluation of interfaces. International Journal of Man-Machine Studies 36:741–773

Preiss M (2006) Ein Reifegradmodell für die Optimierung der Produktentwicklung kundenorientierter Zulieferbetriebe. ETH, Zürich, Switzerland, PhD thesis

Ramon y Cajal S (1999) Advice for a young investigator. The MIT Press, Cambridge MA

Raphael B, Shea K, Smith IFC (1999) A task and software independent CAE course. In: AICIVIL-COMP99: The fifth international conference on the applications of AI to Civil and Structural Engineering, Oxford, UK

Reich Y (1994a) Layered models of research methodologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing 8:263–274

Reich Y (1994b) Annotated bibliography on research methodologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing 8:355–366

Reich Y (1995) The study of design research methodology. Journal of Mechanical Design 177:211–214

Root-Bernstein RS (1989) Discovering: inventing and solving problems at the frontiers of scientific knowledge. Harvard University Press, Cambridge MA

Roozenburg NFM, Eekels J (1995) Product design: Fundamentals and methods. John Wiley & Sons, Chichester

Rossi PH, Freeman HE, Lipsey MW (1999) Evaluation, a systematic approach, 6th edn. Sage Publications, Thousand Oaks

Royce WW (1970) Managing the development of large software systems: concepts and techniques. In: IEEE WESTCON, Los Angeles, pp 1–9

Samuel A, Lewis W (2001) Curiosity-oriented research in engineering design. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). IMechE, Glasgow, pp 37–44

Sarkar P (2007) Development of a support for effective concept exploration to enhance creativity of engineering designers. Indian Institute of Science, Bangalore, PhD Thesis

Saunders J (2008) Computer Supported Collaborative Work. http://www.johnsaunders.com. Accessed 14 December 2008

Schank RC (1990) What is AI, anyway? In: Partridge D, Wilks Y (eds) The foundations of artificial intelligence: a sourcebook. Cambridge University Press, Cambridge, pp 3–13

Schregenberger JW (1997) The further development of design methodologies. In: Frankenberger E, *et al.* (eds) Designers - the key to successful product development. Springer Verlag, London, pp 57–67

Schreiber G, Akkermans H, Anjewierden A, de Hoog R (2000) Knowledge engineering and management: the CommonKADS methodology. MIT Press, Cambridge, MA

Seepersad CC, Pedersen K, Emblemsvåg J, Bailey RR, Allen JK, Mistree F (2006) The Validation Square: How does one verify and validate design methods? In: Chen W, *et al.* (eds) Decision Making in Engineering Design,. ASME Press, New York, pp 303–314

Shah JJ, Finger S, Lu S, Leifer L, C.Cruz-Neira, Wright P, Cagan J, Vandenbrande J (2004) ED2030: strategic plan for engineering design. In: NSF Workshop on Engineering Design in Year 2030 National Science Foundation, USA

Shah JJ, Hazelrigg H (1996) Research opportunities in engineering design. In: NSF Strategic Planning Workshop National Science Foundation, USA

Shneiderman B (1998) Designing the user interface, 3 edn. Addison-Wesley Publishing Company

Simon HA (1981) The sciences of the artificial, 2nd edn. MIT Press, Cambridge MA

Smith AJ (1990) The task of the referee. Computer, IEEE 23:65–71

Smith MF (1991) Software prototyping: adoption, practice and management. McGraw-Hill, London, New York

Smithers T (1998) Towards a knowledge level theory of design process. In: Gero J, Sudweeks F (eds) Artificial Intelligence in Design '98. Kluwer Academic Publishers, pp 3–21

Sommerville I (2006) Software Engineering, 8th edn. Pearson Education

Srinivasan V, Chakrabarti A (2008a) Design for novelty – a framework? In: Marjanovic D, Storga M, Pavkovic N, Bojcetic N (Eds) International Design Conference (Design 2008), Dubrovnik, Croatia, pp 237–244

Srinivasan, V, Chakrabarti, A (2008b) Novelty-SAPPhIRE relationship: an empirical study. In: International Conference on Trends in Product Lifecycle- Modeling, Simulation and Synthesis, Bangalore, India pp 212–220

Stacey M, Eckert CM, Earl C, Bucciarelli L, Clarkson P (2002) A comparative programme for design research. In: Commonground Design Research Society International Conference, Stoke on Trent, UK

Star SL (1997) Anselm Strauss: an appreciation. Sociological Research Online 2:8

Stephenson J (1995) Design for reliability in mechanical systems. Engineering Department. Cambridge University, Cambridge, UK, PhD

