IPEK
Institut für Produktentwicklung
am KIT ▪ o. Prof. Dr.-Ing. Dr. h.c. A. Albers

Forschungsberichte

Leif Marxen

A Framework for Design Support Development based on the integrated Product Engineering Model iPeM

Ein Ansatz zur Entwicklung von Konstruktionsunterstützung auf Basis des integrierten Produktentstehungsmodells iPeM

Band 74
Herausgeber: Prof. Dr.-Ing. Dr. h.c. A. Albers

Satz: Leif Marxen

Druck: Stolzenberger Druck und Werbung GmbH & Co. KG, Leimen
06224-7697915

ISSN 1615-8113

A Framework for Design Support Development based on the integrated Product Engineering Model iPeM

Zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
der Fakultät für Maschinenbau
Karlsruher Institut für Technologie (KIT)

genehmigte
Dissertation

von

Dipl.-Ing. Leif Marxen
aus Heidelberg

Tag der mündlichen Prüfung:	15. Mai 2014
Hauptreferent:	Prof. Dr.-Ing. Dr. h.c. A. Albers
Korreferent:	Professor Peter John Clarkson, Ph. D. University of Cambridge / UK

Vorwort des Herausgebers

Wissen ist einer der entscheidenden Faktoren in den Volkswirtschaften unserer Zeit. Der Unternehmenserfolg wird in der Zukunft mehr denn je davon abhängen, wie schnell ein Unternehmen neues Wissen aufnehmen, zugänglich machen und verwerten kann. Die Aufgabe eines Universitätsinstitutes ist es, hier einen wesentlichen Beitrag zu leisten. In den Forschungsarbeiten wird ständig Wissen generiert. Dieses kann aber nur wirksam und für die Gemeinschaft nutzbar werden, wenn es in geeigneter Form kommuniziert wird. Diese Schriftenreihe dient als eine Plattform zum Transfer und macht damit das Wissenspotenzial aus aktuellen Forschungsarbeiten am IPEK - Institut für Produktentwicklung Karlsruhe (ehemals: Institut für Maschinenkonstruktionslehre und Kraftfahrzeugbau) verfügbar.

Die Forschungsfelder des Institutes sind die methodische Entwicklung und das Entwicklungsmanagement, die rechnergestützte Optimierung von hochbelasteten Strukturen und Systemen, die Antriebstechnik mit einem Schwerpunkt auf den Gebieten Antriebsstrang-Engineering und Tribologie von Lager- und Funktionsreibsystemen, die Mikrosystemtechnik mit dem Focus auf die zugehörigen Entwicklungsprozesse sowie die Mechatronik. Die Forschungsberichte werden aus allen diesen Gebieten Beiträge zur wissenschaftlichen Fortentwicklung des Wissens und der zugehörigen Anwendung – sowohl den auf diesen Gebieten tätigen Forschern als auch ganz besonders der anwendenden Industrie – zur Verfügung stellen. Ziel ist es, qualifizierte Beiträge zum Produktentwicklungsprozess zu leisten.

Albert Albers

Vorwort zu Band 74

Die Forschung auf dem Gebiet der Entwicklungsmethoden und -prozesse sowie der Konstruktionsmethodik hat eine lange Tradition. Sowohl in Deutschland, beginnend in den 1940er Jahren, wie auch im angelsächsischen Raum gibt es erhebliche Bemühungen, mit wissenschaftlichen Methoden das Vorgehen und die Arbeitsweisen bei der Konstruktion neuer technischer Systeme zu erforschen. Die Arbeiten von HUBKA, PAHL, BEITZ, HANSEN, Roth, Koller, WEBER, BIRKHOFER, EHRLENSPIEL, LINDEMANN, CLARKSON, SUH und Albers seien als Beispiele genannt. Ziel ist es dabei immer, mit unterschiedlichen Vorgehensweisen und unter Nutzung verschiedenster Methoden und Ansätze den Prozess der Konstruktion und des Konstruierens besser zu verstehen. Es handelt sich dabei also immer um eine Forschung an und mit dem Menschen. Es gibt einige Ansätze, in denen die Autoren ein strukturiertes Forschungsframework für solche Fragestellungen vorschlagen. Genannt seien hier die Modelle von BLESSING oder auch von CLARKSON. Diese beschreiben auf einer sehr generischen Ebene grundlegende Vorgehensweisen, die sicherlich hilfreich sind, um ein Forschungsdesign zu den hier angesprochenen Fragestellungen aufzustellen. In der praktischen Forschungsarbeit zeigt es sich allerdings, dass diese Ansätze nur einen ersten Hinweis geben können und für die wirkliche Planung und Strukturierung von Forschungsprojekten nicht vollständig genügen. An dieser Stelle setzt die Arbeit von Herrn Dr.-Ing. Leif Marxen an. Er hat eine sehr interessante Idee Grundlage für die Lösung detaillierterer Unterstützung im Forschungsdesign entwickelt: Er nutzt einen neuen, flexiblen Ansatz für die Modellierung von Produktentstehungsprozessen das „iPeM“ – Integriertes Produktentstehungsmodell nach ALBERS um auf dieser Basis ein Framework und eine Vorgehensweise für die Planung und Durchführung von Forschungsarbeiten auf dem Gebiet der Produktentwicklungsforschung zu erarbeiten. Er modelliert die Forschung damit äquivalent zur Produktentstehung basierend auf dem ZHO-System nach Ropohl als Transformation von Forschungszielen – oder Forschungsfragen durch ein geeignetes Handlungssystem – dem Forschungsprozess mit seinem Design und seinen Randbedingungen – hin zu Forschungsergebnissen als den zentralen Outputobjekten des Prozesses. Die vorgeschlagenen Ansätze Ergebnisse können mit großem Nutzen von vielen Wissenschaftlern verwendet werden.

Mai, 2014 Albert Albers

Preface of Volume 74

The research in the field of development methods and processes as well as design methodology has a long tradition. Both in Germany, beginning in the 1940s, as well as in Anglo-Saxon countries, there are significant efforts to explore with scientific methods the procedure and working methods in the design of new technical systems. The works of HUBKA, PAHL, BEITZ, HANSEN, ROTH, KOLLER, WEBER, BIRKHOFER, EHRLENSPIEL, LINDEMANN, CLARKSON, SUH and ALBERS shall be mentioned here as examples.

The goal is always to gain a better understanding of the process of design and of designing, applying different approaches and using different methods and procedures. This always involves research on and with humans. There are some approaches in which the authors propose a structured framework for such research questions, e.g. the models of BLESSING or CLARKSON. They describe on a generic level, basic structures that are definitely helpful to set up suitable research designs to investigate the issues raised here. However, in practical research it turns out that these approaches can only give a first indication and do not fully meet the specific tasks of planning, structuring and conducting research projects. This is the target of Dr.-Ing. Leif Marxen's work. He has developed a very interesting idea as a basis for the solution of more detailed support in research design: He uses a new, flexible approach to the modelling of product development processes, the "iPeM" - Integrated product Engineering model by ALBERS. On this basis, he derives a framework and an approach for the planning and conduction of research projects in the field of design science. Thus he models research equivalent to product development based on the ZHO system according ROPOHL as a transformation of research objectives – or research questions – through an appropriate operation system – the research process with its design and its constraints – into research results as the central output objects of the transformation process. The proposed approach and results can be used to great advantage by many scientists.

May 2014 Albert Albers

Kurzfassung

Ziel dieser Arbeit ist es, Methodenentwickler dabei zu unterstützen, *zuverlässige*, *glaubwürdige* und *valide Methoden zur Konstruktionsunterstützung* zu erschaffen.

Die Konstruktionsmethodik als Wissenschaft beschäftigt sich mit den Konstruktionsergebnissen, Konstruieren als Aktivität und mit Konstrukteuren. Das macht sie multidisziplinär und komplex. Zugleich ist sie eine verhältnismäßig junge Wissenschaftsdisziplin, weshalb ihr noch immer eine eigene, etablierte Forschungsmethodik fehlt. Das führt dazu, dass die Ergebnisse der Konstruktionsforschung erheblicher Kritik ausgesetzt sind. Entwicklungsmethoden mangelt es deshalb häufig an Akzeptanz in der industriellen Anwendung. Einige Kritiker aus dem Bereich der Wissenschaft gehen sogar soweit, anzuzweifeln, ob Konstruktionsforschung überhaupt als Forschung bezeichnet werden kann. Die Konstruktionsforscher haben darauf reagiert. Abstrakte Ansätze wie zum Beispiel „DRM“ oder „The Spiral of applied Research“ wurden entwickelt um Forschungsprojekte methodisch zu organisieren. Zudem sind zahlreiche spezifische Methoden entwickelt worden, die sehr spezielle Forschungsaktivitäten unterstützen sollen. Beide Aspekte sind wichtig und wertvoll für die Erfolgreiche Durchführung von Forschungsprojekten. Dennoch bleibt das Problem, dass verfügbare Methoden entweder abstrakt sind, ohne konkrete Methoden einzubeziehen – wie genau man also von einem zum nächsten Stadium innerhalb eines Forschungsprojektes kommt wird nicht hinreichend adressiert – oder aber die Methoden zielen ausschließlich auf spezielle aber isolierte Aktivitäten in der Konstruktionsforschung ab, die nicht Teil einer übergeordneten Methodologie sind.

In der hier vorliegenden Dissertation wird deshalb ein Ansatz entwickelt, der auf dem „Integrierten Produktentstehungs Model iPeM“ basiert. Er erlaubt die einheitliche Beschreibung, Planung und Durchführung von Projekten zur Entwicklung von Methoden und Werkzeugen zur Konstruktionsunterstützung mit Hilfe eines Sets generischer Schritte. Zusätzlich zu diesem Ansatz wird eine Sammlung spezifischer Methoden bereitgestellt, die aus der Konstruktionsforschung sowie aus anderen Wissenschaftsdisziplinen wie der Management Forschung, der Statistik oder den Sozialwissenschaften zusammengetragen wurden. Der vorgestellte Ansatz stell damit die Verbindung zwischen der abstrakten Organisation von Forschungsprojekten in der Konstruktionsforschung und den konkreten Methoden und Aktivitäten zu deren Durchführung her. Er stellt einerseits eine gemeinsame Sprache für Konstruktionsforscher bereit um ihre Ergebnisse zu vergleichen und zu diskutieren, und um ihre Aktivitäten zu planen. Andererseits enthält er eine Sammlung notwendiger Methoden zur Durchführung der geplanten Aktivitäten. Diese Sammlung kann erweitert werden. Der Ansatz hat damit das Potential mit der Weiterentwicklung der Konstruktionsforschung zu wachsen. Seine Anwendung wird zu neuem Erfahrungswissen führen. Dieses kann in Form von Mustern gespeichert und für zukünftige Forschungsprojekte abgerufen werden.

Abstract

The overall goal of this thesis is to help design support developers in their efforts to provide *reliable, credible and valid design support.*

Design science deals with design outcomes, with designing as an activity, and with the designers, making it multidisciplinary and complex. At the same time, it is still a relatively young discipline and is therefore still lacking its own established research methodology. In consequence, the outcomes of design science are exposed to criticism. Design support often lacks acceptance in industrial practice and some critics in the scientific world even doubt whether design science should be called science at all. Members of the design science community have reacted to this problem. Frameworks like “DRM” or the “Spiral of applied Research” have been suggested to arrange research projects from a methodological point of view. On the other hand, a variety of specific methods supporting very specific research activities have been developed and presented in the past. Both aspects are important and valuable for successful research projects. However, the methodological support available to design researchers today either draws the big picture and does not include concrete methods - how you get from one research stage to the next is not sufficiently addressed - or it represents very specific yet isolated activities in design research that are not part of a larger methodology.

In this thesis, therefore, a framework is developed, based on the “Integrated Product Engineering Model iPeM”. It allows to describe, plan, and conduct design support development projects in a set of generic steps. Alongside the framework a collection of specific methods is provided, drawn together from within design science as well as other disciplines like management science, statistics and sociology. The framework poses a link between the big picture and concrete methods. It provides a common language for design researchers, to compare, discuss and plan research activities, together with a collection of research methods to conduct the necessary steps of a research project. It is open for the integration of further research methods and hence has the potential to grow as design science matures.

Its application will provide experience that can be documented in patterns for use in future research projects.

Acknowledgements

This thesis was written during my time as a research associate at IPEK – Institute of Product Engineering at the Karlsruhe Institute of Technology (KIT). I am especially thankful for my mentor, Prof. Dr.-Ing. Dr. h.c. Albers. The discussions with him, his enthusiasm for my work, and his strong support were more than helpful. I also want to thank my co-examiner, Prof. Clarkson, for his time and effort to help me improve my work.

During this time, I could rely on the support from many colleagues of which a number have become close friends. I would like to thank all of them for the time, for their helpful remarks, and for the support they gave me in my daily work.

Mirko, you had a major role in my starting at IPEK. If it wasn’t for you, who knows where I might have ended up instead. Thomas, I loved working with you. You tought me one of the greatest comunication techniques ever. Which one? You can decide that yourslef. I am sure you know. Petra, we solved a huge pile of small day-to-day problems together. Without you, I would still be stuck in that pile today. Not even SPALTEN would have gotten me out. Jochen, we always challenged one anothers thoughts and ideas without mercy but always with the respect of true friends. Thank you for giving me such a hard time. I truly hope I could and will be of similar assistance to you. Peter, you are one of a few who can truly share my exitement for my new projects. You started some of them. Thank you for introducing me to the art of espresso and thank you for making me hang on at all times. To all my friends and family that patiently accepted my “Sorry, I can’t. I want to work on my thesis this weekend.” for years without any complaint: Thank you for your understanding and support!

To my parents: You have always worked hard to give me the best opportunities. You were always there for me and never let me down. That’s all I could ever have asked asked for. Thank you.

Katrin, thank you so much for believing in me. Your endurance when you had to listen to me complainig on days that weren’t so productive was remarkable. You were happy for me, every time I got a step closer to the goal, although, at times, the necessary steps seemed to become more and more. In the midst of all that, you married me. I love you.

„Counting sounds easy until we actually attempt it, and then we quickly discover that often we cannot recognize what we ought to count. Numbers are no substitute for clear definitions, and not everything that can be counted counts "

William Bruce Cameron
sociologist 1958

"Not everything that can be counted counts, and not everything that counts can be counted."

allegedly, the saying was written on a
chalkboard in Albert Einstein's office

Table of contents

1 Introduction 1

2 LiteratureReview 5

2.1 Acceptance of Design Support 5

2.1.1 Performance Presentation and Process 6

2.1.2 Industry’s unrealistic Wishlist 8

2.1.3 Dimensions of Design Activities 10

2.1.4 Criticism against Design Support 11

2.2 The Dilemma of developing heuristic Design Support 16

2.2.1 Design Research has to involve multiple Disciplines 16

2.2.2 Anticipating the User 18

2.2.3 Quality Assessment of Design Support 19

2.2.4 Rules of Experimentation do not apply - it can only be done once 20

2.3 Do we need Support for Design Support Development? 21

3 Target and Scope 26

4 Status quo – Research Methods for Design Research 28

4.1 Available Approaches within Design Science 29

4.1.1 DRM: A Design Research Methodology 29

4.1.2 The Spiral of applied Research 37

4.1.3 Cantamessa’s empirical Perspective 41

4.1.4 Reich’s layered Model 42

4.1.5 The Validation Square 44

4.1.6 Foundation for the Development of Design Methods 47

4.1.7 Horváth’s Reasoning Model 48

4.1.8 Medical Treatment – Design Method Analogy 50

4.1.9 Method Impact Evaluation 53

4.1.10 Validation of Decision Support Tools 54

4.1.11 Requirements for Method Development 55

4.1.12 General Advice for Conducting (Design) Research 56

4.2 Methods in Design Research 56

4.2.1 Evaluation based Task Design 57

4.2.2 Creativity and Creativity Assessment 60
4.2.3 Research Methods for Content oriented Design Research 74
4.2.4 Research Methods for Process oriented Design Research 83
4.3 Scientific Methods from other Field of Science..103
4.3.1 Data Collection Techniques...104
4.3.2 Statistics 110
4.3.3 Validity...139
4.3.4 Reliability 145
4.3.5 Validation and Verification 147
4.3.6 Summing up 148
4.4 Research Paradigms...148
4.4.1 Rationalism..149
4.4.2 Authoritarianism..150
4.4.3 Empiricism 150
4.4.4 Positivism 151
4.4.5 Foundationalism 151
4.4.6 Reductionism..151
4.4.7 The Relativistic / Holistic / Social School of Epistemology...................152
4.4.8 Summing up 152
5 Supporting the Development of Heuristic Design Support 155
5.1 Comparing Design Support Development and Product Engineering 158
5.1.1 System of Objectives...160
5.1.2 Operation System...161
5.1.3 System of Objects 165
5.2 Expanding the Operation System..167
5.2.1 Activities of Design Support Development...168
5.2.2 Activities Matrix for Design Support Development...............................172
5.2.3 Navigating the Activities Matrix of Design Support Development 179
5.2.4 Dynamic Modelling of Design Support Development...........................180
6 Exemplary Application of the Framework 185
6.1 Project Planning and Controlling 185
6.2 Empirical Research and Analysis of real-world Processes.........................187

6.3 Embodiment of Design Support....190
6.4 Experimental Studies, Evaluation in controlled Environment....198
6.5 Implementation Studies....201
6.6 Transfer Studies....204
6.7 Analysis of Utilization....207
6.8 Summary of the Sample Project....208
7 Conclusion and Outlook....211
8 References....217
9 Appendix....243
9.1 Definitions and Explanations....243
9.1.1 Algorithmic Methods - Heuristic Methods - Heuristics....243
9.1.2 Algorithmic Methods....243
9.1.3 Heuristic Methods....243
9.1.4 Heuristics in other scientific disciplines....244
9.2 Worksheets....245
9.2.1 Code of fair Testing Practices in Education....254
9.3 Tables....260
9.3.1 Table of the Standard Normal Distribution....260
9.3.2 Student’s Distribution....262
9.3.3 Critical Values for Wilcoxon’s signed-rank-sum-test....263
9.3.4 Chi-Square Table....264
9.4 Recommendations for the Evaluation of Tests....265
9.5 Call for proposals “Innovations against Product Piracy”....271
10 Epilogue - Scurvy, an early Experiment....272

1 Introduction

It has been two and a half million years since man started to use tools in order to accomplish the tasks he was faced with every day. Yet, it has only been three hundred years since THOMAS NEWCOMEN invented the steam engine[1] and rapidly changed the world. Industrialization started to spread around the globe. Engineering, thus far a profession strictly associated with the design of military equipment, started to turn into one of the most important professions for a country's economy. Successful engineers were considered artists far into the twentieth century. One of the pioneers who took a scientific interest in what makes a good engineer and which factors determine a prosperous design, was FERDINAND REDTENBACHER (1809-1863), who is today considered the founder of what we call design science.[2] However, the scope of design science was rather limited for another hundred years. Well into the second half of the twentieth century, design science was limited to successful engineers describing very specifically how to design and calculate concrete *technical systems*. Design science that focuses on the *design process*, the human activities within, and the way designers think and behave has only been around for a little more than half a century.[3] Since then, however, it has produced a great variety of valuable insight on engineering processes, technical systems and methods aiming to support engineers in their day-to-day design tasks.

Design support has taken various developments. With the rise of computer technologies, computer-aided engineering has had a large share of the attention of the scientific communities in the recent past. With the ongoing rise in computational power, artificial intelligence and computer-aided engineering were supposed to be the dominating future technologies in the late nineteen eighties and early nineties. While we are still seeing further development in this area, the discussion about artificial intelligence has somewhat died down, at least in the field of engineering design science.

It is the author's belief that the design engineer – a creative, thinking human being – is and will always be the most important element in the design process. Therefore, this thesis focuses on methods that support design engineers. Those methods can be distinguished into two types, *algorithmic methods* and *heuristic methods.*[4] While algorithmic methods do play an important role in engineering in general, heuristic methods are the methods design science has been at odds with. They contain

---

[1] His design, patented in 1712, was the first steam-driven piston. There was an earlier invention of a steam-operated pump in 1698, creating a vacuum through condensation of steam in a cooled vessel. This design is also sometimes referred to as the first steam engine.

[2] Redtenbacher described engineering as a mixture of art and science in Redtenbacher (1852).

[3] 1962 is defined by many as the advent of design research as this is the year in which the first "Conference on Systematic and Intuitive Methods in Engineering, Industrial Design, Architecture and Communications" was held. See Jones (1962).

[4] See chapter 9.1 for a detailed description and definition.

subjective judgment, which is significant for any kind of creative process but hard to describe, understand and analyze from a scientific point of view. In the development of heuristic methods, it is especially challenging to prove their efficiency and effectiveness. Developing valid heuristic methods is an extremely demanding task and will be the central topic of this thesis.

While for algorithmic methods, one can measure the standard deviation of the results achieved with the method after repeatedly applying it under comparable conditions, heuristic methods cannot be reiterated under constant conditions. As they involve a human decision-maker, repeatedly applying such a method would bias the result, as the decision-maker builds up experience with the problem and its possible solutions. Replacing the human decision-maker would bias the result just as much as it completely changes the conditions of the experiment. Exchanging the problem to be solved would lead to non-comparable results, so there is a “logical bug” in the validation of heuristic methods. “No longer are we dealing with repeatable cause and effect, but rather cause, followed by humanly manipulated effect where the human input is variable, leading to many possible outcomes.”[5]

It will be the central goal of this work to develop an approach that helps developers of heuristic methods cope with this dilemma and produce reliable results. Before that, chapter two will sum up and review literature that contains criticism towards design science. It will show that this criticism is one of the reasons, the results produced with design science lack acceptance in industry. Furthermore, chapter two shows the challenges that lie in validating design methods and the parallels with other human-centered sciences such as sociology and psychology. In chapter three, the findings from literature will be interpreted and the motivation of this thesis will be derived along with the corresponding goals. Chapter four will draw together available research methods and approaches both from design science and from other fields such as psychology and sociology. The goal is to present a set of selected approaches and methods design researchers can choose from and to describe them in context. This collection will be one of the two major achievements of this thesis. It will build the foundation for the framework presented later that will help to model, plan and navigate individual research projects as well as choose and combine suitable methods from this collection. This framework poses the second major achievement of this thesis and will be developed and presented in chapter five.

In chapter six, the application of the framework will be described using an example taken from an actual research project. The goal is to show the usability and effectiveness of the framework and to identify possible opportunities for future improvement, which will be presented in chapter seven together with a summary of the major findings and achievements of this work.

---

[5] Fulcher / Hills (1996), p. 185

In order to improve readability in the main chapters, some work-sheets, tables, and explanations will be located in chapter nine as attachments following the list of references that can be found in chapter eight.

2 LiteratureReview

“*Design researchers frequently lament that design research is not scientific and that a methodology needs to be created to put design research on a scientific footing. But most design research is – or should be – grounded in the techniques and methodological rigor of one of several academic disciplines that treat design as another human activity. These disciplines, including cognitive psychology, artificial intelligence, complexity science, and various flavors of sociology, have very sophisticated views of what are effective research procedures, what constitutes adequate methodological rigor, and what is the epistemological status of their findings. While cognitive psychology is certainly science, a lot of valid design research doesn’t fit most philosophers’ definitions of science.*”[6]

The following sections will review current literature and will show that the above statement by ECKERT ET AL. is true, especially for the development of design support in three steps:

- Design research produces methods that are only of limited acceptance.
- Design methods’ effectiveness is difficult to prove.
- Several others have pointed this out repeatedly and call for a research methodology.

2.1 Acceptance of Design Support

“Acceptance describes the willingness of individuals and organizations to apply a methodology in practice.”[7]

Both critics and supporters of engineering design science believe that its success can be determined by observing the degree to which its outcomes are applied in practice. It is also the author's opinion that engineering design scientists can only be called successful if they manage to produce useful support that practice adopts and applies. So, when talking about developing valid design support, showing a method’s usefulness is an important aspect. In the following paragraphs, the goal is to show that there is a serious problem with a lot of design support: They are confronted with very limited acceptance in practice. There is the reason to believe that it is not necessarily the methods which are useless. Much more, their creators have difficulties proving and communicating the methods’ usefulness. Methodological support could improve this situation.

For the past decades, scientists have come up with methods to support engineers in practice. However, only few methods actually make it. Many methods are discarded

---

[6] Eckert et al. (2003), p.3

[7] Albers / Lohmeyer (2012)

and never used.[8] “In industrial design work[9] the use of design methods plays a varying, but mostly minor role. There seems to be a substantial gap between the needs of designers working in a competitive industrial environment and the outcome of design research. The question is why all these results and outcomes of more than 40 years of research do not have more influence on design in its entirety.”[10] Several authors have tried to find explanations for this unfortunate faith of so many good efforts in design science.

2.1.1 Performance Presentation and Process

BADKE-SCHAUB ET AL. group today’s criticism of design methods in three major fields:

- The questionable *performance* of methods
- The ways methods are *presented* and formulated
- *Process*-related problems during the application of methods.

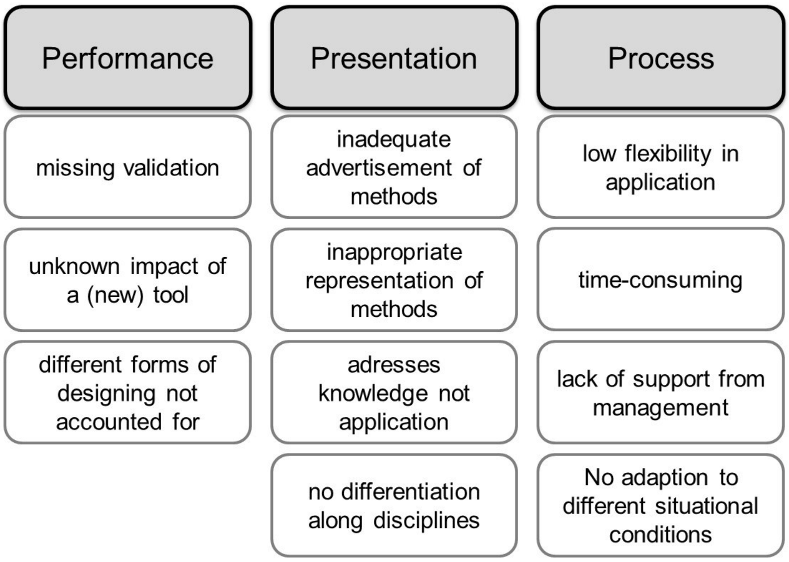


Figure 1 - Criticism about design methods, Badke-Schaub et al. (2011), 183

2.1.1.1 Performance

BADKE-SCHAUB ET AL. relate performance issues to the question whether using a certain method really improves the design results. Inconsequent validation may lead to low performance just as well as unknown or unwanted side-effects. Insufficient performance does not necessarily result only from “bad” methods. It may also be the misuse of methods, i.e. their application in situations they are not intended for or a mismatch between the characteristics of the user and the method. Low performance could be avoided if design support was delivered with information about its proper application and test results about how it performed during validation.

---

[8] See also Frost (1999), Gill (1990), Horváth (2001), Badke-Schaub et al. (2011)

[9] Authors remark: “industrial design work” in the quoted source refers to design work in companies, not the design of buildings.

[10] Birkhofer (2004), 5 f.

2.1.1.2 Presentation

Criticism subsumed in this category relates to the improper representation of methods such as "[…] abstract language used to describe the procedures of methods, which seems to be inappropriate for use in practice."[11]

Unappealing presentation can be avoided if the creators of design support use easy to understand and clear language and leave out unnecessary information for practitioners. To the author's knowledge, there is no established set of rules on how to represent the results of method development for practitioners.

2.1.1.3 Process

Process issues according to BADKE-SCHAUB ET AL. have a multitude of sources. While some authors blame companies' management for not showing enough interest in the methods, mostly time consumption and inflexibility are named as the root causes for lacking acceptance. If the application of a method threatens to postpone milestones or forces an organization to change its established patterns of activities – usually called "the company's design process" – the method tends to be rejected.

Such issues could be avoided if method development included information about ideal situations for the method to be applied within a design process. However, this would require a universal way to describe both methods and design processes. BIRKHOFER has expressed his concern about the way methods are presented to practice. Their documentation delivers knowledge about the method and its roots but is lacking didactic elements that enable the reader to develop competency in applying the method.[12] BIRKHOFER ET AL. have started to publish work aiming to provide a universal way to describe the order of actions that make a method.[13] ALBERS ET AL. have been publishing research on the Integrated Product Engineering Model iPeM since 2007.[14] It provides a universal language to describe design processes.[15] MARXEN AND ALBERS have suggested for creators of design methods to use iPeM as a standardized language to describe potential application scenarios of a method.[16]

According to BADKE-SCHAUB'S explanations, the developers of design methods should not only deliver the method itself. They should prove beforehand that its application will help and will not have unintended side-effects; they are supposed to provide information about when to apply the method within the design process and preferably deliver all the information in a way that is easy to understand, so that practitioners will

---

[11] Badke-Schaub et al. (2011), 184

[12] Birkhofer et al. (2002), 457 ff.

[13] Ibid.

[14] The very first works on the Integrated Product Engineering Model where under the title "IPEMM-Integrated Product Development Process Management Model" Albers / Meboldt (2007a). For comprehensive reading on iPeM see also: Meboldt (2008) and Albers / Braun (2012)

[15] Albers / Marxen (2012), Albers et al. (2011a)

[16] Marxen / Albers (2012)

not have to invest too much time to understand the method. It is, however, not clear whether fulfilling all these demands is realistic.

2.1.2 Industry's unrealistic Wishlist

One might argue that it is nearly impossible to reach industry's demands when it comes to design methods. BIRKHOFER lists the following requirements from industry which he calls *evidently unrealistic:*

"Methodical support in design…

- should need as little effort for learning and training as possible.
- should be easy to use.
- should solve problems 'in no time'.
- should produce convincing results for complex problems.
- should not be islands of support, but integrated in the existing design environment."[17]

Although it is indeed hard to fulfill all the requirements completely, and at the same time. However, not fulfilling them and expecting designers to use the methods seems evenly idealistic. Hence, BIRKHOFER also criticizes developers who make unrealistic promises about their methods as such behavior contributes further to the disappointment on the user side. "There have to be mentioned, too unrealistic promises of researchers, who produce new methods with enthusiasm, but neglecting at the same [time] the effort for learning, for adapting and integrating them in practice."[18]

Other authors, who have a more specific view on the lacking acceptance, list more or less single details of the more comprehensive studies presented thus far. The following paragraphs will give a brief and exemplary overview of typical criticism. Since this is a long and ongoing discussion, many authors have given statements on lacking acceptance. It is not the goal of this thesis to list them all. However, from the author's observation, acceptance factors have become an increasingly popular field within design science. BADKE-SCHAUB, FRANKENBERGER, BIRKHOFER, LINDEMANN and ALBERS have been investigating acceptance factors, and it is to be expected that future work by other researchers will be devoted to this topic.[19]

GEIS ET AL. conducted a *study on teamwork* and use of methods in which they asked practitioners in mechanical engineering to express what they expect from methods.[20] Table 1 shows the requirements as stated by the participants of the study. A similar list resulted from the study "New ways towards product development" initiated by the

---

[17] Birkhofer (2004), 6

[18] Birkhofer (2004), 6

[19] i.e. some of the latest work of Albers: Albers et al. (2012a), Albers / Lohmeyer (2012), Frankenberger / Badke-Schaub (1998), Birkhofer et al. (2002), Lindemann (2002)

[20] Geis et al. (2008)

German Ministry for Education and Research (German: BMBF) and conducted by GRABOWSKI AND GEIGER (see Table 2).[21]

Table 1 – Expectations towards design methods from practice (from Geis et al. (2008)

	Methods should
Interaction	▪ improve speed and effectiveness of communication ▪ support presenting and discussing ideas competently and objectively ▪ help in reaching agreements
Planning	▪ help in planning, organizing and controlling projects or processes ▪ support analysis of the process (in general and actual state) ▪ ensure sustainability of actions and measures ▪ support individual time and project management ▪ help prioritize work quotas
Usage of methods:	▪ be simplified ▪ the flexibility of methods should be improved ▪ focus on the output and have less theoretical ballast ▪ be better integrated in the process ▪ be improved from time to time according to the wishes of the users

Table 2 – Findings from “New ways towards product development”[22]

Things, methods should achieve:	**Problems users see for the application in industrial practice:**
▪ improvement of processes ▪ reduction of iterative loops ▪ visualization of existing knowledge ▪ support in reaching targeted cost and deadlines ▪ savings of time and cost ▪ support of documentation ▪ help regarding technical and organizational decisions ▪ support reaching customer- and goal-oriented decisions ▪ support in accessing linked information	▪ Too much effort needed for some of the methods ▪ Too much “theoretical ballast” for some of the methods ▪ Lack of preparation and support for the application of the methods ▪ Missing computer assistance ▪ Missing support of an individual design process through CAD systems ▪ Missing willingness to apply methods on a trans-sectoral level ▪ Implementation of new methods and possibilities is to slow ▪ Varying applicability of methods depending on company size ▪ Underestimation of the importance of FMEA and QFD ▪ Missing holistic methodology-framework with integrated single methods ▪ Missing variant management ▪ Missing methods for change management (e.g. for the estimation of change induced effort) ▪ Missing standards of solution principals with internationally understandable design and development documentation

---

[21] Grabowski / Geiger (1997)

[22] Grabowski / Geiger (1997), 46 f.

JÄNSCH also deals with the matter and calls it a “transfer-problem” with three main areas of concern:

- Presentation and documentation:
  First of all, the reasons can be found within the design method itself that make it difficult to transfer it to teaching such as heterogeneous terms, different paradigms, high level of abstraction, too theoretical, inconsistent representation, high level of complexity and so on.
- Learning and teaching issues:
  A further reason is mediation. It is unclear as to how design methods could be mediated more efficiently. The questions remain, how application competence could be achieved and whether or not design methods suit human problem-solving behavior.
- Acceptance and application problems:
  There are doubts and skepticism towards the introduction, the need for adaption whether or not there is a demand, the time consumption and terminological misunderstandings and so on. Transfer problems are therefore, not simply acceptance problems but also application, teaching, presentation and documentation problems.[23]

JÄNSCH proposes increased transferability in order to achieve application of newly developed methods in practice as a major success criterion and goal for the advancement of such methods. Hence, she suggests the following list of measures that should help achieve this goal:[24]

- create more operational procedures for application: more focus on practicability, hints, tips and advice, description of benefit of usage, etc.
- creation of user-specific methods: consideration of knowledge/expertise, identification of user- and department-specific needs, etc.
- company specific adaptation of methods: usage of identical and homogenous terms and visualization, adaptation of abstraction degree, adaptation of methods according to product and industrial sector, etc.
- education: usage of existing concepts of education and continuing education in companies for method implementation.

2.1.3 Dimensions of Design Activities

In his research on design processes, DORST noted: “Design Methodology has always had something of a blind spot for design problems: the focus in Design Methodology has almost exclusively been on the support of the *process* of designing. However, any method for aiding design activities necessarily contains statements or assumptions about all three *dimensions of design activities*” (Figure 2).[25] The three dimensions, the

---

[23] Translated from Jänsch (2007), 50

[24] Jänsch (2007)

[25] Dorst (2003), 1

dynamics of a design process, the designer and the design problem have been mentioned earlier by ROOZENBURG AND CROSS.[26]

DORST criticizes that "most process-focused design methods seem to incorporate strong assumptions about what design problems are (e.g. concerning the independence of sub-problems, the objectivity of problems, the possibility to create an overview of a design problem, etc.)."[27]

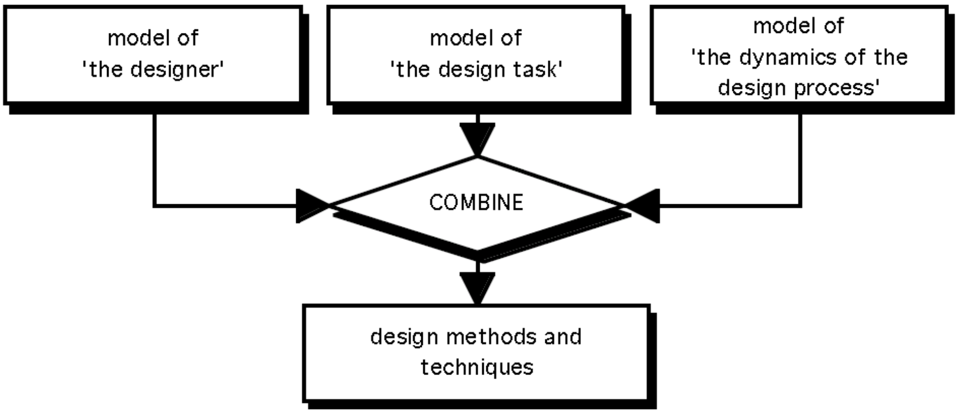


Figure 2 - The three dimensions of design activities, Dorst (2003), 1

Such criticism reminds of the "presentation-issues" discussed by BADKE-SCHAUB and can be avoided if the developer of a design method made such assumptions explicit.

2.1.4 Criticism against Design Support

HUTTERER gives a detailed overview on existing criticism.[28] He explicitly names *lack of performance* due to complexity and the methods being too *overloaded with theory*, rigidity and lack of flexibility, effort and late return on invest, their prescriptive character and finally misinterpretations that lead to wrongly expected generality. He also names what he calls a *subjectively felt lack of availability of methods.* It is caused by both lack of knowledge about which methods are actually ready for use and lack of understanding of the available methods accompanied by individual misjudgments. Further, he identifies *missing options to estimate a method's capability*, as well as *missing approaches to choose adequate methods*. The reasons he gives are methods generally being difficult to assess and missing ways to estimate effort, benefit and extend in advance. Finally, he names *insufficient adaption of methods to a specific problem* that lead to rather general solutions for not so general problems and disregard of intuitive parts of method application.[29]

---

[26] Roozenburg / Cross (1991)

[27] Dorst (2003), 2

[28] Hutterer (2005)

[29] For the complete and detailed overview see Hutterer (2005), 15-28

2.1.4.1 Unsuitable for individual creative Style

WÖGERBAUER noticed in the early days of design science that the benefit of a method not only depends on its quality but is also strongly affected by human traits and individual acceptance.[30] It is therefore important to take into account designers' characteristic style of thinking, problem solving or just the way they engage in creative activities. Some authors have shown that the character of methods rarely fits the users' thinking style. Hence, potential users criticize that methods cannot be applied intuitively. "Acceptance problems are observed in industrial practice, due to the collision of the logical-systematic character of methods with the individual thinking styles and behavior of human individuals."[31]

JORDEN argues that design methods are too complicated, too theoretical, too hard to memorize, and too time-consuming for practitioners. He suggests that a design methodology should be introduced, which aims at problem-solving rather than at logical procedures[32]. SCHREGENBERGER also focuses on the problem solving aspects[33] and requests that scientific inquiry on the topic should be more psychologically and heuristically oriented.[34] "[Design methods] are practically bulky, ignore individual working styles of designers and from case to case slap his/her experience in the face. Their heuristic power is limited and they omit the difficulties of division of work as it occurs within design teams"[35]

2.1.4.2 Unsuitable for the Problem at Hand

ZANKER argues that design support is developed for a certain purpose, for specific situations and constraints. Users have to check whether this suits their personal constellation or not, which again makes it difficult to use the methods, diminishing acceptance.[36] "The first issue relates to the performance of methods and addresses the question of whether it is proven that design methods really lead to superior design performance. Even when methods are applied, the design performance can still be low. This is due to poor use of methods, or the quality of the method itself.[37] Weak performance can be caused by a mismatch between characteristics of the chosen method and the task or problem at hand, or due to incorrect timing in the process."[38] LINDEMANN makes this problem one of the key topics in one of his works and suggests a set of approaches to adapt methods to situation and preferences, individually. PAHL AND BEITZ, as well as EHRLENSPIEL do not explicitly discuss the issue as a problem.

---

[30] Wögerbauer (1942), 173; cited in Jänsch (2007), 46
[31] Pahl (1994), 8ff., authors translation
[32] Jorden (1983), 494
[33] Schregenberger (1980)
[34] Schregenberger (1983), 524
[35] Ibid., author‘s translation
[36] See Zanker (1999),149 f.
[37] Cp. 2.1.1.1
[38] Badke-Schaub et al. (2011), 183

However, in their work on methods for engineering design, they warn the user not to apply them without the courage to adapt them to the problem at hand.[39]

This seems to show that future success of design methods can most likely be achieved, if both parties, the creators as well as the users of design methods are aware of the merits and the limitations of methods and act accordingly smart. As BIRKHOFER put it: *"We have to accept that simple methods and tools probably produce no more than simple results. In consequence, we have to expect, if methods and tools are successfully used in real design work, designers must contribute to an adequate intellectual level. And without the appropriate amount of time for learning and training and without the motivation to modify their own behavior in problem solving, the success of designing methodically can hardly be achieved."*[40]

2.1.4.3 Difficult to use and understand

FROST argues that industry does, in fact, use many of the design methods suggested by scientists but do not know it. The reason he gives is that the methods are actually based on observations from industry. Scientists describe what designers are doing, call it design method and then "give it back" to them to apply it. This is nothing spectacular but simply describes empirical research. Nevertheless, FROST criticizes: "*Very often, however, these descriptions are also couched in such abstract terms as to be almost unrecognizable against the methods and activities which they represent, especially by industry itself, whose major focus of attention is inevitably upon achieving pragmatic outcomes, rather than on esoteric abstraction.*"[41]

BIRKHOFER repeatedly criticized that methods are described too theoretically. Their documentation carries a lot of information about the method and its theoretical background. However, it does not carry application competence.[42] This makes methods difficult to use mainly because they are hard to understand in the first place. Suggestions to meet this problem are e.g. new teaching methods that include practicing methods and hence deliver competency rather than pure knowledge. We have to acknowledge that using methods is a type of competency.[43] The criticism can be taken as a reminder for scientists to describe their findings in a way that is understandable rather than imposing, although this is nothing particular to the field of design science but can generally be considered good practice.

When you wish to instruct, be brief; that men's minds take in quickly what you say, learn its lesson, and retain it faithfully. Every word that is unnecessary only pours over the side of a brimming mind.

---

[39] Ehrlenspiel (2007), Ehrlenspiel et al. (2005), Pahl et al. (2005), Pahl et al. (2006)
[40] Birkhofer (2004), 6
[41] Frost (1999)
[42] Birkhofer et al. (2001), Birkhofer et al. (2002), 17; Jänsch / Birkhofer (2004)
[43] acatech (2012)

Marcus Tullius Cicero 106 BC – 43 BC; Roman philosopher, statesman, lawyer, orator, political theorist, consul and constitutionalist

2.1.4.4 Time consuming Effort for Learning and Training

One of the reasons for the lack of acceptance identified e.g. by JORDEN[44] or RUTZ[45] is that potential users are skeptical concerning value to the cost. It is necessary to spend time and money in order to learn a new method. How does one know that it will actually work?

LINDEMANN states: "Up to now there are hardly any methods to analyze methods in terms of effort and benefit. Trial and error as well as consultancy based on experience [are] the mostly used ways of gathering information about methods."[46] He suggests applying SWOT analysis or TRIZ based functional analysis before using a method in order to avoid waste of time.[47]

The effort users have to invest to learn a new method has also been mentioned by BIRKHOFER. He provocatively claims that industry demands for methods, which need no training whatsoever.[48] Engineers in practice, on the other hand, claim that they are willing to invest time to learn new methods if it doesn't exceed an acceptable level, and if they can see a return on invest. There is no published research that has explicitly dealt with the question of what determines the maximum acceptable level.

2.1.4.5 Imbalance between Evaluation and Generation of new Design Support

A majority of scientists seem to prefer generating brand new tools over evaluating existing tools. This was empirically recognized by CANTAMESSA[49] and again pointed out by ECKERT.[50] The majority of publications at ICED '97 and ICED '99 suggested support, they had developed, while only few publications deal with evaluation. This seems not surprising – considering that design research is done by scientists and engineers, known to strive for new findings. This is of course not true for every design support development project. Nevertheless, even if only a few projects lacked to prove their performance, it would still pose a serious problem for the community as a whole, further diminishing the acceptance in practice for design methods in general. The lacking validation of design methods is commonly criticized and should be fixed. Among the

---

[44] Jorden (1983)
[45] Rutz (1994)
[46] Lindemann (2002)
[47] To the author's knowledge, a detailed description on how to do this has not been published
[48] Birkhofer (2004), 6
[49] Cantamessa (2003)
[50] Eckert et al. (2003)

critics are BLESSING, CHAKRABARTI & WALLACE[51], HORVÁTH[52], ZANKER[53], SCHREGENBERGER[54], JÄNSCH[55], BIRKHOFER[56], and ANDREASEN[57].

2.1.4.6 Questionable Results for complex Problems

Not all methods are published without validation, but a large portion is validated in overly simple design tasks such as a ball pen, or a hole-puncher. There are several reasons to choose simple exemplary applications.