Strauss A (1970) Discovering new theory from previous theory. In: Shibutani T (ed.) Human nature and collective behaviour: papers in honor of Herbert Blumer. Prentice-Hall, Englewood Cliffs, NJ

Suchman L (1987) Plans and situated action. Cambridge University Press, New York

Suh NP (1998) Axiomatic Design as a basis for universal design theory. In: Grabowski H, *et al.* (eds) Universal Design Theory. Shaker, Aachen

Takeda H, Tsumaya A, Tomiyama T (1999) Synthesis thought processes in design. In: Kals H, van Houten F (eds) CIRP International Design Seminar. Springer, University of Twente, Enschede, the Netherlands, pp 249–268

Tashakkori A, Teddlie C (1998) Mixed methodology: combining qualititve and quantitative approaches. Sage Publications, Thousand Oaks

Taylor MM (2008) Cite references. http://library.open.ac.uk/help/howto/citeref. Accessed 19 December 2008

ten Have P (1977) Sociologisch veldonderzoek. Boom Meppel, Amsterdam

Thomson GP (1957) The strategy of research. University of Southampton Press, Southampton

Tomiyama T, Yoshioka M, Tsumaya A (2002) A knowledge operation model of synthesis. In: Chakrabarti A (ed.) Engineering design synthesis: Understanding, approaches and tools. Springer, London, pp 67–90

Trochim WM (2006) The Research Methods Knowledge Base. http://www.socialresearchmethods.net/kb. Accessed May 25 2008

Truesdell C (1984) An idiots fugitive essays on science. Springer Verlag, New York

Upton N, Yates I (2001) Putting design research to work. In: Culley S, *et al.* (eds) International Conference on Engineering Design (ICED'01). IMechE, Glasgow

VDI (1993) VDI Guideline 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte. Verein Deutscher Ingenieure VDI, Düsseldorf

Verschuren PJM (1997) Ontwerpgericht onderzoek: een methodologische bezinning. Bedrijfskunde 69:82–89

Visser W (1990) Evocation and elaboration of solutions: different types of problem-solving actions. In: Cognitiva'90, Paris, pp 689–696

Vitruvius (1960) The ten books of architecture. Dover Publications Inc, New York

Vogt WP (1999) Dictionary of statistics and methodology: a non-technical guide for the social sciences, 2nd edn. Sage Publications, Thousand Oaks

Vroom R (2001) Zicht op product- en procesontwikkeling. Delft University of Technology, Delft, PhD Thesis

Waldron MB, Waldron KJ (1988) A time sequence study of a complex mechanical system design. Design Studies 9:95–106

Wallace KM, Blessing LTM (2000) Observations on some German contributions to engineering design. in memory of Professor Wolfgang Beitz. Research in Engineering Design 12:2–7

Weber C (2005) CPM/PDD - an extended theoretical approach to modelling products and product development processes. In: Bley H, *et al.* (eds) 2nd German-Israeli Symposium on advanced in methods and systems for development of products and Processes. Franhofer-IRB-Verlag, TU Berlin / Franhofer-Institut für Produktionsanlagen und Konstruktiontechnik, pp 159–179

Weinert S (2001) Evaluation of a programme to improve acquisition of design strategy in engineering design industry. Department of Psychology. University of Dresden, Dresden, MSc Thesis

Wester F (1987) Strategieen voor kwalitatief onderzoek. Coutinho, Bussum NL

Winston PH (1993) Artificial Intelligence, 2nd edn. Addison-Wesley, Reading MA

Wood KL, Greer JL (2001) Function-based synthesis methods in engineering design. In: Antonsson E, Cagan J (eds) Formal Engineering Design Synthesis. Cambridge University Press, Cambridge, pp 170–227

Yates I (1998) Yes, Prime Minister, but it is also investment, investment, investment,. In: 14th Lord Nelson of Stafford Lecture. Institute of Electrical Engineers, London

Yin RK (1994) Case study research. Sage, London

Yoshikawa H (1980) General design theory and a CAD system. In: Sata T, Warman E (eds) IFIP WG 5.2-5.3 Working Conference. North-Holland, Tokyo, pp 35–57

Yourdon E (1989) RE-3: re-engineering, restructuring and reverse engineering. American Programmer 2:3–10

Yourdon E (1990) Auld Lang Syne. Byte 257–265

Zanker W (1999) Situative Anpassung und Neukombination von Entwicklungsmethoden. Technische Universität München, München, PhD Thesis

Index