- They are easier to set up. Hence less of the research budget is used up for the validation.
- Less time is needed to understand the design problem (both for the scientist and for the experimentees).
- The scientist might be unfamiliar with real-life design problems as he/she is not a designer himself/herself.
- Realistic application situations can only be found in industrial settings which the scientist might have no access to.
- The examples are designed to later be taught in a classroom setting which demands for quick to deliver design problems the student can identify with.
- With a complex problem, their newly developed method might not work.[58]

Critics are skeptical whether a design method that is presented to them with an exemplary problem, easy to understand, can actually be applied with more complex problems. ALBERS criticizes that too many design methods are developed using trivial and artificial example problems that do not resemble the real-world complexity of a designer's every-day problems.[59] Similar criticism is directed towards validation experiments conducted with the test persons being students. Whether the results of such a test can be transferred to industry is unclear.

Summing up

Due to the empirical nature of all the collected criticism presented above, it is not advisable to regard it as comprehensive. The aspects all address different levels of abstraction. To the author's knowledge, there is no study today that collected the criticism against design methods and/or proposes any operational help for scientists who want to develop design support. However, such operational assistance could be useful for fellow researchers, for example, a checklist in order to determine the

---

[51] Blessing / Chakrabarti (2009)
[52] Horváth (2004)
[53] Zanker (1999)
[54] Schregenberger (1980)
[55] Jänsch (2007)
[56] Birkhofer (2004)
[57] Andreasen (2010)
[58] It should be the scientist's duty to choose a problem likely to provoke the method's failure. However, scientific funding and evaluation systems put pressure on scientists to deliver "successful methods." I will not go into the discussion of such systems. Readers may be referred to the American Journal of Evaluation (http://aje.sagepub.com/) for a comprehensive set of literature as well as to Frey (2007), to Hornbostel / Schelling (2011) or to Gill (1990).
[59] Albers (2010a)

relevance of a new design support or whether it tends to provoke any of the typical criticism that could easily be avoided proactively.

2.2 The Dilemma of developing heuristic Design Support

As shown in chapter 2.1, design support developed by design scientists is not applied as much as its creators intend. The developed methods are often too complex, too theoretical, and not user-friendly while it is unclear whether or not they will deliver the expected performance. Their effectiveness is rarely proven, and their presentation does not meet industry’s likes and preferences.

The simple and straightforward way to react would be to take a step towards industry and try to meet their expectation in order to gain the users’ trust. However, the following paragraphs will show that developing design methods is done by scientists within a scientific community that has very different expectations. This puts design scientists who want to create design methods in a difficult position. As ECKERT ET AL. put it: “[…] design research has two dominating characteristics […] driven by the twin goals of understanding designing and improving it – two goals that require very different research methods.“[60]

After the long list of criticism in chapter 2.1, the following subchapters will show the principal difficulties when it comes to developing design methods. It is not a simple task at all and even though the criticism is based on empirical observation and undeniable, it would be wrong to blame it merely on design scientists’ ignorance or lack of scientific education.

Additionally, the development of design support is after all development. Resources are just as limited as in product development projects. Scientists have to apply for funding and manage to finish their research within the granted time and money. While ideal research might lead to optimal solutions, prosperous design support development, in reality, creates the best solution under given constraints. In short: Successful development is strongly dependent on effective project management.

2.2.1 Design Research has to involve multiple Disciplines

To ensure the success of design support, design scientists need to meet the users’ demands. The questions that have to be answered are:

- Who are the individuals that will use the support?
- What is it that makes them want to use it?
- In what type of situations will they be when they need the support?
- What are their expectations toward the support?

---

[60] Eckert et al. (2003), 1

Sociologists deal with similar problems. Their task is to understand and describe developments, interactions and cause and effect within social groups.[61] At the same time, they don't have access to all of society or the collective they are studying but only to a limited group of individuals who represent and resemble their social group of interest. Very similar, design scientists never have access to engineering design in general. They too have to conduct their research with few individuals. Since sociologists are so familiar with such limitations of research and bias resulting from them, one could ask: "Why not have sociologists do engineering design research?" The answer: "They do not have engineering design experience."

Any type of qualitative research will include interpretation of human behavior. In social research methodology, it is explained that in order to interpret an observed individual's actions, the researcher has to understand the context. The following thought experiment will make the problem more obvious:

Suppose a German male sociologist is interviewing a Chinese woman about her pregnancy. It is a narrative interview situation.[62] The sociologist takes notes about the woman's actions and gestures while a tape recorder records what she is telling him. Although both speak English, there is quite obviously a lot of room for false interpretation concerning her attitude towards certain inquiries. Naturally, one would think: Would it not be better to have someone with her cultural background do the interview if, for example, facial gestures and body language are to be interpreted as well? Would not a woman be more empathetic and know to a greater degree how to interpret some of the comments and reactions? Would not maybe a woman that has experienced pregnancy be better suited for the job? Yes, she would, since a Chinese female interviewer, preferably with the experience of pregnancy, would understand her interviewee's actions in the context of being Chinese and pregnant.

So, in order for a scientist to reveal what works and what does not work, to support a design engineer, it is just the same as with the Chinese woman in the example. In order to correctly interpret what helps a designer and what does not, an observer must have knowledge about the subjective context that might affect the designer. (S)he has to have some type of conceptualized knowledge about using design support and how it affects a design engineer's work and working situation.

People that have experience in engineering design have this conceptual knowledge. However, they generally do not have a degree in sociology and therefore, lack a comprehensive education in human-centered research methods. The result is a lack of strategies and approaches for design researchers to validate their work, which, in consequence, leads to design support that is insufficiently validated.

---

[61] Merriam-Webster's dictionary defines sociology as follows: "The science of society, social institutions, and social relationships; specifically : the systematic study of the development, structure, interaction, and collective behavior of organized groups of human being."

[62] I.e. no closed questions, the woman is encouraged to tell her story through an opening question, compare chapter.

This results in major criticism towards design research. Not only do fellow scientists challenge researchers' works but also practitioners in industry meet new design support with skepticism leading to limited acceptance and thus discouraging further development or improvement of those methods, creating the dilemma of developing heuristic methods.[63]

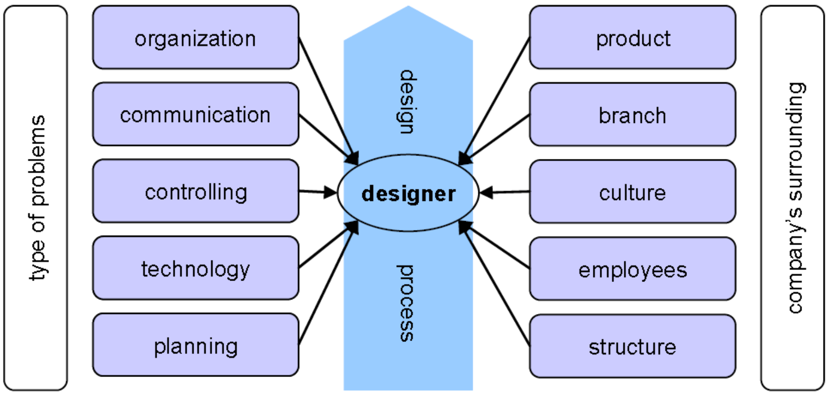


Figure 3 - Influencing factors on designer and design method, Jänsch et al. (2005), 3

2.2.2 Anticipating the User

Furthermore, it is difficult to anticipate the user, even for those with engineering design experience. Designers do not generally want support at all times, but only in certain situations. Which situations call for methodical support cannot be generalized, as it also depends on the designer's personal experience related to the problem at hand, as well as his or her individual preferences. What might be a routine situation for one engineer might be novel and difficult for another because of different experience. But even in difficult and novel situations, it is not always a methodological approach that will help a designer. Redtenbacher went as far as calling engineering design partly an art.[64] However, it is not predictable, when exactly an individual will approach his/her or her problem at hand from a methodological or from an artist's perspective.

BIRKHOFER states that product success is not exclusively depended on methodological procedure but also strongly depends on experience and qualification of the engineer as well as the organizational surroundings. Hence, the success of an attempted support through design methods cannot generally be expected but will only be achieved with certain methods in certain suitable situations.[65]

A lot of work – at least for experienced designers – can be regarded as routine activities with only few situations that are more complex and can thus be regarded non-routine. Even in non-routine activities a designer will not always want support through a method. So, according to BADKE-SCHAUB, it is important to ask: "When does the

---

[63] Compare literatures criticism in chapter 2.1
[64] Redtenbacher (1852)
[65] Compare Birkhofer (1991), 226 ff., cited in Jänsch (2007), 46

designer want to be supported? Which situations do designers experience as non-design situations and how do they deal with these situations?”[66]

It is hard to predict which type of situations will be experienced by a designer as non-routine. BADKE-SCHAUB ET AL. did an interview study with 16 design engineers. “The interviews were designed to find out what kind of situations designers describe as non-routine and in which they felt 'inefficient or ineffective' or 'out of routine'.”[67] The study shows that 'non-routine' is closely related to uncertainty. “Although uncertainty will always be a part of designing, it is only of major influence when it increases to a level that overwhelms the designer and thus adversely affects their performance and prevents them from achieving their goals.”[68] The authors found out that designers’ uncertainty mainly results from either *task-related* uncertainty or *socially induced* uncertainty. I.e. the task is complex, and the designer has to choose an appropriate way to solve the problem at hand, not knowing what the ideal actions would be, or the designer has to cooperate with other individuals but can’t predict all their actions. “In these situations, uncertainty is associated with the interaction between the designer and other team members.”[69] What is most interesting about the study is that only very few situations were identified in which the uncertainty results from perceived limited capability of the designers themselves. According to the findings by BADKE-SCHAUB ET AL., designers tend to use methods in situations of uncertainty, usually caused by the task itself or by their social surroundings, expecting the methods to help them cope with such *non-routine situations* better.

2.2.3 Quality Assessment of Design Support

As the occurrence of non-routine situations is difficult to predict and is strongly influenced by the individual perception of the designer, it is demanding to proof 'that it works', especially for heuristic design support. Subjective judgment is, while important for any kind of creative process, hard to describe, understand, and analyze from a scientific point of view. If it was obvious to a designer that learning and applying a design support is more efficient and therefore, to be preferred over solving problems without the particular support, it is safe to assume the designer would be easily motivated to use it. However, even if the application of heuristic design support seems to be effective, the question remains: Would one have come up with a similar solution without support? Heuristic design support only *helps to find a good solution*. There is no guarantee for optimal solutions.[70] It is principally possible that the same idea or solution is developed, with or without the use of an heuristic design support.

---

[66] Badke-Schaub et al. (2011), 188

[67] Badke-Schaub et al. (2011), 188

[68] Loc. cit.

[69] Loc. cit.

[70] See chapter 9.1.3, page 229

This indicates that the *result alone* – a product, an idea, a solution to a problem – cannot be used to show the effectiveness of a method. Other indicators for the quality of a method can possibly be found when looking closer into the process of its application to compare factors such as the number of alternatives generated, the time consumed to come up with the alternatives, the number of iterations and so on. Furthermore, the personal rating by the users themselves is a possible source of information. Questions regarding how secure a user felt while using the method and how the user assumes (s)he would have felt without the method at hand are subjective. However, combined with further observations they might still provide valuable information. So, it seems necessary and feasible to validate the various aspects of a design support with different approaches. Their combination will allow making a sound judgment about a method’s value for potential users. The main aspects of quality are efficiency, robustness and value to the cost. The core questions for each of those aspects are given in the table below.

Table 3 - Dimensions of quality for heuristic design support

Efficiency	▪ Is it possible to achieve comparable results with or without the use of the method but with reduced effort when applying the method? ▪ Is it possible to achieve better results with the method investing the same effort (time and money)?
Robustness	▪ (How much) do the results differ when applied by different users (changes in the operation system)? ▪ (How much) do the results differ when applied for different goals (change in the system of objectives)? ▪ (How much) do the results differ when applied under different constraints (change of resource system)?
Value to the cost:	▪ Is there a maximum effort that limits the applicability or acceptance? ▪ Could a comparable result be achieved, without the method at question? ▪ Is it ultimately worth it, to invest the time to learn the method?

2.2.4 Rules of Experimentation do not apply - it can only be done once

In experimental studies to evaluate a design support’s quality, some of the parameters are inevitably interdependent. In other branches of science like the natural sciences, an experiment is repeated several times while keeping certain conditions constant. The more often one repeats the experiment and the less the results vary, the more confident the scientist can be about his or her findings eventually deciding whether or not the findings can be trusted to predict what would happen when the experiment was conducted in the future – in other words, high predictive validity (4.3.3.3) is achieved.[71] When experiments include decision-making, subjective judgment and actions of human beings, things get more complicated. Experiments that try to evaluate design support quality simply by repeatedly applying the support, will always encounter some

---

[71] For statistical relevance and confidence see chapter 4.3.2

type of bias. It is impossible to repeat an experiment several times without changing the conditions, as the individuals get more familiar with the context, e.g. a problem to be solved or a design task. Over time, they will focus on a preferred solution inevitably biasing the results as it becomes unachievable to connect the results to the design support. Many design scientists have pondered over this problem:[72] "From a traditional reliability aspect, in order to be reliable, the comparison between two methods should be carried out by the same developers using the same method and developing the same product twice. However, when developing the first edition of the product, the developers will learn *what* to do, which will have effects on the development of edition 2. As 'de-programming' is impossible, the same developers simply cannot compare two methods with trustworthy conclusions, as a lot of uncontrollable aspects may influence the outcome. Neither would the use of varied developers give a comparative situation, as the individuals have different backgrounds, competence, capacity, and so on."[73]

2.3 Do we need Support for Design Support Development?

"Sadly, although design is one of the fastest growing areas of research, the status of research into its own research methodology is, with a few exceptions, poor. In effect, little guidance exists as to how to do design research, leaving it to the individual to find a hopefully efficient, effective and rigorous approach."[74]

What BENDER ET AL. stated above seems to be as true today as it was about ten years ago. The number of exceptions might have risen a bit; however, a general research methodology that is both accepted and even more importantly known to design scientists is still missing[75]. The following paragraphs will review some of the important statements found in literature, which criticize design science in regard to lacking research methodology or rigor, followed by a section that will introduce the approaches BENDER ET AL. called the "few exceptions" as well as those that followed in the recent past in chapter 4.1.

As mentioned, BENDER was one of the strong critics. He also mentions fellow scientists' concerns about the efficiency of design research and the effectiveness of its outcomes. He demands a research method and guidelines that will help to apply and to adapt research methods, and he suggests applying research methods known from the social sciences. However, this should be done carefully because of the following typical characteristics:[76]

- Complex units of analysis while only small sample sizes available

---

[72] Zanker, Bender, Jänsch, Birkhofer, Keller, Andreasen, Blessing, … just to name a few

[73] Björk / Ottosson (2007), 200 f.

[74] Bender et al. (2002), 7

[75] The efforts by Blessing, Chakrabarti, Andreasen and some others are not being ignored here. Their DRM (Blessing / Chakrabarti (2009)) is very helpful for those who know about it or have been to the authors' summer school. Unfortunately it is not well known enough, so it is still lacking general acceptance.

[76] For the complete and original list see Bender et al. (2002), 8

- Lacking theory makes hypothesizing difficult.
- Difficult definition of variables
- Influences cannot be isolated.
- Inter connectivity of influences and variables makes it difficult to identify and determine causality.
- Many variables cannot be observed directly.
- Field testing is often impossible; all cases and participants are different. Industrial practice does not allow for identical tasks or identical situations.
- Difficult to determine influencing factors that are different or the same, because the influencing factors have not been established yet.
- Difficult to form control groups consisting of test persons not exposed to the hypothetical influencing variables; limited motivation of participants when they realize that they do not take part in the ‘real’ experiment.
- Pure experiments involving identical pre-tests and post-test or at least similar tests are practically impossible. Learning and conditioning will bias the results: a design task cannot be done twice.
- Design success as a parameter of high interest is difficult to define, to quantify and therefore, to evaluate.
- Designers as participants are experts with often limited time, motivation, and willingness to give information.

ANDREASEN, BIRKHOFER AND WALLACE have called out for consolidation. In a keynote speech at NordDesign in 2010, ANDREASEN addressed issues such as the inflationary use of keywords as an indicator for lacking consolidation.[77] “We need 1049 unique keywords for 390 papers.” As possible reasons, he gives the extreme complexity of engineering design. He argues that “Most contributions are based on *speculations*, *concepts* and *models* that create a pattern of support for explanation of and support for synthesis – there is an almost *endless number* of such proposals.”

ANDREASEN identifies possible reasons, as to why design researchers do not feel the pressure to consolidate. He states that there are no barriers to prevent them from new speculations and that there is little tradition within this particular community for building on previous contributions. This is again observable in the common reviewing practices which miss discouraging the growth of such speculations. Finally, according to ANDREASEN, researchers in engineering design often miss applying the results of their research in practice and therefore, they:

- do not *acquire direct insights* into best practice.
- do not *appreciate a designer’s mindset*, range of tasks and tools.
- fail to *convince designers* of the value of their ideas.[78]

HORVÁTH also discusses consolidation and suggests a typology to describe design science because in his opinion „Engineering design research shows a rather

---

[77] The presentation to the keynote speech can be downloaded at: http://www.norddesign2010.se/MMA_presentation_NordDesign_2010.pdf; 24/11/2012; 13:06

[78] Loc. cit.

fragmented, if not a chaotic, picture.“[79] BLESSING AND CHAKRABARTI criticize a “ […] lack of scientific rigor, 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 […]”[80] FULCHER & HILLS find design science to be still immature: „Design research […] is new in comparison […] to the established natural sciences." CANTAMESSA adds to that, stating that there is “no specific field of the natural sciences […] from which research methods and tools have been inherited.”[81]

There are still no rules that researchers can play by as in older, more established sciences. ECKERT ET AL. write “In the design research community everybody agrees that design is a highly complex and extremely multifaceted endeavor. There is much fewer consensus about how one should go about studying design and what the aim of any such study should be.“[82] FULCHER and HILLS demand “a clearer understanding of the field as a whole and a sharper definition of goals and agreement on methods.”[83] They strongly believe that this could enhance performance of design research. GILL complains that there are too many proponents and too few exponents within the publishing community due to the pressure on scientists to publish, i.e. too many scientists come up with new ideas instead of working with other’s ideas to validate or falsify them, to improve their research by building upon it. “[…] this scramble to publish is immoral since it creates a 'dust cloud' through which the practitioner is unable to see the small percentage of material that would inform and improve his or her practice.”[84] This criticism has been empirically proven by CANTAMESSA. “At both ICED97 and ICED99, the development of new methods and tools appeared to be the dominant research theme.”[85] HORVÁTH observes that “research into engineering design 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.”[86]

This shows that design science is young, has not yet developed established rules, and this again leads to a flood of publications – most of them suggesting new ideas instead of testing available approaches – within which it is hard to identify relevant articles for oneself since no established language to describe the research results is attainable. As a reaction to such statements, there have been publications of abstract thoughts from a rather philosophical point of view. They debate what the proper research methodologies in design sciences should be, and whether or not they should stick to the “orthodox sciences’ approaches." Furthermore, they deal with the question whether

---

[79] Horváth (2004)
[80] Blessing / Chakrabarti (2009)
[81] Cantamessa (2003)
[82] Eckert et al. (2003), 1
[83] Fulcher / Hills (1996), 184
[84] Gill (1990), 290
[85] Cantamessa (2003), 5
[86] Horváth (2001), 1; cited alike in Blessing / Chakrabarti (2009), 7

design science should be seen as science at all.[87] All conclude that design science does belong to the scientific world and is especially difficult since it has not yet fully matured and there have been parallel streams of investigations, not using a common terminology or a shared methodological framework.

---

[87] Reich (1994), Reich (1995), Fulcher / Hills (1996)

3 Target and Scope

"Most practicing engineers look at design processes as sequences of activities to generate solutions to newly identified needs; sociologists look at design as a socially negotiated process; psychologists as the sum of individual mental processes. However, we all know that design is really all of these things at once."[88]

Three decisive findings from reviewing literature lay ground for the goals of this thesis:

Finding Number I: Design Science is young and its results lack acceptance

Design science – the research that deals with methods to support designers in the centre of product engineering – is a relatively young science. Compared to other, more "grownup" branches of science like physics or chemistry, design science does not have an established and widely recognized scientific methodology for researchers to build their work on. This lack of a tailored methodology leads to a lack of acceptance of its outcomes.

Finding Number II – There is a demand for research methodology in design science.

Literature analysis shows that there is no commonly accepted and applied approach for the development and validation of design support. Established design scientists have articulated the demand for a better research methodology. Chapter 2.3 is titled 'Do we need Support for Design Support Development? Several publications dedicated to research methodology in the field of design science criticize the lack of a common methodology. Therefore, the answer to this question is: "Yes, we do."

Finding Number III – The central role of humans causes need for empirical research.

The difficulties in the development and validation of heuristic design support are caused mainly by human factors. Experiments are not simply repeatable, so their effects usually have to be shown empirically, a problem that is very similar to the difficulties researchers in empirical sociology, and psychology have to face. However, the methodologies for the validation of methods in those fields cannot easily be applied in design science, so it will be necessary to adapt them.[89]

Working Statement

If we had a framework that was widely accepted within the design science community and gave it time to mature, design science's outcomes will be more successful in terms of acceptance due to improved reliability. Such a framework would incorporate methods from design science and other disciplines, which have experience in human-centered research. This way, design science's outcomes could ensure that it captures human needs better, increasing the relevance of its outputs.

The target group of the framework must be scientists with experience in engineering and its processes. Those often did not receive a thorough education in statistics and

---

[88] Eckert et al. (2003), 2

[89] See also Bender et al. (2002)

empirical research. Therefore, assistance must be delivered with the framework on how to apply the methods it incorporates, it needs to be operationalized.[90]

System of Objectives of this Thesis[91]

The goal of this work is therefore, to transfer and match acknowledged methods from (empirical) human-centered research to the approaches already available in design science, see Figure 4. Design science researchers should be provided with practices and lines of action that will expand their operation system in order to develop reliable and relevant design support. The main elements the operation system of design science has to be expanded with within the scope of this work are:

- A preselected set of suitable methods and practices for the development of heuristic design support,
- together with information about the required resources for their application to support project management of design support development projects,
- instructions for their correct application,
- indications to relevant, comprehensive literature, and
- strategies for the selection of a particular set of methods from the pool.

The expansion of the design science's operation system[92] must be in accordance with good scientific practice. This leads to a further requirement: The approach must build on and therefore, be compatible with existing research methodology.

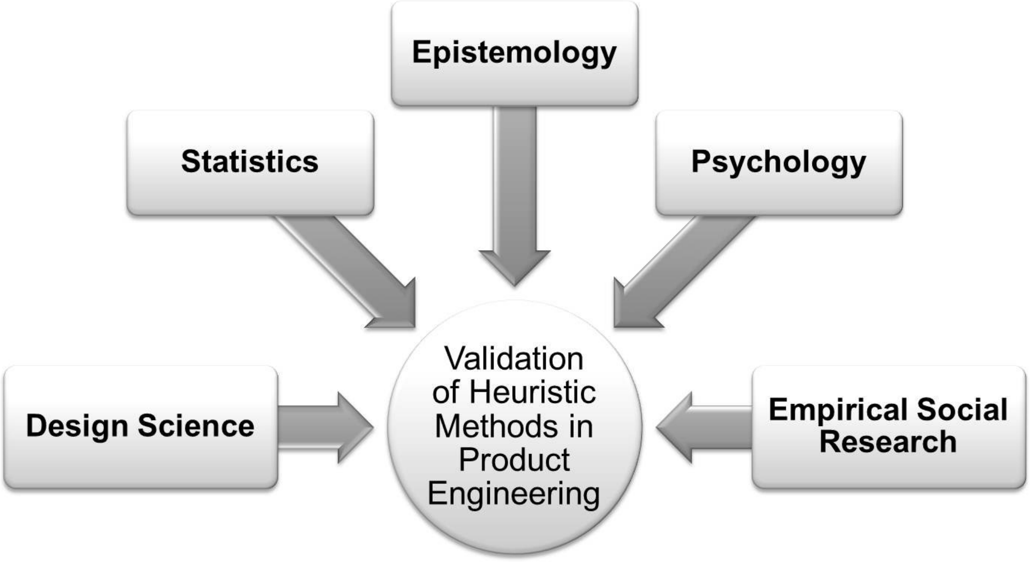


Figure 4: For the development of heuristic design support in product engineering, practices from design science, statistics, epistemology, psychology and empirical social research have to be incorporated into one framework, (source: own illustration)

---

[90] These statements have all been based on and derived from literature. They correspond with the author's observations during five years of working closely with design science researchers. They represent the target group of this framework.

[91] See chapter 5.1.1 for an explanation of the term "System of Objectives" in this context

[92] See chapter 5.1.2 for an explanation of the term "Operation System " in this context

4 Status quo – Research Methods for Design Research

The goal of this chapter is to collect and present what is principally available for the community of design science. This collection is the foundation for the framework later presented in chapter 5.[93] This is the reason why those approaches are found here and not in chapter 2.

The search and selection process for the literature presented here resulted into a number of different layers. First, alternative frameworks and approaches to the one developed in this thesis where identified and are summarized in chapter 4.1. Those frameworks that claim to serve the development of design support are described comprehensively. However, just how short that list is, can be regarded as part of the justification for the decision to develop a new framework.

The main goal of this section was to find and summarize concrete methods for the development of design support. Methods that are explicitly targeted for application to design support development by their authors are typically found in design science literature. Therefore, a majority of the methods in chapter 4.2 originated from within design science. Methods created to be applied for other (or more general) purposes that could be used for design support development are much harder to find. It is impossible to scan all potential literature. For this thesis, the search was therefore limited to references made by the authors of the methods explicitly developed for design support. However, an additional search was conducted in selected fields that could be derived from the criticism against design science and the result it produces (chapter 2). These were management science (which can arguably be placed either within economics or the social sciences) and the social sciences. Some further references were made to clinical research methods in medicine. In addition, critics point out that a lot of design research is statistically weak; hence, a whole sub-chapter has been devoted to selected statistical methods. All these methods are summarized in chapter 4.3. It is a fact that the results of such an approach cannot be considered complete. This thesis does not make such a claim. Instead, the resulting framework is designed as an open architecture. Anyone can improve the framework adding further methods he/she happens to know or has found. In the search for literature, philosophical literature can also be found. Since this thesis' target is to support design scientists and not to contribute to a philosophical discussion, only a small sample of this literature is included. The philosophical discussion about how scientific procedure should be conducted, what is true and what is not, how knowledge can be derived from observation and so on led to different, competing paradigms, a selection of which is presented in chapter 4.4. Apart from these paradigms, the philosophy of science as a

---

[93] As announced in the system of objectives in chapter 3

field of its own was deliberately excluded from this thesis. A short, exemplary list of references from the philosophy of science is given in chapter 4.3.5.

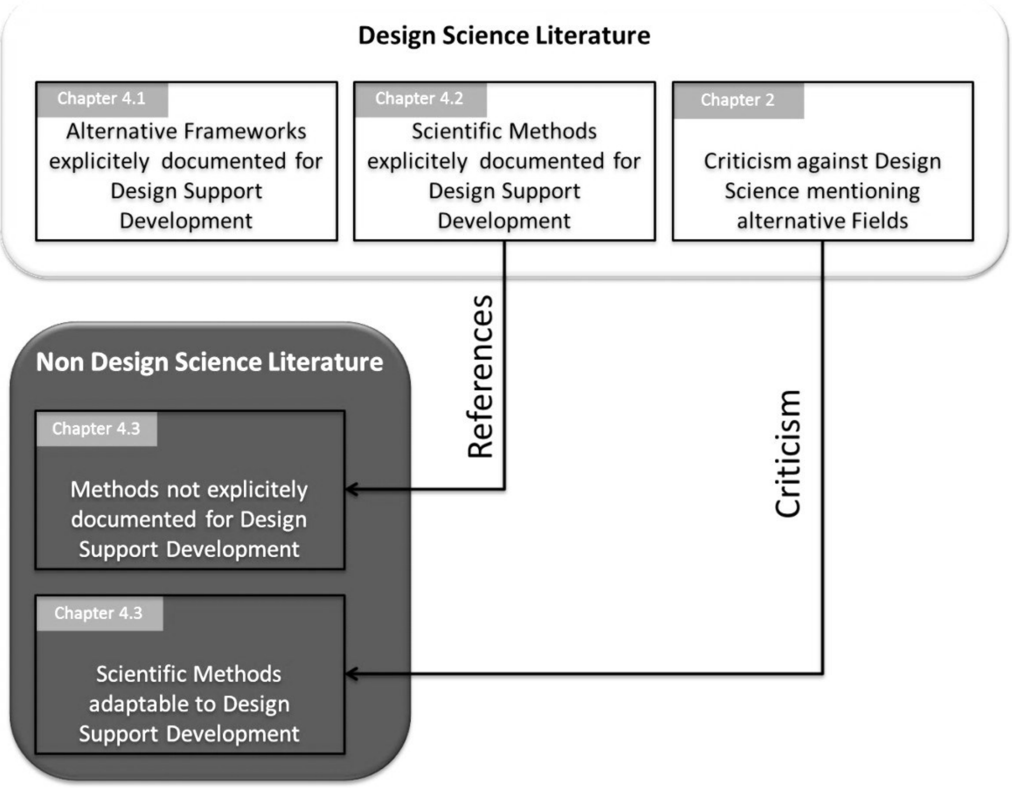


Figure 5 - Search process for the literature presented in chapter 4

The disadvantage of such a “loose” collection of methods is the lack of structure and the lack of an obvious logic that links the methods together. However, at the same time this is what makes the framework flexible for application in various support development projects instead of being only suitable for a very limited category of development projects. So for this thesis, the merits of flexibility were given higher priority at the cost of the lack of structure in the pool of methods.

4.1 Available Approaches within Design Science

While many authors in the field complain about the missing rigor, only few have dared to take a position on how design research should be done. There have been first recommendations for general research approaches and strategies within design science. They will be discussed in this chapter.

4.1.1 DRM: A Design Research Methodology

One of the most comprehensive approaches is the design research methodology as proposed by BLESSING AND CHAKRABARTI.[94] They call their methodology “DRM."

---

[94] Blessing / Chakrabarti (2009)

In DRM, BLESSING AND CHAKRABARTI give an overview on the current state of research methodology in design science. Apart from theoretical background and explanations, they also provide concrete methodological support to a certain extend in terms of checklists and methods such as the “Reference Model” and the “Impact Model." DRM organizes design research in four major stages:

- Research Clarification
- Descriptive Study I
- Prescriptive Study
- Descriptive Study II

An overview of the main objectives and deliverables of each stage is given in Table 4. The steps do not have to be conducted sequentially. "As we indicated […], the example does not show the many iterations and the parallel execution of stages that is 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.”[95]

---

[95] Blessing / Chakrabarti (2009), 17

Table 4 - The four stages of DRM (compare Blessing / Chakrabarti (2009))

	Objectives	**Deliverables**
Research Clarification (RC)	▪ Identify goals, focus, main research problems, questions and hypotheses, relevant disciplines and areas to be reviewed, and the area in which the contribution is expected; ▪ Develop Initial Reference Model and Impact Model ▪ Identify the preliminary set of Success Criteria and Measurable Success Criteria for later evaluation ▪ Provide the focus for DS-I in finding main contribution factors to hinder or prohibit success; ▪ Help focus the PS stage on developing support that addresses factors likely to have the strongest influence on success; ▪ Provide a focus for the DS-II stage for evaluating the effects of the developed support against the goals of the research.	▪ 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

Descriptive Study I (DS I)	▪ Obtaining 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;
	▪ Complete the Reference Model, including the Success Criteria and Measurable Success Criteria;
	▪ Suggest 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;
	▪ 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;
	▪ Provide detail that can be used to evaluate the effects of developed support in the DS-II.
		▪ Completed Reference Model
		▪ Success Criteria
		▪ Measurable Success Criteria
		▪ 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
		  ▪ point at the factors that are most suitable to address in order to improve the situation
		▪ Updated Initial Impact Model;
		▪ Implications of the findings for the development of support and/or for the evaluation of existing support.

Prescriptive Study (PS)	▪ 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; ▪ 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; ▪ Select the part of the Impact Model to address and to determine the related Success and Measurable Success Criteria; ▪ Develop the Intended Support that addresses the Key Factors in a systematic way, and to realize this to such a level of detail that an evaluation of its effects can take place against the Measurable Success Criteria; ▪ 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; ▪ Develop an Outline Evaluation Plan to be used as a starting point for the evaluation in DS-II.	▪ Documentation of the Intended Support: ▪ Intended Support Description: what it is and how it works; ▪ Intended Introduction Plan: how to introduce, install, customize, use and maintain the support as well as organizational, technical, infra structural 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.
Descriptive Study II (DS II)	▪ 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); ▪ Identify whether the support indeed contributes to success (Success Evaluation), e.g., whether the expected impact, as represented in the Impact Model, has been realized; ▪ Identify necessary improvements to the concept, elaboration, realization, introduction and context of the support; ▪ Evaluate the assumptions behind the current situation represented in the Reference Model, and the desired situation represented in the Impact Model.	▪ Results of the ▪ Application Evaluation ▪ 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, customization, use and maintenance issues; ▪ The Actual and Intended Impact Model; ▪ The Reference Model; ▪ The criteria used.

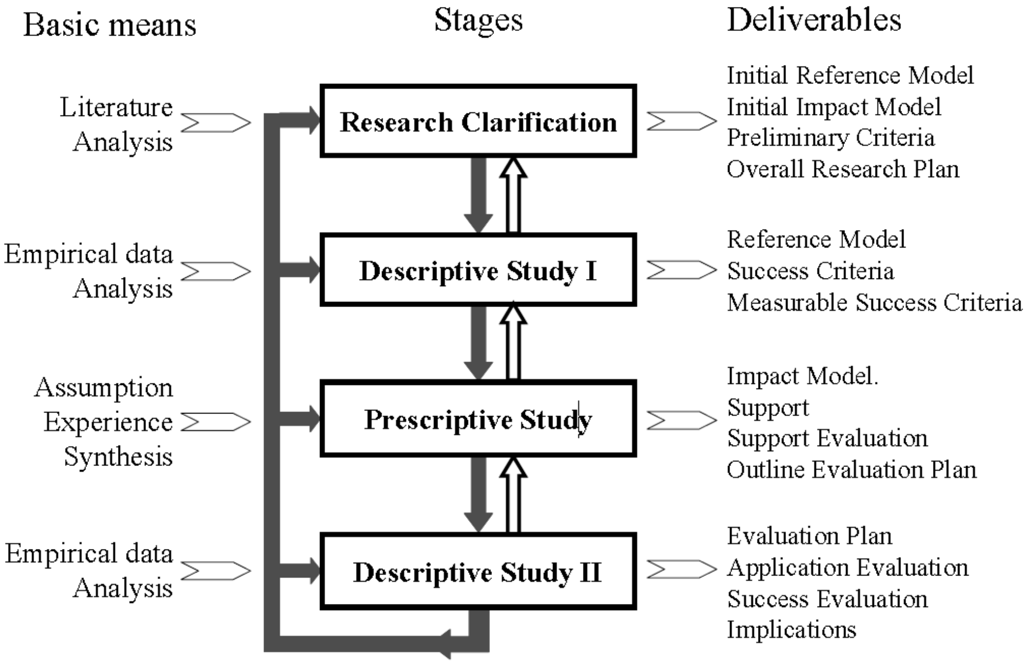


Figure 6 - Basic means and deliverables for the stages of DRM, Blessing / Chakrabarti (2009), 39

4.1.1.1 Graphical representations in DRM: Reference Model and Impact Model

DRM includes two models to visualize the relevant aspects of a research project:

The *Reference Model* which shows the current understanding of the situation as-is, and the *Impact Model* which visualizes the intended situation.[96]

Both models are basic cybernetic models that include influencing factors described with an attribute and an element. If one lacks to describe the influencing factors in such manner, this can lead to ambiguity or even bias the research built upon the model. Another important aspect is that no value should be included within the expression. Influencing factors will be connected resembling statements. Those will include values expressing how one factor (A) relates to another (B), e.g. high A leads to low B or increase of A leads to increase of B and so on. Influences included in either model should be observable, measurable or assessable. The authors warn not to overload the model. It is easy to find yet another factor that also somehow influences the situation. However, the researcher should always ask whether it really makes a difference. As with any model, it is most useful if a suitable degree of reduction is found.[97]

---

[96] For a comprehensive explanation see Blessing / Chakrabarti (2009), 20 ff.; An example of an impact model is shown in Figure 7

[97] I will not go into the details of model theory for which one can read further at Stachowiak (1973a) / Stachowiak (1973b). For further reading about models and model theory in the context of engineering design, refer to Meboldt (2008) or Oerding (2009)

Table 5 - Influencing factors in Reference or Impact Model

	Factor	**Element**	**Attribute**
	Ideation rate	Ideas	Number of
Good example: ➔ Measurable ➔ Quantifiable ➔ No qualitative direction	“The numbers of ideas per session.”		
Bad examples: ➔ “Ho is many” defined? ➔ What is “good” and what is not? ➔ Missing attribute	“Many ideas per session” “Good Ideas” “Ideas” ….		

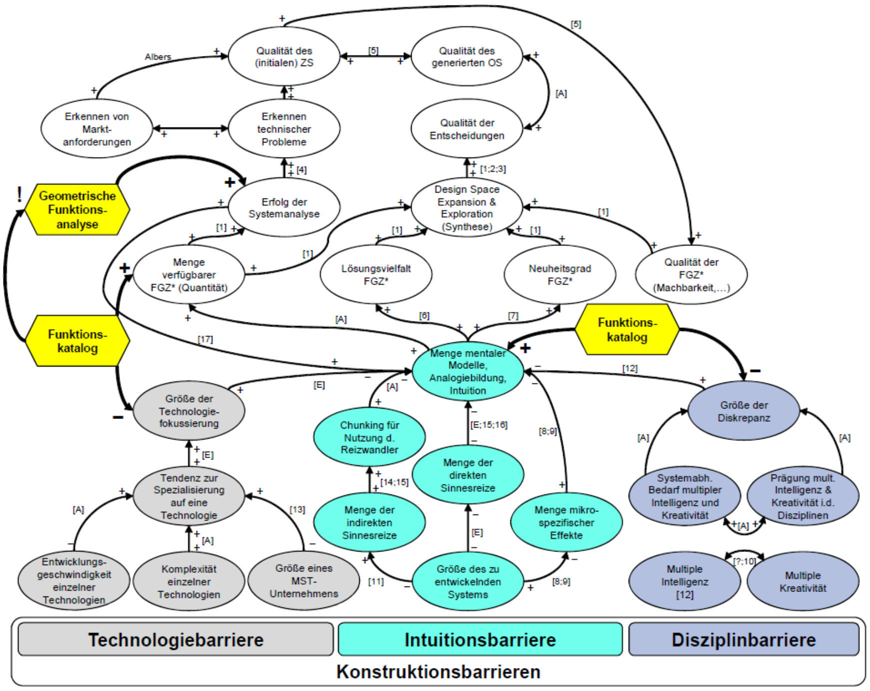


Figure 7 - Example of an Impact Model, Boersting (2012), 128

4.1.1.2 Describing Research Projects with DRM

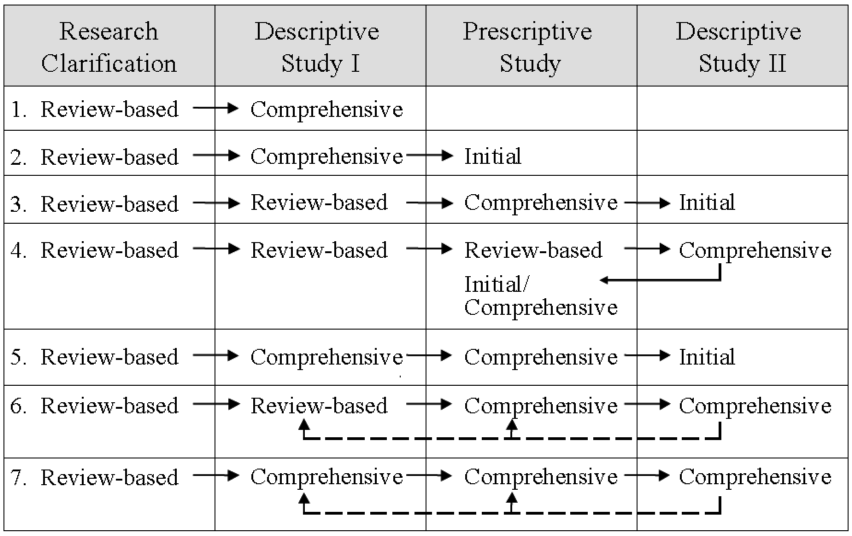
On a rather abstract level, the authors of DRM suggest seven possible types of research one might also call reference processes (Figure 8). They result from three different types of studies that can be conducted in the four stages of DRM: initial, review based and comprehensive (see Table 6).

Table 6 - Three types of studies as proposed in DRM (Blessing / Chakrabarti (2009),

18)

review-based study	A study which is solely based on reviewing literature.
comprehensive study	A study that includes a literature review, but in addition, the researcher produces new findings/observations/results.
initial study	A study that 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.

As shown in Figure 8, Research Clarification is always review-based, while Descriptive Study I can be either comprehensive or review based. The Prescriptive Study can be conducted in any of the three ways, while DS II cannot be review based. Some combinations do not make sense, which leads to a final of seven reference processes for conducting a research project. Reference processes can be used for planning and describing a project. They may be interpreted as a suggestion or recommendation.[98] For planning a project, both research questions as well as available resources will determine which of the seven types suits the problem best. However, DRM does not include any suggestions on how to choose and select between the seven reference processes.



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 8 - Types of design research projects according to DRM, Blessing / Chakrabarti (2009), 18

4.1.1.3 Validation of methods with DRM

DRM does not use the term validation. The authors prefer the term evaluation. However, DRM emphasizes the importance of evaluation and breaks it down into three types of evaluation (Table 7). While there is some guidance as to when, how and why the Reference Model and the Impact Model need to be updated, there is no concrete support on how to set up evaluation experiments for the three stages of evaluation. „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.“[99]

---

[98] Cp. Meboldt (2008)59 f.

[99] Blessing / Chakrabarti (2009), 178

Table 7 - DRM divides evaluation into three steps (compare Blessing / Chakrabarti (2009), 184)

Support Evaluation	“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. […]
Application Evaluation	Application Evaluation, […] aims at assessing the applicability and usability of the support.”
Success Evaluation	Success Evaluation aims at assessing the usefulness of the support, e.g., how successful the support is in achieving the formulated aims.”

4.1.1.4 Success criteria and measurable success criteria

A further important element of DRM is its emphasis on measurable success criteria. “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.”[100]

More important than the question whether qualitative or quantitative research is appropriate is the distinct differentiation between success criteria and *measurable* success criteria. Some types of criteria need to be established in order to compare research results with its goals. Literature suggests typical success criteria for design support development (e.g. 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).[101] Such success criteria are usually difficult to assess. Increased sales, e.g. can only be noticed long after a certain design support during product development has been introduced. Furthermore, such factors at the same time depend on a hugely complex network of other influences. Determining whether a particular design support or some other changes in a company’s environment have ultimately led to a desired effect is a challenge no one has mastered so far. This is the reason DRM’s authors suggest to regard such success factors as a preliminary set of success factors and in a second step to identify those success factors which are actually observable by the researcher within the scope of his or her research. Additionally, factors that have a strong possibility to influence the preliminary success factors and can be directly observed by the researcher can serve as success factors. For example, if positive feedback about a prototype is given by a representative group of potential customers, it can be assumed that replacing an unsuccessful product by such a product will increase sales.

To sum up, in the 2009 version, DRM had evolved for almost two decades. “A preliminary version of DRM was developed as early as 1991 […]”[102]. It is the only

---

[100] Note that the authors of DRM emphasize that by using the term “measurable” success criteria they never intended to reduce DRM to a strictly quantitative research approach. This seems to be a common misunderstanding (cp. Blessing / Chakrabarti (2009), vi).

[101] Blessing / Chakrabarti (2009), 26

[102] Ibid. vi

framework which has been developed specially for design research and it is the first published attempt to deliver a sound research methodology for design scientists. It has many useful aspects and is most likely very helpful for design support development.[103] Although DRM gives a broad background and solid information on good scientific practice in design science, it still lacks sufficient operationalized support for design scientists in the development of design support. It is still relatively young compared to research methodologies in other disciplines and therefore should still be subject to scientific discussion and further improvement.

4.1.2 The Spiral of applied Research

Partly, as a reaction to DRM, ECKERT ET AL. published an alternative model of design research: *The Spiral of applied Research.*[104] "We were to some extent provoked into articulating our methodological position more formally by seeing DRM invoked in a more rigid and naïve form than BLESSING or CHAKRABARTI would use."[105] While it does have some similarities to DRM, readers should be aware that the Spiral of applied Research aims at a different perspective on design research. It describes large-scale research projects or the work of research groups with a long term agenda. It is not useful to try to enforce the model on a single PhD thesis. DRM, on the other hand, seems to address PhD's with their thesis.

However, a single PhD thesis (or similar single study) can deliver one or a few elements (two or three maximum, according to the authors) of a larger research effort. As long as the researchers within a greater research effort agree on the same model such as the Spiral of applied Research, it can be very useful not only to position one's own research within the model but also to pinpoint the research one mentions in literature review as well as activities one recommends for future research. The spiral incorporates and connects eight main stages.

Four fundamental research efforts:

- Empirical studies of design behavior
- Development of theory
- Development of tools and procedures
- Introduction of tools and procedures

For each of the four fundamental research efforts, the spiral also contains an accompanying evaluation stage (compare Table 8).

The strong emphasis on evaluation by separating it from the study itself and hence giving it an equivalent importance as generating new findings is extraordinary. Most of

---

[103] There are no studies to my knowledge that have investigated how useful DRM is.

[104] Eckert et al. (2003); The authors of "The Spiral of Applied Research" refer to the early publications on DRM: Blessing et al. (1998), Blessing / Chakrabarti (2002), Blessing (2002) It could not refer to or criticize the more comprehensive book published six years later (Blessing / Chakrabarti (2009)).

[105] Eckert et al. (2003), 7

the available literature pinpoints evaluation or validation as something like the final step of a study. Another unique feature is that ECKERT ET AL. explicitly include both research that aims at *understanding* design as well as research the aims to *improve* design. Most other authors will side with one of the two aspects. Similar to DRM, the authors do not prescribe a specific order, although they claim that ideal research will follow a clockwise order.[106]

Why is it called a spiral? “In healthy research groups, research on tool building and tool introduction leads to new research questions. All good design research raises as many questions as it answers – we should accept this as a positive force.”[107] The authors emphasize that research questions along with criteria that define whether or not a design support will be helpful can be the result of other research or a first empirical study and must not necessarily be its starting point. “Empirical studies engaging with industrial practice need to be opportunistic. []…we see criteria for the success of tools and methods as desired results of an empirical study, rather than a starting point.”[108] In contrast, DRM assumes that research projects always start with some type of empirical research that is started after success criteria have been verbalized for it.

What is meant with applied research? ECKERT ET AL. do not give a definition of what they meant by “applied research”. However, context and explanations lead to assume that they mean “applied *design* research." It refers to “the twin goals of understanding designing and improving it.”[109] They do not claim the spiral to be suitable to engineering research projects with a strong technical focus like e.g. the analysis of airflow in a clutch housing.

---

[106] Eckert et al. (2003), 7

[107] Ibid. 6

[108] Ibid. 8

[109] Ibid. 1

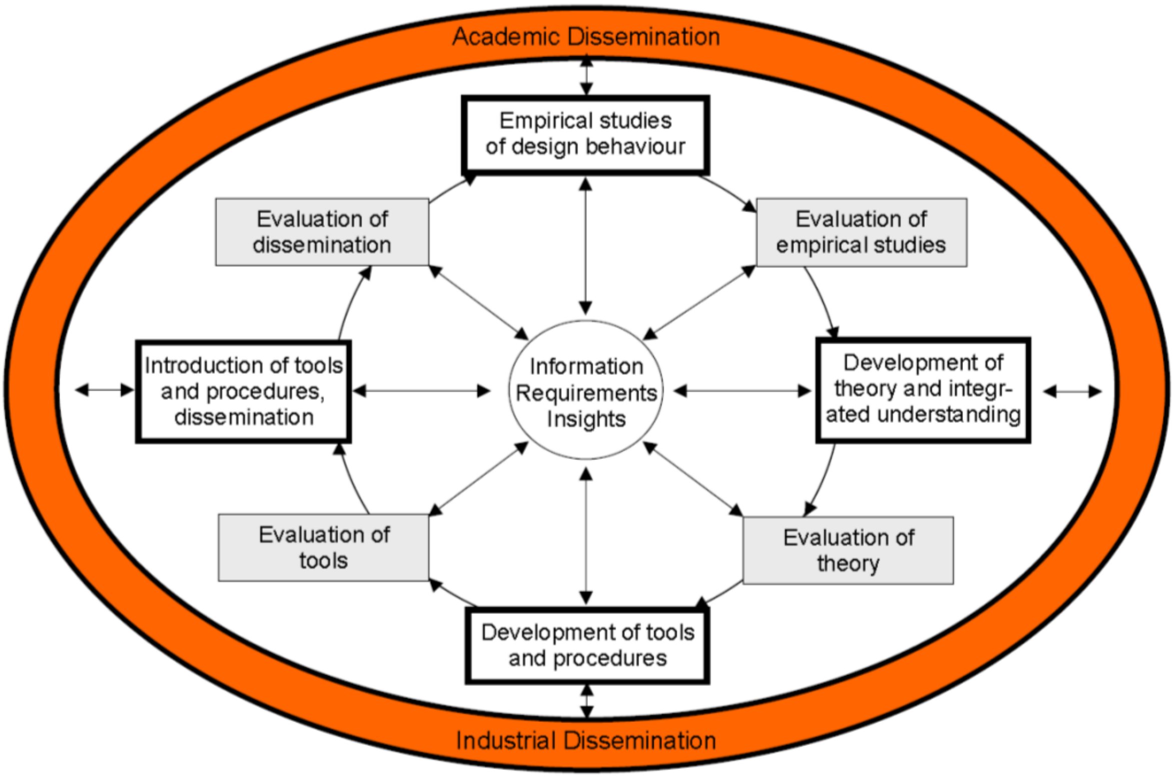


Figure 9 - The Spiral of applied Research, Eckert et al. (2003), 4

Table 8 - The eight stages of the Spiral of applied Research (Eckert et al. (2003), 5)

Fundamental stages	Accompanying evaluation stages
Empirical studies of design behaviour … can include case studies employing different observational methods (ethnography, soft systems methodology, analytical approaches, cross-process comparisons, experimental studies of individual design activities, protocol analysis, …). The results of such studies are findings about how exactly design proceeds in certain conditions.	**Evaluation of empirical studies** This includes assessing the validity of the research results, how far the results can be generalized, how they relate to other studies and how they fit or conflict with theories of design behavior.
Development of theory Empirical research should lead to the development of our understanding of design practice whether this takes the form of theories of aspects of design, mathematical models of processes, theories in contributing disciplines such as psychology, or more local analyses of particular types of designing.	**Evaluation of theory** Theoretical analyses should be compared with existing empirical data, and assessed both in terms of their philosophical and methodological assumptions and their grounding in more general theoretical frameworks, and their relationship to analyses grounded in different conceptual frameworks.
Development of tools and procedures These are design activities that depend on the developers' objectives. As design researchers and software engineers well know, understanding peoples' real needs for procedures and software support is very difficult. Computer tools for designers, and techniques such as design methodologies, thinking techniques and management procedures, will only be effective if they are grounded in a good understanding of the thinking processes and work practices of their users.	**Evaluation of tools and procedures** The development of tools and procedures should be an iterative activity interlaced with evaluation of interim products, as users' and developers understanding of the real requirements change when the users get to test prototypes. And a lot of usability testing is needed to identify and correct glitches and situations where the users do not interact with the system in the anticipated ways. The same principles apply to formal procedures and techniques that designers are expected to learn and apply. The discipline of human computer interaction provides a range of useful analytical techniques.
Introduction of tools and procedures Successful tools and procedures should be tested in serious industrial use. This is dissemination of research and an opportunity to conduct useful research on design practice and the process of introduction as well as the tool itself. In the social sciences studying the consequences of changing how an organization works is called *action research*. *Soft systems methodology* is essentially a procedure for thinking in systems terms about how the participants in a work culture might achieve their goals more effectively and then effecting changes to that culture.	**Evaluation of dissemination** The results of studying the introduction of a tool and its subsequent use can be assessed for validity and for how they fit into our general understanding of design practice.

4.1.3 Cantamessa's empirical Perspective

CANTAMESSA conducted a content analysis of ICED papers published at ICED '97 and ICED '99. He managed to group the content of the published papers into five categories (Table 9).

The first category is empirical research dedicated to observing, analyzing and understanding real-world design processes. This type of research is not aimed at improving design. However, its findings can produce relevant problems in design, research questions and hypotheses on how to improve design. Improvement is usually achieved through tools and methods for supporting design processes, in short design support. Publications which suggest new design support build the second category. CANTAMESSA notes that this is the most common category. The third category deals with design experiments. These usually aim at testing new design support (category II) or testing conclusions from empirical observations (category I). CANTAMESSA notes that the third category seems surprisingly uncommon. Considering that findings from the first two categories *should* be tested, the numbers of publications should be much closer in all three categories. This also applies to the fourth category. It contains publications that deal with the implementation of new design support in real-world design processes. The fifth and last of CANTAMESSA'S categories summarizes all other publications do not fit in any of the first four, such as education studies.

Table 9 - Five Categories of design research according to Cantamessa (Cantamessa (2003))

I	Empirical research, in which researchers analyze real-world design processes.
II	Development of new tools and methods for supporting the design process or elements of it.
III	Experimental research, in which researchers purposely set up design processes in a controlled environment.
IV	Implementation studies, in which researchers discuss the real-world deployment of innovative methods and tools.
V	Other, which includes papers dedicated to theory and education.

Although the empirical foundation is limited to ICED publications, the five categories seem well suited to describe one's own design research on an abstract level.[110] However, while CANTAMESSA'S findings have been cited since their publication, the model has not become standard practice using them to classify, describe or tag research results. Neither have any consecutive studies been conducted to compare the '97-'99 findings to later ICEDs.

---

[110] The author acknowledges this potential source of bias in his conclusion, Cantamessa (2003), 14

4.1.4 Reich's layered Model

REICH deals with research methodology in the context of what he calls artificial intelligence research and engineering design analysis and manufacturing problems, in short AIEDAM. He recognizes the need for a research methodology and points out a problem of engineering design research: It is stuck between two scientific or philosophic world views, namely scientism and practicism.

"First we observe that there are differing viewpoints on the role of AIEDAM research some researchers think that AIEDAM research is about gaining an understanding of some phenomena (e.g. what is design) while others stress the practical relevance of research (e.g. how can we aid design). These differing objectives are originating from the two perspectives of research […] scientism and practicism."[111]

The two worldviews are not the only possible views. It is a matter of philosophy and beliefs. REICH draws from other authors, who have pointed out, come up with, and named many more.[112] In REICH'S view, however, they can all be categorized into two distinguishable groups: "There may be several worldviews of science. There are, however, two worldviews that outline the range of possible worldviews: scientism and practicism."[113]

Scientism stands for scientific inquiry, based on facts and absolutely pure and neutral observation. Scientists devoted to this paradigm can be found in natural science department. Their goal is knowledge. Scientists who believe in *Practicism* aim to improve human actions. Their observations are usually subject to interpretation since they are embedded in some type of context. Typical representatives of this paradigm can be found in the sociology corner of science. Engineering design research as pointed out by others as well is hard to pin to either of the two competing paradigms.[114]

This motivates REICH'S layered model. He claims that research and especially AIEDAM research is a complex matter and introduces two further layers: *Research heuristics* (these are methods for modelling and solving problems in a particular manner) and *specific issues* (s. a. methods for evaluating hypotheses and criteria for such evaluations). According to REICH, sticking to one layer is not useful, because different situations during the conduct of a research project call for distinctive actions on different levels of abstraction (layers of abstraction). However, it is mandatory to make sure that the chosen research heuristics and specific issues are compatible to the chosen worldview. Table 10 gives a brief overview of REICH'S understanding of the two worldviews.

REICH gives a list of literature examples of former combinations of worldview, heuristic and specific issues. However, he fails to give any rule or method on how to decide on

---

[111] Reich (1994), 7
[112] See for example: Guba (1990),Rowan (1981); Smith (1991)
[113] Reich (1994), 5
[114] See e.g. Eckert et al. (2003), Bender et al. (2002), Horváth (2004)

a certain combination or how to justify a chosen combination. He claims human-centered research cannot be combined with a scientism worldview as an example for incompatibility: “Scientism - Human centered: this combination is incompatible.”[115] In consequence, if this should be true and one believes that engineering design research is human centered, this will mean that it is useless to try to defend scientific rigor according to a scientism worldview within one's research.[116]

Table 10 - The two worldviews: Scientism and Practicism (from Reich (1994), 7)

Dimension	**Scientism**	**Practicism**
Researcher‘s relationship to setting	Detachment, neutrality	Immersion
Validation basis	Measurement, logic, reliability, external validity	Experiential
Researcher‘s role	Onlooker	Actor
Source of categories	A priori	Interactive emergent
Aim of inquiry	Universality and generalizability	Situational relevance
Type of knowledge acquired	Universal, theoria, precise, causal, cumulative, reductionistic	Particular, praxis, imprecise, multiple causation, problematic, holistic
Nature of data and meaning	Factual, context free	Interpreted, contextually embedded
Status of science as a field of knowledge	Privileged, progressive, autonomous	Not separated from other fields of knowledge
Value content	Value free	Value laden
Aim of science	Prediction and control	Promotion of human development

Table 11 - Layered model of research methodology (from Reich (1994), 10)

Layer	**Examples**
Worldviews	Practicism
	Scientism
Research heuristics (sources of theories or hypothesis)	Cognitive science
	Decision science
	Formal methods
	Human centered
	Software engineering
	Systems science
Specific issues (evaluation or goodness criteria)	Formal representation
	Parsimony
	Practical relevance

Summing up, one should decide on a worldview (within each research project) and determine whether one’s personal preferences and beliefs are compatible to it and continuously question the research methods one has chosen regarding their compatibility with the chosen worldview. If the research is questioned or even attacked

---

[115] Reich (1994), 15

[116] See e.g.: Albers / Lohmeyer (2012)

by fellow scientists, it might then be quite useful to ask them, which worldview they believe in.

4.1.5 The Validation Square

PEDERSEN ET AL. pose the question "Why [does] an approach solely based on 'formal, rigorous and quantifiable' validation [constitute] a problem?"[117] They review literature on epistemology summing up the essentials of different worldviews:

- The Foundationalist/Formalist/ Reductionist School of Epistemology
- The Relativistic/Holistic/Social School of Epistemology

Famous representatives of the *foundationalist* worldview are PLATO, ARISTOTLE, and WITTGENSTEIN. They believed that objectivity and absolute truth exist. Consequently, showing the truth, making it visible is a rigorous, formal, and logical process, comparable to what REICH calls 'scientism'.

The opposing worldview is the *Relativistic/Holistic/Social School of Epistemology*, in which truth is regarded less static. Truth evolves and is part of a social context. It is not the same as REICH'S practicism, but has some similarities. Famous representatives of this worldview are KANT, HEGEL and KUHN.[118]

Different to REICH'S claim a scientist should stick to one worldview and be careful that the chosen methods are compatible with it, PEDERSEN ET AL. claim that a bit is needed from both worldviews. Internal consistency, however, can be proven in a formal way, based on logic and rigor. Showing that a method works, i.e. showing its relevance for practice, cannot be proven alike and is a matter of context. "*Accordingly, we assert that formal, rigorous and quantifiable validation (i.e., based on logic) can be applied to a design method's internal consistency but fails to validate its external relevance (i.e., its usefulness). Hence, formal, rigorous and quantifiable validation is necessary but not sufficient, and we therefore suggest including the validation of a method's usefulness with respect to a purpose as well.*"[119] The solution PEDERSEN ET AL. propose is what they call *the Validation Square* (Figure 10):

---

[117] Pedersen et al. (2000), 1

[118] See chapter 4.4 for a more detailed view on some of these aspects.

[119] Pedersen et al. (2000), 4

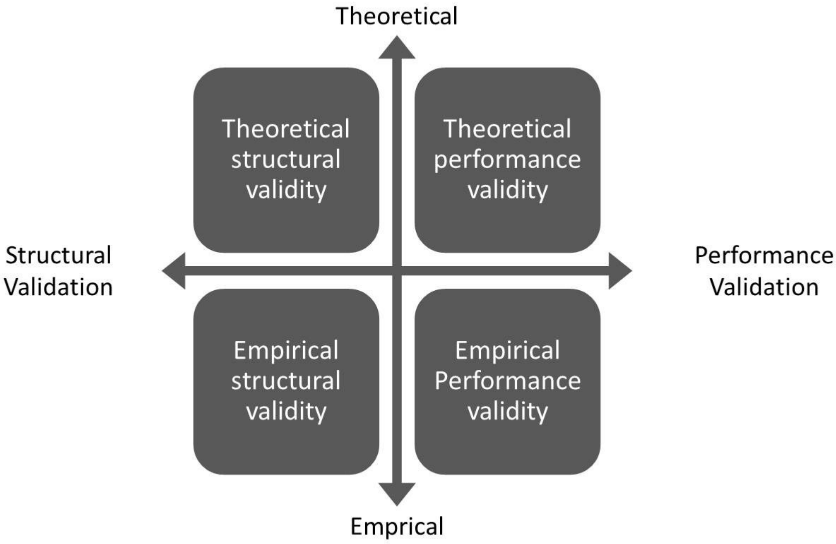


Figure 10 - Validation Square, derived from Pedersen et al. (2000), 5

They suggest establishing credibility in six steps (Table 12). In total these steps cover the four fields of the Validation Square. Although some hints are given as to how one should apply the six steps, some questions arise that are not discussed by the authors. If in step 1, all constructs of a method need to be proven from literature, the validation square is limited to methods that only represent a recombination of steps from other methods. It is not clear, how new constructs should be addressed.

In step 5, the authors suggest to compare with and without each single construct to allow for quantitative evaluation. While this is theoretically correct, it ignores the dilemma of design methods. It is impossible to vary the constructs and leave everything else constant as the participants of the evaluation will grow familiar with the example problem and cannot be switched as this will induce further bias (see chapter 2.2).

Finally, it seems questionable that inductive step 6 will be generally accepted and whether true induction will ever be achievable due to the flaws in step 1 and 5, mentioned earlier. However, users of the validation square should be aware once again that the goal of such a framework model is not absolute truth. It should be left to philosophers to discuss whether or not this is even possible. The goal is to establish confidence. The better one applies the framework, the more likely potential users and fellow scientists will be convinced that one has done valuable work that is worth building on (in the scientific community) or trying and applying it (in industrial practice).

Table 12 - Six steps to cover the Validation Square

	What is to be demonstrated?	**How should it be done?**
1	Show that individual constructs constitute the method as part of: Theoretical structural validity	Literature based: For each step (construct) of your method, show sources in literature. ▘
2	Show internal consistency of the way the constructs are put together in the method. As part of: Theoretical structural validity	Flowchart representations of information- Demonstrate for each step (construct): ▘ ▪ adequate input is available, ▪ anticipated output from step is likely to occur based on input ▪ anticipated output is an adequate input to another step
3	Show appropriateness of the example problems that will be used to verify the performance of the method. As part of: Empirical structural validity	Documentation in stages: ▖ ▪ that example problems represent actual problem for which method is intended. ▪ that example problems are similar to the problems for which the method constructs are generally accepted ▪ that the data associated with the example problems can support a conclusion.
4	Show that outcome of the method is useful with respect to the initial purpose for chosen example problem(s). As part of: Theoretical performance validity	Use representative example problems and apply metrics to make transparent to what degree an articulated purpose has been achieved. ▝
5	Show that the achieved usefulness is linked to applying the method. As part of: Empirical Performance validity	Evaluating contributions of each construct individually ▗ ▪ Compare solutions with and without the construct, → quantitative evaluation ▪ Compare with existing design approaches. ▪ Look critically if rival theories provide alternative explanations for the observed effect.
6	Show that usefulness of the method is beyond the chosen case studie(s). As part of: Empirical Performance validity	Use induction. You have demonstrated that: ▗ ▪ the individual constructs are generally accepted for some limited applications (1). ▪ internal consistency is given for the way the constructs are put together in the method (2). ▪ constructs are applied within their accepted ranges (3). ▪ method is useful for chosen example problem(s), which in (3) are demonstrated to be appropriate for testing the method (4). ▪ usefulness achieved is due to applying the method (5).
Based on this you can claim generality.		

4.1.6 Foundation for the Development of Design Methods

Teegavarapu conducted a cross case study on three projects with the goal to identify patterns and rules for the development of design methods. The project members had to develop design methods alongside an engineering problem to be solved. The research was focused on the following questions:[120]

- Does a systematic method exist to develop design methods?
- What are the user requirements for a design method?
- Do developers of design methods follow a process which is intuitive and tacit?
- Is a systematic method needed to develop design methods?
- Can a systematic design process be used to develop design methods and design tools?
- How can a meta-method be validated?

Teegavarapu observes that neither him nor the participants who were to develop the different design methods found any documented, systematic method for developing design support. However, the participants of his study seemed to follow some type of structured approach intuitively. Teegavarapu believes to have observed a generic sequence of steps taken during the development of design support. Nevertheless, the evidence is rather weak. Two out of the total of three teams followed a similar sequence of steps. Teegavarapu further notes that user requirements were not tracked during development. [121]

While the research seems to be rather prefixed on the idea that case study research is the best method for it is unclear what the author's definition of a design method is. The methods developed in the three case studies seem to be adoptions of existing approaches of varying complexity rather than developments of new methods. Therefore, assumptions about the generality of the findings should be made very carefully. The research does not show clear indications regarding scope of applicability of the resulting "meta-method".

The author observed in his point of view a sharp focus on validation. However, the validation in all three cases was rather a demonstration of feasibility than an organized validation as stated by the author himself: "Validation has been a strong focus in the process of development of design methods […]. Whether or not the developers follow a systematic approach, they are aware of the significance of validating a design method. The choice of validation strategy seems to be consistent across the cases, with demonstrations being dominant […]."[122]

---

[120] Teegavarapu (2009), 29 ff.

[121] Ibid. 204

[122] Ibid. 208

Besides DRM, it is the only such extensive work dedicated to the question on how design research is done. The author shows the high relevance of the topic and shows that case study research – while criticized by many fellow scientists – is a feasible approach for many types of design science research projects.

4.1.7 Horváth's Reasoning Model

HORVÁTH proposes a "gnoseology oriented approach"[123] to give structure to engineering design research: "A conceptual scheme that arranges (and explains) the universe of engineering design research"[124] since it contains a multitude of issues. According to the author, these issues are e.g. "exploration, description, structuring, rationalization, and application of design knowledge and technologies, in combination with the designed artifacts and processes."[125]

Figure 11 shows a visualization of a systems theory approach describing the contents of design science that HORVÁTH uses to argue the complexity of design science.[126] In his model, HORVÁTH arranges design science in three general categories (

Figure 11): *source categories, pipeline categories* and sink categories.

"The source categories of engineering design research are the categories that endow with the fundamental mental capacity for engineering design. The pipeline categories establish links between scientific/ theoretical knowledge categories and pragmatic/ technical knowledge categories by structuring, deriving and dedicating knowledge. The sink categories are concerned with eliciting knowledge that is necessary for the ultimate utilization of the entirety of engineering design knowledge."127 Within each of those three categories, further subcategories can be found (see Figure 12).

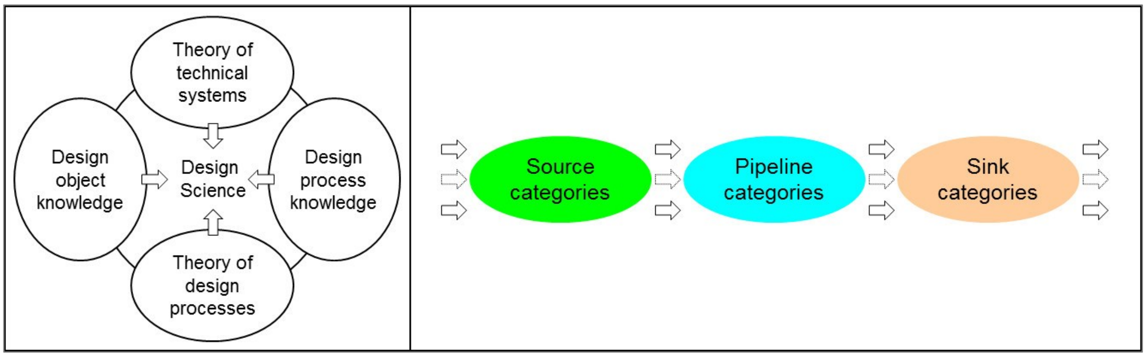


Figure 11 - Knowledge categories of engineering design on the basis of technical system theory (left) and "The natural stream of knowledge in engineering design" (right), Horváth (2001), 2

---

[123] Gnoseology: the philosophic theory of knowledge, inquiry into the basis, nature, validity, and limits of knowledge (http://www.merriam-webster.com/dictionary/gnoseology, 3/12/2012, 20:17) Compare also epistemology: the study or a theory of the nature and grounds of knowledge, especially with reference to its limits and validity (http://www.merriam-webster.com/dictionary/epistemology, 3/12/2012, 20:18)

[124] Horváth (2001), 2

[125] Ibid. 2

[126] Horváth (2004), 156

[127] Horváth (2001), 2

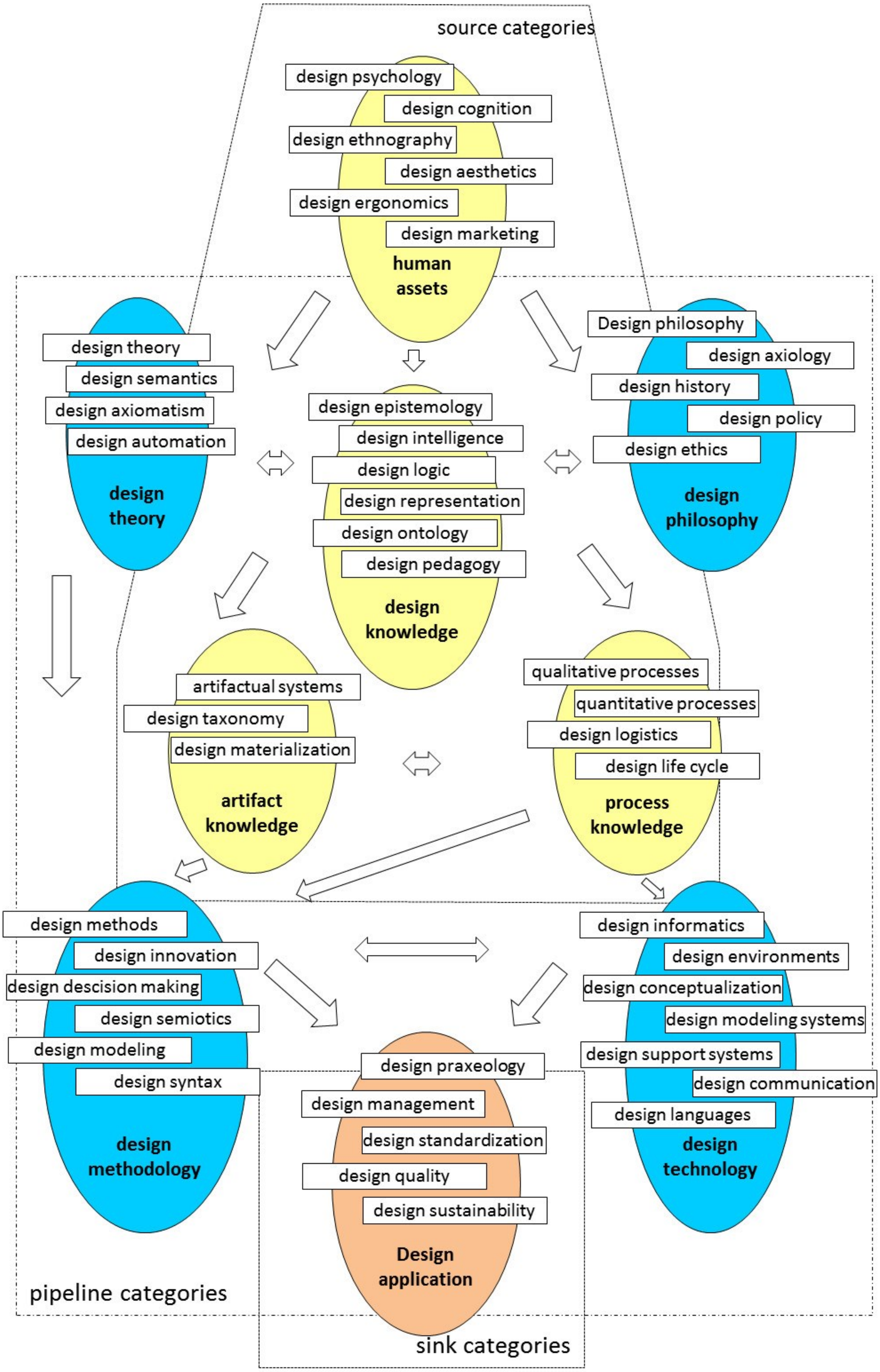


Figure 12 - The Gnoseology-Oriented Reasoning Model showing the Research

Domains in the contextual Categories, Horváth (2001), 3

HORVÁTH'S reasoning model is not normative, i.e. it does not tell a scientist how to arrange his research. However, it can be useful if one wants to describe and or compare his type of research with that of similar or alternative research. It can be used to identify research trajectories or mainstream developments. It can help a researcher clarify what (s)he is about to do during the planning and situation analysis of the early stages of a research project. Hence, it could later be valuable to see whether the research results still fit the domains and categories, one originally aimed for.[128]

4.1.8 Medical Treatment – Design Method Analogy

FREY AND DYM analysed if and how validation methods could be adopted from medical research as they assume several parallels. They call it the "medical treatment – design method analogy"[129]

Table 13 - The medical treatment - design method analogy (compare Frey / Dym (2006), 48)

	Medical research and development	**Design theory and methodology**
What is validated	Medical treatments	Design methods
Entity affected	Human patient	Engineering organization
Outcomes evaluated	Health, side effects, etc.	Quality, time to market, profitability, etc.
Developers	Academic researchers, pharmaceutical companies, etc.	Academic researchers, industry practitioners, consultants, etc.
Professions involved	Medical doctors, nurses, technicians	Engineers, statisticians, managers
Standards for validation	Food, Drug, and Cosmetics Act, and so on	IEEE definition of validation and so on

FREY AND DYM are aware that such an analogy also has many flaws and can easily be argued: "While the association among the entities compared is not perfect, it is worth continuing the effort to extend the medical treatment-design method analogy to compare validation of medical treatments and validation of design methods."[130] They present analogies in five different categories: clinical trials, natural experiments, in vitro experiments, animal models, and theory.

In *clinical trials* in medicine, a closely controlled group of patients is exposed to the treatment to be tested. Usually a control group is given an alternative treatment or a placebo. To avoid bias, blinding is applied, so the patient does not know if (s)he is

---

[128] See chapter 4 for a more detailed suggestion on how to apply Horváth's framework.

[129] Frey / Dym (2006), 48

[130] Frey / Dym (2006), 49

given the real treatment or the placebo. Double blinding can further avoid bias as it remains unknown to the medical researcher which patients are given which treatment or placebo.

The design method analogy would be a group of designers (“patients”) given a design method and a control group which does not have the design method both treating the same design problem (“disease”). Difficulties arise concerning blinding. The designer must actively apply the method, hence, (s)he will always know whether or not she was in the placebo-group. How do you find participants (such as companies) that will voluntarily be in the placebo group? While in medicine, a naturally-occurring problem – the disease to be treated – is given, how can an adequate design problem be chosen for a clinical design study? How do you achieve tolerable sample sizes – for several companies, designers, methods and products? Who pays for it?

To cope with some of the problems, using *surrogate variables* is suggested. Surrogate variables are applied whenever effects in medicine cannot be observed directly or in acceptable time. In such cases, the researcher looks for variables that are thought to affect the desired outcome. For example, it is believed that high blood cholesterol levels increase the long-term risk of heart-diseases. If a treatment has a positive effect on the almost immediately measurable cholesterol levels, the treatment is believed to lower the risk of heart diseases. In this example, cholesterol level would be the surrogate variable. However, there are examples where the application of surrogate variables was misleading and eventually no long-term effects of the treatment could be shown. For decision support tools, OLEWNIK AND LEWIS proposed criteria for evaluation. That could be viewed as surrogate variables.[131] For other types of design support, no such list exists. Furthermore, the correctness of OLEWNIK’S AND LEWIS’ list has not been shown.

Similar to clinical studies are *natural experiments*. Those experiments are not controlled, i.e. no special treatment is compared, but the natural occurrence of variables is observed. Examples are smoking habits and the occurrence of lung cancer, weight and the occurrence of heart diseases and so on. Since many factors can bias such an observation and the groups are not randomized, statistical relevance is much harder to establish. The answer to the question whether smoking and cancer are related or not took a lot more than one study.[132] The application of natural experiments is not uncommon in design science[133]. However, it is limited to those methods that have already been accepted in industry. The effectiveness of FMEA, for example, might be subject to a natural experiment. The same problems arise as in medicine. A direct link is hard to establish. FMEA is recommended in the ISO

---

[131] Olewnik / Lewis (2005)

[132] To the knowledge of the author no clinical study was ever conducted that would have forced people to smoke or quit smoking. The smoke- cancer causality is all based on natural experiments.

[133] See for example the studies about the use of systematic design methods in industry that were quite popular in the early 1990s such as Birkhofer (1991) Beitz et al. (1992), Lohse (1993),

9000/9001 as an adequate method to ensure quality.[134] A natural experiment that would try to establish a link between the application of FMEA and quality of companies' design processes, production processes or products would probably come to the conclusion that among companies applying FMEA, quality is at a more mature level. What such a statement neglects is the fact that in a group of companies, which apply FMEA regularly, the number of companies that also apply many other tools to improve quality is higher: These are all ISO 9000/9001 certified companies. As in medicine, several studies would improve the reliability of the statement.

In vitro experiments are applied in medicine, when treating a human (in vivo) is not (yet) justifiable due to risks of harming the person.[135] It may, however, be an option to treat human tissue or cells outside the human body. Possible analogies in design would be:

- "The subject of the experiment is removed from the usual context, in this case, the corporation where most authentic engineering practice takes place.
- Cooperation of an entire engineering enterprise is generally not needed. The human subjects can volunteer individually.
- Closer observation and control of experimental conditions may be possible."[136]

Looking at today's design research, many studies can be compared to in vitro studies. They are often conducted with single probands, isolated from the time pressure, resource limitations, and interdependencies encountered in an industrial company.

Another analogy is the comparison with *animal models*. Animals like mice or rats are used for controlled experiments instead of humans. The control group can even be manipulated. In case of mice e.g. there are genetically manipulated exemplars where certain genomes have been deactivated – so called "knockout mice." While using animals is not applicable in design research, one might go so far as to compare experiments conducted in classroom settings to animal models.[137] Similar reductions from reality are made and related difficulties arise. To what extend experiments with students can be generalized to a real-world setting is still unclear. Assumptions about a method in real practice from student experiments can only be made with lots of experience. The same difficulties apply to findings about how certain mice react to certain treatments under laboratory conditions.

Finally, FREY AND DYM point out that in medical treatment research, sometimes effects cannot be explained with the basic theories (like chemistry or biology). They quote that

---

[134] DIN EN ISO 9000 (2005); DIN EN ISO 9001 (2008)

[135] Latin: "in vitro" - in glass; "in vivo" - within the living

[136] Frey / Dym (2006), 51

[137] I do not intent to compare students to animals, neither do Frey and Dym

psycho-stimulants have a calming effect on humans with ADHD[138] while nobody knows why, their positive effect could be shown in clinical studies and therefore, psycho-stimulants are being prescribed to patients with ADHD. Similar, engineering has encountered situations where the working understanding has preceded the fundamental explanations. Steam engines have been around longer than the thermodynamic models needed to explain why and how exactly they work.[139] Although this is against scientific rigor, it leaves the idea of "Just because you cannot explain it, it does not mean it won't help." While such a thought should be allowed, one has to be careful not to use it as an excuse for sloppy scientific work.

To sum up their analogy, it can be stated that FREY AND DYM do not deliver anything applicable, a framework or a set of rules that can be "borrowed" from medicine and simply applied to engineering design research. It is an interesting game of thought. Keeping medical researchers' methods in mind, and trying to understand them, can still be useful for a design researcher, especially when he/she comes to a point, when certain effects of a design support cannot be explained.

4.1.9 Method Impact Evaluation

STETTER and LINDEMANN deal with the evaluation of the impact a method has.[140] They present a layered framework that includes more detailed approaches from various authors. They list five steps and argue that the completion thereof will, in sum, lead to impact evaluation. The five steps are listed in Table 14 .

The authors applied the framework in a case study. The following quotes summarize some of the findings relevant for this thesis: "During all phases of the method implementation process, it was observed that the chance that a method or tool would be used in the longer term was greatly increased if it was initially promoted by someone who was deeply convinced of the benefits of the method."[141]

"A central success factor was trust – the designers in the product development department needed to trust the external method sources. They needed to believe that the proposed methods were suited to their actual situation and problems that they were easy to use, and that they would have a positive impact."[142]

While this is relevant and interesting for both industry and science, it does not address the problem of establishing confidence in a new design method when it is being developed to promote its implementation in industry but assumes that the method has

---

[138] Attention deficit hyperactivity disorder - "Attention deficit disorder (ADD) and attention deficit hyperactivity disorder (ADHD) are characterized by the inability to self-regulate focused attention. Children with hyperactivity are impulsive and behaviorally disinhibited. The condition is developmentally disabling which, if left uncontrolled, persists into adolescence and adulthood." Barabasz / Barabasz (1995), 1

[139] Frey / Dym (2006), 52

[140] Stetter (2000), Stetter / Lindemann (2005)

[141] Stetter / Lindemann (2005), 449 f.

[142] Ibid. 450

already been accepted. The question, how one should proceed in order to evaluate the impact remains unanswered.

Table 14 - Evaluation of method impact (compare Stetter / Lindemann (2005), 447; notes of sources are from the author)

Step	**Literature sources for additional information**		**Notes on the sources**
Assessing the product development productivity	Concept of product development productivity:	Duffy (1997)	Debate on "What is design productivity and how can it be measured"
	Problems complicating the evaluation of the impact:	Wildemann (1993)	Discussion about time pressure on R&D departments.
		Reichwald / Conrat (1995),	Relate cost for design changes to design quality
		Giapoulis (1998)	Thesis on models for successful design processes
Estimation of the impact using indicators	Networked efficiency thinking:	Weichselbauer (1996)Weichselbauer (1996)	Controlling instruments to evaluate and implement organizational restructuring.
Developing indicators	Goal-Question-Metric (GQM):	Basili / Rombach (1988)	Findings from software engineering and according process models
Using qualitative criteria as indicators	Qualitative criteria:	Weichselbauer (1996)	Controlling instruments to evaluate and implement organizational restructuring
Using quantitative measurements as indicators	Quantitative measurements:	Fromm / Haist (1989)	Collection of techniques to improve organizational quality in companies
		Briand et al. (1998)	Linking qualitative measures for software maintenance to organizational structure of software engineering companies.
		Usher (1996)	Concurrent Engineering for SMEs

4.1.10 Validation of Decision Support Tools

Olewnik and Lewis introduce a model for the validation of decision support tools which they base on reasoning and literature findings.[143] They reason that in order to be valid, a decision support has to fulfill three criteria (Table 15).

Defining "logical" as "in accordance with intuition" (Criteria I) conflicts with other authors, who criticize that intuitive test cases are being applied to evaluate complex design support. It is not quite clear whether this is Olewnik and Lewis's intention or whether "logical" refers to something like a pilot study, a pre-check before more complex problems are addressed with a design support. Summing up, the three criteria

---

[143] Olewnik / Lewis (2005)

seem helpful to keep in mind as a general guideline but the authors do not offer suggestions on how the criteria are made operational on a concrete level.

Table 15 - Criteria for validity of decision support tools (Olewnik / Lewis (2005), 112 f.)

It must be logical	The results that come from the model make sense with the intuition. Testing can be accomplished using test cases for which the results are intuitive and checking if the model results agree with intuition.
It must use meaningful, reliable information,	Information incorporated into the model should be meaningful in the sense that it provides insight into interdependencies among system variables and reliable in the sense that the information comes from appropriate sources. Another important consideration regarding the reliability of information is the level of uncertainty associated with it. Understanding the uncertainty in information leads to a better understanding of the possible errors in the achieved results and gives a feeling for the level of confidence one can have in the results.
It may not bias the designer	No matter the methodology, the preferences of the designer utilizing the methodology should not be set by the method itself. Forcing a preference structure on the designer parallels the notion that the process used in decision making can influence the outcome. Rather, the decision method should allow to use own set of preferences.

4.1.11 Requirements for Method Development

KELLER and BINZ have presented a list of requirements for design methodologies.[144] They have based their list on a literature review, identified the items shown in Table 16, and ranked them by counting interdependencies. The top five requirements that interrelate with at least two others are: objectivity, reliability, comprehensibility, learnability, and efficiency. Table 16 can be useful as a general checklist for design support development.

Table 16 - Requirements on Design Methodologies (Keller / Binz (2009), 2-205)

Requirement category	Grouped items
Revisability	Validation; Verification
Practical Relevance & Competitiveness	Innovativeness; Competitiveness
Scientific Soundness	**Objectivity; Reliability;** Validity
Comprehensibility	**Comprehensibility;** Repeatability; **Learnability;** Applicability
Usefulness	**Efficiency;** Effectiveness
Problem Specificity	Problem Specificity
Structure & Compatibility	Handling Complexity; Problem Solving Cycle; Structuring; Compatibility
Flexibility	Flexibility

---

[144] Keller / Binz (2009)

4.1.12 General Advice for Conducting (Design) Research

BENDER ET AL. give a list of general advice to follow conducting design research. While the list is not operational in a sense that it can be applied as a design support development method it is still useful as a summary of basic rules to follow in design science in general. A researcher who follows these rules will very likely produce valuable results for which that (s)he can establish confidence.[145] For more convenient application of the rules, they have been summarized in a worksheet (Worksheet 1 : General advice for design researchers), which can be found in the appendix (9.2)

4.2 Methods in Design Research

In the underlying system of objectives[146] this thesis (Chapter 3), one of the necessary elements for the expansion of the design scientists' operation system[147] was argued to be a preselected set of suitable methods and practices for the development of heuristic design support along with indications to relevant literature. The following chapters will present such a collection. Secondary targets also formulated in the system of objectives are:

- Information about the required resources for the application
- Support for project management of design support development projects,
- Instructions for their correct application,

This chapter, therefore, summarizes a collection of related work that has originated from within the engineering design community. They have been designed for specific purposes, so none of them address the development of design support holistically. All the work presented within this chapter is strictly related to design research. Methods from other disciplines will follow in 4.3. In order to address the secondary targets, each method’s characteristics will be summarized in a method profile under the following viewpoints:

- Field of application within design research
- Answers questions of the type
- Advantages/Disadvantages
- Necessary preparation and follow-up work
- Related techniques
- Synonyms

“Field of application within design research” and “Answers questions of the type”

The development of design methods and techniques applied in this context can be of various flavors. There are content related techniques such as the quantification of quality criteria of design outcomes. On the other hand, there are process-related

---

[145] For the complete list, refer to Bender et al. (2002), 14 f.
[146] See 5.1.1 for a definition
[147] See 5.1.2 for a definition

techniques that focus on how things are done and less on the results. Since comparing such different techniques is neither easy nor convenient, a brief orientation is given for each method, in which types of research situations it is most useful. Some exemplary questions are formulated to make the field of application more obvious to potential users.

“Advantages/Disadvantages”
In most cases, more than one method will be available to the researcher. In such situations, an overview of advantages and disadvantages is useful to help decide, hence addressing the demanded support for the selection of suitable methods from the pool of available methods.

“Necessary preparation and follow-up work”
In any kind of research, resources are limited. Scientists must have an idea of the necessary effort connected to the application of a certain technique. This not only helps decide if more than one method is available but also supports project management as has been demanded in the system of objectives in chapter 3.

“Related techniques” and “Synonyms”
This section will hint to similar and related techniques to help search for potential candidates and generate a variety of suitable methods for design scientists when planning and managing design support development projects. Many techniques that are essentially the same come with different names. This is criticized as a weakness, however, the problem exists. To assist further reading and make the search for suitable methods easier for the readers, the synonyms for each method are listed in this section.

4.2.1 Evaluation based Task Design

As has become apparent in chapter 2.2, in the context of design research, experiments are easily biased and very difficult. Researchers, therefore, have to pay close attention to possible influence to keep bias to a minimum. A major influence in experiments is the task given to the participants. It is hence important to design it according to the goals of the study at hand. BENDER has adapted *test design* from psychological tests to the problem of *task design* in design science. He addresses the following questions:[148]

- “How can validity and comparability of experiments be increased in empirical design research?
- Which *objectives* shall be met by design tasks for empirical research?
- Which *demands* have to be fulfilled by experimental design tasks?
- How can appropriate tasks be designed deliberately?
- How can design tasks (not only) for empirical research be analysed and categorized?”

---

[148] Bender (2003), 400

He concludes that researchers should apply fundamental quality criteria as they exist for test design as suggested by LIENERT AND RAATZ:[149]

- The real-life context of the task should be reproduced in a laboratory design as closely as possible to enhance *empirical relevance of the research*. Here the support of professional design experts is recommended.
- *The task must be appropriate* to measure what it intends to and therefore has to be adapted to the scope of research carefully (e.g. to the design stage to be investigated).
- Setting up precise *performance criteria* and evaluation procedures makes sure that good design results can be distinguished from poor ones."[150]

Also adapted from LIENERT AND RAATZ, BENDER gives a set of six demands, a good task design should meet summarized in Table 17: Objectivity, Reliability, Validity, Empirical relevance, Adequate difficulty and Efficiency.[151]

BENDER points out that adequate difficulty is hard to predict for a researcher as it will be subjectively experienced by the participants of the design experiment.[152] He therefore suggests applying an objectified scheme for the evaluation of a task's degree of difficulty. Such a classification scheme has been developed by RÜCKERT and SCHRODA for use in design practice they use six criteria for the evaluation as shown in Table 18.[153] SCHRODA provides a questionnaire for the assessment of these criteria.[154] A worksheet has been derived from the questionnaire (see 9.2).

SCHRODA'S questionnaire for design task evaluation uses a five-point ordinal scale for all categories, reaching from very weak to very strong or from very few to very many.[155] Consequently, this method will not allow a quantitative value, characterizing a design task. However, in most cases, a researcher will try to show that the chosen design tasks for the validation of an heuristic design support are comparably difficult and hence the degree of difficulty as a source for possible bias is brought to a minimum. For this goal, an ordinal scale is absolutely relevant and sufficient. SCHRODA'S questionnaire includes six categories, one of them being the dynamics of a task. While this category is extremely relevant in engineering design, it is not for engineering design science. A researcher planning a task in order to validate a design support should avoid changing conditions during the validation anyways. Therefore, this category is left out.

It is therefore suggested to applying design task evaluation as a combination of BENDER'S AND SCHRODA'S approaches. The general process would be to design a

---

[149] compare Lienert / Raatz (1998), 29 ff.
[150] Bender (2003), 406 f.
[151] the original list is from Bender (2003), p.401 ff., drawing from Lienert / Raatz (1998), p. 29ff.
[152] Bender (2003), p. 403
[153] Rückert et al. (1997); Schroda (2000)
[154] Schroda (2000), 160ff.
[155] Ibid. 160 ff.

series of tasks based on the general guidelines from Bender and then assessing them in order to determine whether or not the tasks are similar in their degree of difficulty. If some of the tasks are noticeably easier or harder than the others, they can be varied and reassessed until they achieve the desired degree of difficulty.

Table 17 - Six quality criteria for a task design (compare Bender (2003))

Objectivity	Test evaluation should lead to the same results when different persons evaluate its outcomes. Therefore, valid methods for the assessment of test performance have to be applied (e.g. value analysis).
Reliability	*Re-test reliability:* A test has to be formulated such that the *same test person* being confronted with *the same test twice* understands it identically and achieves the same results. *Parallel-test reliability*: A test person confronted with different versions of the test must achieve similar performance (e.g. test score). Design tasks cannot be performed twice by the same person.[156] For a longitudinal approach, one therefore has to focus on parallel-test reliability by creating different tasks of *comparable characteristics.*[157]
Validity	*Perceptibility:* It is important that good test performance can be distinguished from poor test performance with sufficient certainty. Design task have to allow the formulation of *precise and operational performance criteria*.
Empirical relevance	*Transferability* (from laboratory to practice): To ensure objectivity and reliability of test results, a 'synthetic' design task for a laboratory study has to be designed in such a way that adequate observation of variables is possible, while at the same time being as near to practice as possible for optimum transfer of results.
Adequate difficulty	To sustain the motivation of participants, a task has to be designed with adequate difficulty: *Do not ask too much of the participants but don't be too trivial!* The task must be formulated so that the participant can cope with it within the scheduled period of time; with his/her individual qualifications (knowledge, faculty, skills) with the provided resources. Verification of this fundamental requirement for tests is subject to a pilot study.
Efficiency	Test design has to ensure that a sufficient number of potential test persons is willing to participate, although they have limited time to spare. The expected amount of captured data must be kept manageable.

---

[156] compare chapter 2.2.3

[157] 0contains a worksheet based on the works of Schroda that will help in characterizing design tasks.

Table 18 - Schroda's six criteria for task evaluation (compare: Schroda (2000), 41 ff.)

conflicting aims	▪ overall number of aims ▪ number of conflicting aims ▪ strength of the conflict
complexity	▪ number of sub functions ▪ number of relations between the sub functions ▪ strength of the relation
transparency	▪ availability of information on the initial status ▪ availability of information on boundary conditions
degrees of freedom	▪ number of potential solution variants ▪ number of potential solution paths
dynamics	▪ variability of the initial status ▪ the predictability of decisions ▪ the predictability of interventions ▪ external influences
necessary knowledge	▪ required subject-specific knowledge, ▪ required problem adapted procedures ▪ required common strategies for problem solution

Table 19 - Profile of ”Evaluation based Task Design”

Evaluation based Task Design	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment.
Answers questions of the type:	▪ Is the design task appropriate to evaluate the design support? ▪ Are the design tasks for repeated tests or different design teams comparably difficult?
Advantages	▪ Reduces bias resulting from different design tasks. ▪ Can be done by the researcher without a pre-study ▪ Can lead to a “database” of common design tasks.
Disadvantages	▪ Only qualitative comparison between tasks is possible.
Necessary preparation	▪ For assessment, the task design must be complete and detailed.
Follow up work	▪ If degree of difficulty differs between tasks, redesign and reassess the task
Related techniques	▪ Test design (psychology(sociology) ▪ Creativity assessment
Synonyms	▪ Task Design; Task evaluation; Test design

4.2.2 Creativity and Creativity Assessment

Creativity is considered a vital factor for innovation success. KANTER writes: “Creative ideas are the raw material necessary for innovation, and a strong competitive advantage is conferred upon organizations that are adept at eliciting creativity from their employees.”[158] Consequently, a lot of effort is put into the development of design support supposed to help designers exploit their full creative potential.[159] In order to

---

[158] Kanter (1988), cited in Audia / Goncalo (2007), 3

[159] See e.g. Maier et al. (2007)

evaluate how successful such a design support is, creativity assessment is a useful approach. Before the details on creativity assessment are presented, it is important to clarify the concept of *creativity* itself.

4.2.2.1 The Term Creativity

Creativity is the ability to produce novel and valuable ideas.[160]

The above definition – including small variations of it – is widely agreed upon and common. Still, it does not mean that this is the ultimate definition. Perhaps it is so popular because it is simple and easy to remember. Digging deeper into the topic reveals a confusing and multi-faceted state of research. The following paragraphs will present some different aspects of the term creativity.

For Psychologists, creativity is a *human trait* of great interest, a phenomenon that allows individuals to generate new ideas. However, not every person shows this trait to the same extend or under the same circumstances. What makes people creative? In which situations are they creative? Does creativity correlate with other human traits? These are typical questions coming from psychologists and determining the nature of research within the field.[161]

Educators take a different interest in the topic asking: “How creative are my students?” “What actions can I take to make them more creative?” “Are there gifted students in my classroom that demand special attention?” “Has the implementation of a certain program any influence on the creativity of my students?”[162] It is important to keep in mind that in creativity research, 'education' mostly refers to the education of children from pre-school to high-school level. However, in management and design education, for example, creativity can play an important role. In the Karlsruhe Education Model for Product Development by ALBERS, creativity is one of the five high level teaching goals side by side with elaboration, professional skills, methodological skills and social skills.[163] In management science, creativity is studied as a valuable asset leading to invention, innovation, and finally economic growth.[164] Human-resource managers might ask: Is a potential employee creative? Are our engineers and designers creative? Does our company provide a creative environment or culture?[165] Design science often

---

[160] Some claim that this definition was formed by AMABILE. It is so common, it is nowadays usually used without quoting any original author.

[161] See e.g. Amabile (1983), Amabile (1996a), Dacey (1989); Guilford (1950); Guilford (1959); Guilford (1960); Lee (2005); Lubart (2001); Rhodes (1961); Runco et al. (2010); Sheldon (2006); Sternberg (1998); Sternberg (2006); Torrance (1995); and many more.

[162] For reading on education related literature about creativity, see e.g.: Treffinger et al. (2002); Torrance (1965); Torrance (1981); Buhler / Guirl (1963); as well as current research being conducted by the “Neag Center for Gifted Education and Talent Development”.

[163] See also Albers et al. (2006); Albers et al. (2008a)

[164] Schlicksupp (1977), 20ff.; Howard et al. (2008)

[165] See e.g. Amabile (1998), Isaksen (2007); Isaksen / Ekvall (2010); Rickards / Moger (2000)

studies creativity focusing on the originality of different solutions suggested by the designers focusing more on the outcomes than the human trait. They might ask questions such as: Is this version of a product more creative than another one? Is this a more creative solution to a design problem? Does this method help generate creative solutions?[166] DEIGENDESCH gives a comprehensive overview of those different domain-specific viewpoints on creativity as well as BÖRSTING.[167]

A widespread, more detailed explanation of the term creativity than the definition at the beginning of this chapter to describe the term creativity is RHODES' 4P concept, which dates back to the early nineteen sixties. He summarizes the description of creativity in the four classes *person*, *process, press* and *product* (short: 4P) after having reviewed many different treatises on creativity: "In time I had collected forty definitions of creativity […] But as I inspected my collection I observed that the definitions are not mutually exclusive. They overlap and intertwine. When analysed, as through a prism, the content of the definitions form four strands. Each strand has unique identity academically, but only in unity do the four strands operate functionally."[168]

Person: "The term person, as used [by Rhodes], covers information about personality, intellect, temperament, physique, traits, habits, attitudes, self-concept, value systems, defense mechanisms, and behavior."[169]

Process: "The term process applies to motivation, perception, learning, thinking, and communicating. Essential questions about process include: What causes some individuals to strive for original answers to questions while the majority is satisfied with conventional answers? What are the stages of the thinking process? Are the processes identical for problem solving and for creative thinking? If not, how do they differ? Can the creative thinking process be taught?"[170]

Press: "The term press refers to the relationship between human beings and their environment. Creative production is the outcome of certain kinds of forces playing upon certain kinds of individuals as they grow up and as they function. A person forms ideas in response to tissue needs, sensations, perceptions, and imagination. A person receives sensations and perceptions from both internal and external sources. A person possesses multi-factorial intellect, including ability to store memories, to recall and to synthesize ideas. Each idea that emerges reflects uniquely upon the originator's self, his sensory equipment, his mentality, his value systems, and his conditioning to the everyday experiences of life. Each person perceives his environment in a unique way; one man's meat is another man's poison and vice versa."[171]

---

[166] Howard et al. (2008)
[167] Deigendesch (2009), 48ff.; Boersting (2012), 56 ff.
[168] Rhodes (1961), 306 f.
[169] Ibid. 307
[170] Ibid. 308
[171] Ibid. 308

Product: "Product describes the view on creativity that focuses on the result of a creative act. The word idea refers to a thought which has been communicated to other people in the form of words, paint, clay, metal, stone, fabric, or other material. When we speak of an original idea, we imply a degree of newness in the concept. When an idea becomes embodied into tangible form it is called a product. Each product of a man's mind or hands presents a record of his thinking at some point in time. Thus an idea for a new machine reflects the inventor's specific thoughts at the moment when the concept was born. And by probing backward from the moment of inspiration it may be possible to trace the thoughts and the events leading up to the idea. Products are artifacts of thoughts. […] A system is needed for classifying products according to the scope of newness."[172]

Especially the "*process*" aspect of creativity is widespread and of particular interest within *design science* in two different ways: Scientists take a great interest in describing the creative design process[173]. They also develop methods to support creative activities along the creative design process.[174] Literature has a large number of different process-models of the creative process. HOWARD ET AL. have compared a large number of models and recombined it to a resulting meta-process of creativity consisting of the four major phases:[175]

Figure 13 - Meta Process of Creativity, Howard et al. (2008), 167

HOWARD ET AL ., however, argue that "the communication/implementation phase should be deemed a design activity."[176]

It becomes obvious that the term creativity is elusive and hard to grasp. TREFFINGER ET AL. have collected and sorted aspects of creativity definitions according to different authors (Table 20).

Table 20 - Different authors and their focus on the term creativity; (cp. Treffinger et al.

---

[172] Rhodes (1961), 309

[173] E.g. Howard et al. (2008); Albers et al. (2010); Albers / Braun (2011); Clarkson / Eckert (2004); O'Donovan et al. (2004)

[174] E.g. Albers et al. (2008b); Albers / Alink (2007); Eckert et al. (2009)

[175] See also Howard et al. (2008), 167

[176] Ibid., 167

(2002), p.9)

Source for Sample Definition	**Emphasis in Definition**	**Primary Focus Implications for Assessment**	**Identify creativity through**
Fromm (1959) Khatena / Torrance (1973) MacKinnon (1978)	Person	Characteristics of highly creative people	Assessment of creative personality traits
Gordon (1973) Guilford (1959), Mednick / Mednick (1965), Torrance (1964), Treffinger et al. (2001), Wallas (1926)	Cognitive process or operations	Skills involved in creative thinking or in solving complex problems	Testing for specific creative thinking and problem solving aptitudes or skills
Maslow (1976) Rogers (1954)	Lifestyle or personal development	Self-confidence, personal health and growth; self-actualization; creative context or setting	Assessing personal adjustment, health, and self-image; assessing the climate that nurtures or inhibits creativity
Gardner (1993), Khatena / Torrance (1973)	Product	Results, outcomes, or creative accomplishments	Assessing and evaluating products or demonstrated accomplishments
Amabile (1983), Rhodes (1961)	Interaction among person, process, situation, and outcomes	Multiple factors within specific contexts or tasks	Assessing multiple dimensions in a profile, with various tools

4.2.2.2 Creativity and Product Development Success

The assumption that creativity has a beneficial influence on the success of product development projects has been proven true in several studies in different industries. LOCH et al. showed a positive correlation between design quality and sales growth, while the design quality is strongly believed to be influenced by creativity of the design team.[177] STEVENS ET AL. found similar positive correlations between profits and creativity. They studied “New Product Development Projects” in a chemical company and found that project analysts with an above average degree of creativity found the more profitable opportunities when evaluating new business ideas within the company.[178] HEUNKS finds his hypothesis that “Innovation and success of a small firm depend on creativity, particularly” to be partially true, as the correlation could only be shown for older companies greater 32 years).[179] His findings are purely empirical, based on a survey among 200 entrepreneurs in six countries. They agree with what NYSTRÖM claimed some 20 years earlier: „Innovative companies should recruit and

---

[177] Loch et al. (1996), focus of the study was on the computer industry

[178] Stevens et al. (1999)

[179] Heunks (1998), 267

stimulate intuitive individuals and individuals who can switch between intuitive and analytic patterns of thought.“[180] Furthermore, studies focusing on product failure instead of business success have been conducted leading to the result that the primary reason for a new product’s failure is the lack of its uniqueness.[181]

4.2.2.3 What makes a Person creative?

A strong scientific interest, mainly driven by psychology, has been on possible traits of character or personality characteristics that correlate with creativity. As creativity itself is very hard to measure, the question for many psychologists is: Are there any typical other characteristics in a person that indicate his or her creativity?

A common perception used to be that there is a positive influence intelligence has on creativity. Newer studies have shown that this is only true within a limited scope. GETZELS AND JACKSON showed that the correlation is merely apparent up to an intelligence quotient of *120*.[182] Above this threshold, increased intelligence does not necessarily result in increased creativity and vice versa.[183]

This led to further diversification in the search of human traits indicating creativity and additional indicators were found. Table 21 gives an overview of some of these traits as summarized by SCHULER AND GÖRLICH.[184]

Table 21 – Traits of character that promote creativity, Schuler / Görlich (2007)

Trait	**Associated characteristics**
Intelligence	Handles complexity, intuition, insight, fantasy and imagination, education, integration ability
Intrinsic motivation	Ambition, stamina, concentration, achievement motivation, energy, achievement pleasure, drive, deferred gratification
Nonconformity	Originality, unconventionality, strive for autonomy, individuality, independence of judgment, independence of mind
Self-esteem	Emotional stability, self-perception: ‚creative‘, risk-taking
Frankness	Curiosity, enjoyment of new, aesthetic demands, intellectual values, need for complexity, wide interest, flexibility, ambiguity tolerance
Experience	Knowledge, mindset, metacognitive abilities (planning, monitoring, feedback, self-control, self-judgment)

Another trait closely related to creativity is 'problem-solving', emphasized by GUILFORD: “The very definitions of these two activities show logical connections. Creative thinking produces novel outcomes, and problem solving involves producing a new response to a new situation, which is a novel outcome."[185]

HOVECAR and BACHELOR relate the measurement of creativity to the detection of the abilities *divergent thinking* or *fluency*.[186] In extreme cases, divergent thinking can result

---

[180] Nystrom (1979), 57

[181] Crawford (1977), Cooper / Kleinschmidt (1987), Kleinschmidt / Cooper (1991), Cooper (1999)

[182] Getzels / Jackson (1962)

[183] Ibid.; Sternberg (1995)

[184] The table is translated. For the original, German version see Schuler / Görlich (2007)

[185] Guilford (1977) quoted by Treffinger et al. (2002), 6

[186] Hocevar / Bachelor (1989), from Redelinghuys / Bahill (2006), 122

in over-inclusion, which is the inability to accept and maintain conceptual boundaries. EYSENCK found that creativity is related to those psychotic forms of divergent thinking.[187]

4.2.2.4 What makes a Product creative

The findings about creativity, creative individuals, and processes presented thus far have repeatedly contained the construct of novelty or originality. It is an essential aspect when discussing what makes a product creative. Copying an existing solution is not considered a creative act, hence the copy of the original is usually not deemed to be creative. Therefore, for a product to be considered creative, it has to be new/novel/original.

BODEN also considers whether an idea is original to an *observer with limited information*. The idea might have occurred to someone else, somewhere else at a different time, but the observer does not know about it.[188] She calls this type *psychologically creative* (*P-creative*). If the idea is absolutely new and the world has not seen it before, it is the first time in history it has come up. Therefore, such an idea is titled *historically creative* (*H-creative*).[189]

4.2.2.5 Creativity Assessment

A slightly different viewpoint on creativity is taken in *creativity assessment*. The question at hand is not simply *whether* a person or product is creative or not but much rather *how* creative the product or the person is. What can we do in order to come up with reproducible, quantifiable values, which allow us to compare different individuals in regard to their degree of creativity?

While some authors take a management science perspective aiming at support for human-resource management, e.g. when setting up project teams, another relevant perspective is to assess the creativity of individuals or groups of individuals in order to determine whether two individuals or teams are comparably creative. Such a situation might appear when one group is given a problem and is asked to apply a design method and the other is supposed to solve the same problem without the method. Critics will always argue that in case of the first group being more successful, influences like incomparable degrees of creativity between the two groups might have biased the outcome. Hence, scientists need a means to compare and adjust the groups prior to any type of such tests. “A clear definition of, or a metric for assessing creative design outputs consisting of measurable elements, would enable researchers to gauge the effectiveness of any new creativity tools or methods proposed.” HOWARD et al. show that “the classification of ‘design outputs’ […] in the domain of engineering design –

---

[187] Eysenck (1994)

[188] The observer can also be the person having the idea

[189] Boden (1990), 43 ff.

often referred to as design types – closely relates to the research performed by psychologists involving ‘creative outputs’ […].”[190]

“No single assessment instrument or test provides evidence about all the possible meanings or elements associated with the construct of creativity.”[191] This is a result of the many definitions (cp. chapter 4.2.2) of the term creativity. “As much as we might yearn for precise, objective categories, the reality of the complexity of creativity, its attendant characteristics, and our assessment tools remind us that such precision is seldom attainable at the highest levels of human behavior.”[192]

As various as the definitions, as various are the different methods for assessing creativity. REDELINGHUYS and BAHILL reviewed them, developing their “framework for the assessment of the creativity of product design teams.“[193] Furthermore, TREFFINGER ET AL. present a three-page list of tools that can be used for creativity assessment.[194]

Criticism and Limitations of Creativity Assessment

Is Fluency the right measure? In several psychometric approaches to quantify creativity, it is actually fluency that is being measured.[195] It is assumed that high fluency in divergent thinking tests can be interpreted as creativity. However, a number of authors state a lack of validity in research showing that 'divergent thinking' equals 'creativity'.

Is creativity content specific? There is an ongoing debate within the creativity research community arguing whether or not creativity is content-specific. Some state that it is task related within the content domain. If we asses a group’s creativity – e.g. in the field of sporting goods – a snowboarding-related task might lead to completely different results than a tennis-related task, simply because one or more individuals might be quite interested in snowboarding, but dislike tennis. Among the critics who argue creativity to be content- and task-specific is e.g. BAER.[196] Theoretical and empirical evidence that support the notion that creativity is content specific have been presented by CSIKSZENTMIHALYI, GARDNER, RUNCO, and STERNBERG & LUBART.[197] However, other researchers have presented results, concluding that creativity is only partly domain or task-specific, and that a large portion of creativity is a general human ability.

Predictive validity: Furthermore, some authors argue that creativity assessment is lacking predictive validity.[198] Researchers who have explicitly addressed this problem state that the lack of predictive validity exists. However, it is not resulting from the psychometric approaches themselves, but from weaknesses in methodology, e.g.

---

[190] Howard et al. (2008), 170

[191] Treffinger et al. (2002), xiii; also Benedek in Dresler / Baudson (2008)

[192] Ibid. xiv, ff.

[193] Redelinghuys / Bahill (2006)

[194] Treffinger et al. (2002), 58 ff.

[195] Cp. Chapter 4.2.2.3

[196] Baer (1994a); Baer (1994b); Baer (1994c); Baer (1994d); Baer (1996); Kaufman / Baer (2004)

[197] Csikszentmihalyi (1988); Gardner (1993); Runco (1989); Sternberg / Lubart (1995)

[198] Predictive Validity is explained in chapter 4.3.3.3 as a variation of criterion validity

limited duration of the studies or statistical errors.[199] PLUCKER & RUNCO state that those who have explicitly addressed the weaknesses named above have collected positive evidence, pointing to a much better predictive validity.[200]

As a consequence, it seems impossible to name the ideal *creativity test* that could be used in a pre-test to group participants in comparably creative teams. Much more important, it does not seem advisable to simply use one single test, after studying the large number of assessment techniques available. No matter how one assesses the participants' creativity, if another technique was used, the result is likely to be quite different, so parallel-test reliability is not given, hence biasing the experiment more than randomizing the teams would. If assessment techniques are applied, it is possible to use various techniques, or to switch to team-assessment techniques that do not so much focus on creativity alone like e.g. Myers-Briggs type indicator, which is a personality assessment technique. A similar argumentation is also used by BAER, who advises to use "Consensual Creativity Assessment" rather than any of the metrics available in college and university settings.

"Assessment of creativity presents a unique challenge in higher education. Although there are tools on the market for assessing creativity, most are designed for young children, and all tend either to lack sufficient validity and reliability or to assess only rather trivial aspects of creativity (or, in many cases, both). If creativity is to be assessed in college settings in a meaningful way, divergent-thinking tests like the Torrance Tests of Creative Thinking and other commonly used creativity tests are inadequate because they fail to meet even the loosest standards of validity."[201] Alternatively, related personality assessment techniques can be used to group test-teams, some of which are also presented in the following paragraphs.

Consensual Creativity Assessment Technique

The Consensual Creativity Assessment Technique is a straightforward and simple alternative to the lengthy list of metrics available for creativity assessment resulting from the equally long list of assumptions and models about creativity. It does not rely on any of these models. One could go so far and say it ignores them. For a full description, refer to BAER AND MCKOOL.[202]

Consensual Creativity Assessment Technique is basically the same as what a jury in a contest does to decide, which contribution wins the price:Subjects are asked to create something, and experts from the domain are then asked to evaluate the creativity of the things they have made.

---

[199] Hocevar / Bachelor (1989); Plucker / Renzulli (1999); Torrance (1979), all also quoted and discussed in: Plucker / Runco (1998), 3ff.

[200] Hong et al. (1995); Milgram / Hong (1993); Okuda et al. (1991); Plucker (1999); Sawyers / Canestaro (1989); all also quoted and discussed in: Plucker / Runco (1998), 3ff.

[201] Baer / McKool (2009)

[202] Baer / McKool (2009)

Some basic rules have to be followed, however. It is important to equip all participants with the same task and materials since their creations are being compared relative to one another. Then, the experts have to judge the creations. It is important that they do their judgement independently to avoid group dynamics among the jurors. In cases, where the judges literally sit together, it is practical to instruct them to write their decisions down before it is being discussed. Furthermore, it is important to expose the judges to all the creative products first and only then let them decide. Results can be improved if they are given a scale. The goal is to compare the results within the group, so even if all results impress a judge very much, the most creative should get the highest score and the least creative the lowest, even if it were likely to rank high within a different group. The judges are *not* asked to explain or defend their ratings in any way, and it is important that no such instructions be given.

BAER also discusses reliability and validity of the consensual creativity assessment technique. It has been shown by different researchers in independent studies that expert ratings are similar, even if they do not always agree on all levels. This is why a *group* of experts is needed. It could also be shown that the larger the group of experts, the better their results concerning the overall inter-rater reliability correlations. As a rule of thumb, 10 judges seem to be a good number.[203]

Summing up, Consensual Assessment Technique seems a practical method to judge relative creativity within a group of subjects. In an experimental setting, it seems easier than any of the complicated metrics and it can be conducted without a psychologist present. Yet, it seems sufficiently accurate and reliable. On the negative side, the researcher has to put together a group of about 10 experts. However, in design research, these experts can just as well be fellow researchers, grading homework, or a pre-test assignment of students. For further reading on the consensual assessment technique refer to the works of AMABILE, BAER, and HENNESSEY.[204]

---

[203] Compare e.g. Amabile (1996a), Baer (1994d), Baer / McKool (2009)

[204] Baer / McKool (2009), Amabile (1996a), Amabile (1996b), Baer et al. (2004), Hennessey et al. (1999),Hennessey (1994)

Table 22 - Profile of "Consensual Assessment Technique"

Consensual Assessment Technique	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies in real-world deployment of design support ▪ Evaluating the design results produced with/without a sophisticated method.
answers questions of the type:	*How good is a solution* compared to the other solutions produced in the experiment?
Advantages	No metrics are needed, easy to apply.
Disadvantages	▪ Arguably subjective. ▪ Effort, as several experts are needed.
Necessary preparation	▪ Engage a group of experts in the field ▪ Prepare/duplicate solutions so that experts can evaluate independently from one another
Follow-up work	none
Related techniques	▪ Metric for assessment of design outcomes
Synonyms	▪ Jury rating; expert judgment; expert rating

Meyers-Briggs Type Indicator

A popular and well known instrument for personality assessment is the Myers-Briggs Type Indicator (MBTI). It is commonly applied in commercial settings. *CPD Inc.* is the company that holds the MBTI trademark today. They offer various types of data sheets and MBTI-related applications as well as consulting services for leadership development. *CPD Inc*. claims to have several million customers every year.[205] ALBERS uses a personality test which is based on the MBTI to set up student design teams in larger design projects.[206] STEVENS et al. used the MBTI in their study on New Product Development projects in the chemical industry (see also p. 64).

When KATHARINE COOK BRIGGS and her daughter, ISABEL BRIGGS MYERS, developed the MBTI, the motivation was to match women in the American workforce to support the war industry to jobs according to their personality. It was assumed that the women would be more efficient and effective. The MBTI is based on psychological findings from JUNG in the early1920s. He believed that humans are born with natural differences in their behavioral preferences similar to the way people are either born right- or left-handed. The key elements of his work that were adopted for the MBTI were the 'psychological functions'. JUNG proposed the existence of two dichotomous pairs of cognitive functions:[207]

---

[205] For more information see: https://www.cpp.com/products/mbti/index.aspx, 12/08/12, 15:42

[206] Albers et al. (2006)

[207] Jung (1921)

- The "rational" (judging) functions - *thinking* and *feeling*
- The "irrational" (perceiving) functions - *sensing* and *intuition*

Briggs-Myers and Briggs added 'attitude' and 'lifestyle'. MBTI summarizes the differences in personal preference in four opposing pairs called dichotomies (see Table 23). An individual's preference – his or her psychological type – can now be described in either side of the four dichotomies. One prefers either extraversion or introversion, prefers sensing over intuition or vice versa, and so on. This results in a total of sixteen combinations. Since an individual's personality cannot merely be specified through one out of sixteen four-letter combinations, more comprehensive descriptions and interpretations of the sixteen types are available in the corresponding literature as well as online.[208]

Table 23 - The four dichotomies for MBTI (left) and typical profile of creative individuals in MBTI (right)

	Dichotomies		
Attitude	Extraversion	⟷	Introversion
Perceiving function	Sensing	⟷	Intuition
Judging function	Thinking	⟷	Feeling
Lifestyle	Judging	⟷	Perception

Creative people in MBTI		
[✓ Extraversion]	⟷	Introversion
Sensing	⟷	[✓ Intuition]
[✓ Thinking]	⟷	Feeling
Judging	⟷	[✓ Perception]

It is important to note that MBTI indicates what a person *prefers*. It does not predict how that person will *act*. The two do not necessarily mean the same. This is considered to be one of the main reasons why sometimes people cannot identify themselves with the result of the test hence questioning its validity.[209] The MBTI is generally subject to a lot of criticism. Studies have shown that e.g. its construct validity, internal consistency, and test-retest reliability are quite convincing. [210] Other studies support the assumption that MBTI lacks credibility.[211]

Gough developed the MBTI further suggesting the "Myers-Briggs Type Indicator Creativity Index." He conducted the MBTI assessment with individuals from different domains that were deemed to be highly creative through suggestion by peers. Table 23 (right) also shows the tendencies of creative individuals taking the MBTI assessment that Gough found empirically.

Table 24 - Profile of "Myers-Briggs Type Indicator"

---

[208] Myers (1962); Myers-Briggs (1962); Carlyn (1977); Gough / Library (1981); Myers et al. (1985); McCrae / Costa Jr. (1989); Furnham (1996); A comprehensive description with redirects for further studying can be found on the English version of Wikipedia: http://en.wikipedia.org/wiki/Myers-Briggs_Type_Indicator, 12/08/12, 17:29 and https://www.cpp.com/, 12/08/2012, 17:30

[209] Carskadon / Cook (1982)

[210] Thompson / Borrello (1986a); Capraro / Capraro (2002)

[211] Stricker / Ross (1964); Carlyn (1977);Thompson / Borrello (1986b); McCrae / Costa Jr. (1989); Danmin et al. (2000); Hunsley et al. (2003);

Myers-Briggs Type Indicator	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies in real-world deployment of design support
answers questions of the type:	How can comparably creative design teams be put together for evaluation studies.
Advantages	Well known method with worksheets and support available.
Disadvantages	▪ Arguable predictive validity ▪ Being commercialized, application of MBTI costs money
Necessary preparation	▪ Acquisition of worksheets or MBTI survey software
Follow-up work	Mixing the teams according to the results
Related techniques	▪ MBTI Creativity index ▪ KAI
Synonyms	▪ Personality test

Assessing Problem Solving Style

In engineering design, the individual problem solving style is an issue discussed in many contexts. Neglecting peculiar problem solving style has been part of the criticism against design methods (compare chapter 2.1.4.1). The British psychologist MICHAEL KIRTON developed a theory based on the assumption that everybody is creative but that there are different ways of developing "new things".[212] In his Kirton-Adaption-Innovation inventory (KAI), he places individuals between two extremes: innovators and adaptors. "Adaptors desire to do things better; innovators seek to do things differently."[213] Table 25 gives a more comprehensive overview of typical characteristics of the two extremes as well as some indications about what one extreme might find difficult to deal with about the other extreme.

For the assessment of a test person's KAI, a questionnaire is filled out. According to the answers, a score between 32 and 160 is assigned to the test person.[214] "A person with an adaptive style will usually score in the 60–90 ranges, whereas a person with an innovative style will score between 110 and 140. Persons with scores in the middle of a group have some of both characteristics, and under some circumstances, they can function as 'bridgers'. This inventory has been found to be extremely accurate and has been globally validated across many cultures over decades."[215]

---

[212] Kirton (1976)

[213] Stum (2009)

[214] The questionnaire along with instructions on it evaluation was published in Kirton (1976), a new and refined version can be found in Kirton (2003). The questionnaire may not be reprinted within this thesis.

[215] Online source: http://pubs.acs.org/subscribe/archive/ci/31/i11/html/11hipple_box3.ci.html, 17/11/2012, 16:35

Table 25 - Characteristics of Adaptors and Innovators[216]

Adaptor	**Innovator**
Efficient, thorough, adaptable, methodical, organized, precise, reliable, dependable	Ingenious, original, independent, unconventional
Accepts problem definition	Challenges problem definition
Does things better	Does things differently
Concerned with resolving problems rather than finding them	Discovers problems and avenues for their solutions
Seeks solutions to problems in tried and understood ways	Manipulates problems by questioning existing assumptions
Reduces problems by improvement and greater efficiency, while aiming at continuity and stability	Is catalyst to unsettled groups, irreverent of their consensual views
Seems impervious to boredom; able to maintain high accuracy in long spells of detailed work	Capable of routine work (system maintenance) for only short bursts; quick to delegate routine tasks
Is an authority within established structures	Tends to take control in unstructured situations
How the “other side” often sees extreme adaptors and innovators	
Dogmatic, compliant, stuck in a rut, timid, conforming, and inflexible	Unsound, impractical, abrasive, undisciplined, insensitive, and one who loves to create confusion

Originating from psychology, KAI has been popular with management science, trying to answer the question which constellations make successful teams in industrial practice for different tasks or which types of creativity can be expected from certain teams, but also geographical and gender differences have been investigated. The Following table gives an overview of studies related to KAI.[217]

Table 26 - Profile of "Kirton-Adaption-Innovation inventory"

Kirton-Adaption-Innovation inventory	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies in real-world deployment of design support
answers questions of the type:	▪ How can comparably creative design teams be put together for evaluation studies.
Advantages	▪ Well known method with worksheets and support available.
Disadvantages	▪ Being commercialized, application of KAI costs money
Necessary preparation	▪ Acquisition of worksheets or KAI survey software
Follow-up work	▪ Mixing the teams according to the results
Related techniques	▪ MBTI
Synonyms	▪ Personality test

---

[216] Online source: http://pubs.acs.org/subscribe/archive/ci/31/i11/html/11hipple_box3.ci.html, 17/11/2012, 16:35

[217] Stum (2009), 70f.

Table 27 - Studies related to Kirton's Innovator-Adaptor theory

Source	**Subject of Study**
Buffinton et al. (2002)	Entrepreneur's problem-solving styles: empirical study using KAI
Buttner / Gryskiewicz (1993)	Entrepreneur's problem-solving styles: empirical study using KAI
Chan (2000)	KAI inventory using multiple-group mean and covariance structure analysis
Foxall / Hackett (1994)	Styles of managerial creativity: KAI comparison of United Kingdom, Australia, and United States
Goldsmith (1984)	Personality characteristics and KAI
Hutchinson / Skinner (2007)	Self-awareness and cognitive style: KAI, self- monitoring, and self-consciousness
Jabri (1991)	Educational and psychological measurement: modes of problem solving
Kaufmann (2004)	Two kinds of creativity
Kubes (1998)	KAI in Slovakia: cognitive styles and social culture
Kwang et al. (2005)	Values of adaptors and innovators
Meneely / Portillo (2005)	Personality, cognitive style, and creative performance
Mudd (1996)	KAI inventory: evidence for style/level factor composition issues
Schilling (2005)	Network model of cognitive insight
Shiomi / Loo (1999)	Cross-cultural response styles and KAI
Skinner / Drake (2003)	Behavioral implications of KAI
Taylor (1989)	KAI: re-examination of inventory factor structure
Tullett (2011)	KAI cognitive styles of male and female project managers
Woodman et al. (1993)	A theory of organizational creativity

4.2.3 Research Methods for Content oriented Design Research

Designing can be regarded from mainly two different viewpoints: The process and activities, on the one hand, and the outcomes of those activities on the other.

The content of those outcomes includes concepts, sketches, and prototypes of verbal or sentential descriptions of products. When the effectiveness of a design support is to be measured, literature suggests several possibilities, the most common being: Quantity, variety, novelty, quality, and feasibility. Experts agree that using only one of the factors is misleading since in industrial practice, all those factors are relevant. A sizeable quantity of ideas is not impressive if they are basically just variations of one idea. Even if a large variety is presented, it is important that the ideas are novel. At the same time, ideas that are not feasible eventually get discarded. Therefore, a combined application for the evaluation of design outcomes is suggested in the following sections. Alternative information on design outcome evaluation is documented by Duffy, who published the results of the First International Engineering Design

Debate.[218] In the debate, it was discussed what exactly design productivity is, how it can be measured what are the elements that make design effective, how effective are they and how do they relate.

4.2.3.1 Metric for Quantifying Design Outcomes

To evaluate newly developed design support, design tasks and example projects are set up. The most common success criterion of a design project would be the quality of its result. Mind that the quality is not exclusively linked to the application of a method.[219] However, one important step is to define criteria for the evaluation of the design outcome. Common criteria are *originality*, *complexity* and *creativity*. There have been numerous attempts to define sets of criteria.[220] One should be careful to apply any of these without adaption, since the success criterion that defines a positive outcome after application of a method is very much dependent on the goal of the method. Therefore, it is advisable to take great care when defining one's original system of objectives.

Quantity

The quantity of ideas generated in a given time is believed to be a good indicator for the effectiveness of a method. Studies have shown that design processes, which generate many ideas during the process achieve better solutions as the final outcome of the process.[221] Hence, methods that aim to increase the number of ideas generated are generally believed to have a positive influence on the design outcome.

The difficulty with counting the quantity is that the researcher has to decide at which point two ideas are different enough to be considered as two separate ideas instead of just counting them as one. For this reason, SHAH and VARGAS-HERNANDEZ suggest a combined metric.[222] They avoid the problem by introducing as a further measure *variety*, in which this factor is accounted for. This way, the researcher is allowed to count the number of documented ideas and thus appoint a quantity-score M_{quan} accordingly. In order to be compatible to the other three metrics for later combination, the Quantity score should be normalized on a scale from 1 to 10.

$$M_{quan} = \frac{N_i}{n} * 10 \qquad (1)$$

N_i: number of ideas produced by team i

n: total number of ideas in the set

Variety

Variety addresses the number of *different* solutions. It indicates how well someone has explored the design space. One has to be careful though on how to decide whether

---

[218] Duffy (1997)

[219] Cp. chapter 2.1

[220] See e.g. Blessing / Chakrabarti (2009)

[221] Parnes (1961)Osborn (1963), Basadur / Thompson (1986), Kumar et al. (1991), Candy (1996), Cross (1996)

[222] Shah / Vargas-Hernandez (2003)

two solutions are to be considered different or not. It is highly recommended to use some type of predefined measure in order to avoid bias.

Figure 14 - How many different solutions of devices that show the time do you count? (own illustration)

SHAH and VARGAS-HERNANDEZ suggest a suitable metric: a "genealogy tree" *for each function* (Figure 15). [223] The first step is to determine all the functions that have been embodied. Distinguishing functions on such an abstract level is not trivial. There is no common concept of what exactly a function is. While this might be a secondary issue as long as one sticks to the same mental model of functions when comparing a set of ideas, it might be challenging to agree on a mental model at first place. Even if only one researcher is involved, (s)he should spend some time on this issue. For further reading on functions and their mental models refer to the works of ALBERS, ALINK ECKERT, and GERO among others.[224] One option is to use the contact and channel and connector approach by ALBERS to determine the different functions.

After the functions are set, they can be assigned a weight f_j to express differences in their importance. The nodes in the tree carry the count of ideas in each category in each level. The levels are also assigned weights s_k.[225] Variations on a detailed level (e.g. when two ideas use the same geometry just with two different materials) now attain a smaller score than variations between different physical principals (e.g. when the same function is realized with hydraulics in one idea and electro-mechanically in another). The example shows four levels to distinguish ideas. It is possible that fewer or more levels are needed to distinguish within a given set of ideas.[226] The variety score M_{var} assigned to the analysed set of ideas calculates to:

---

[223] Shah / Vargas-Hernandez (2003), 126 ff.

[224] Albers et al. (2004), Keller et al. (2007), Eckert et al. (2011), Boersting et al. (2008), Alink et al. (2011), Alink (2010), Gero / Kannengiesser (2002)

[225] In the example in Figure 15 these are: s_1=10, s_2=6, s_3=3, s_4=1

[226] The number of levels should be as small as possible, as large as necessary. The effort for the analysis rises dramatically with the number of levels!

$$M_{var} = \sum_{j=1}^{m} f_j \sum_{k}^{4} s_k\, b_k/n \qquad (2)$$

b_k: number of branches at level k
s_k: weights for each level on a scale from 1 to 10 [227]
m: total number of functions counted in the set
n: total number of ideas in the set

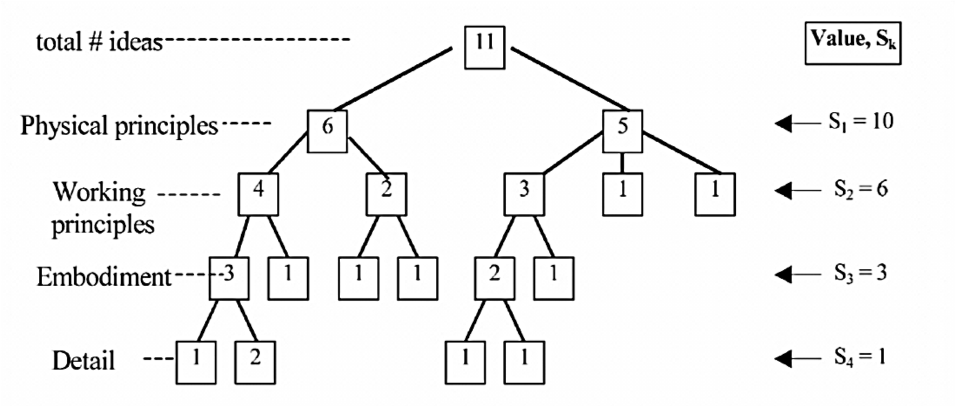


Figure 15 - Genealogy Tree to measure variety within a set of ideas, Shah / Vargas-Hernandez (2003), 126

Novelty

Since the goal of designing is to generate something new, the degree of novelty of an idea is a useful measure to determine the effectiveness of an applied method. If it produces more novel ideas than other methods, it is a preferable method.

The problem is that novelty is something relative to context. An idea can be new to an individual, a design team, a company and so on but known to society, industry, science and so on. Before assessing novelty, one should therefore, consciously determine how novelty should be defined in the concerned context. When e.g. a design support is tested in a creativity lab, it is irrelevant whether or not ideas that are generated during an experiment are known somewhere else in the world. If the ideas are new to the participants, they can be deemed novel. On the other hand, in an industrial context, it is important that innovation processes lead to ideas that are novel to the industry or the market.

Determining novelty can therefore be done either in an absolute perspective or on a relative perspective. The first method would call for a way to collect all solutions for the problem available (in the whole world!). This will lead to additional questions such as: “How does one make sure they are finished collecting?” Just because one does not find a solution in his or her research, it does not mean that the solution has not been come up with before.[228] Such a measure appears impractical and further discussion shall be left to those dealing with epistemology or other traits of philosophy.

---

[227] In the example (Figure 15): s_1=10, s_2=6, s_3=3, s_4=1

[228] Compare Historical creativity in chapter 4.2.2.4

Alternatively, personal novelty could be taken as a measure. An idea is novel to an individual if (s)he has never had that idea before, no matter if others have come up with the same idea. This type of novelty is subject to a lot of potential bias. One has to rely on self-estimation of participating individuals if such measure was used in experiments. It is also difficult to assess ideas generated in a group if some of the individuals state they were familiar with the idea beforehand, and others state they were not. Hence, idea evaluation on a *relative scale within the group* is suggested. The ideas produced during a design session are compared among each other and ranked from common/obvious to surprising/novel. An idea everyone in the group had (high count) is obvious and gets a low score. An idea counted only once or very few times is less obvious and attains a higher score. Since the degree of novelty is determined in relation to the group of individuals that take part in the experiment, it is referred to as societal novelty.

SHAH AND VARGAS-HERNANDEZ propose the following procedure to achieve such a ranking.[229] It is derived from psychology and creativity analysis and has been applied in those contexts successfully.[230] They suggest decomposing the given problem into its key functions or characteristic. In a second step, each idea is analysed determining:

- Which of the functions does it satisfy?
- How does it fulfill the function at a conceptual level?
- How does it fulfill the function at the embodiment level?

$$M_{\mathrm{nov}} = \sum_{j=1}^{m} f_j \sum_{k=1}^{n} S_{1jk} p_k \tag{3}$$

M_{nov}: overall Novelty score for the idea
m: number of key functions of the problem
f_j: weight of function j
n: number of design stages[231]
S_1: novelty score
p_k: weight for the stage n

S_1 can be gained in two different ways. If it is expected that all solutions can be closely predicted beforehand, it is possible to define the preliminary total set of ideas and assign a novelty value to each idea, e.g. through expert discussion. S_1 may then be determined for each idea by finding a closest match. This approach is questionable and should be used with great care as it is extremely biased. A more objective yet more complex way to determine S_1 is given with *(4).*

$$S_{1jk} = \frac{T_{jk} - C_{jk}}{T_{jk}} \times 10 \tag{4}$$

T_{jk}: total number of ideas produced for function *j* and stage *k*
C_{jk}: count of the solution within the set of produced ideas for function *j* and stage *k*
Multiplying by 10 normalizes on a scale from 0 to 10.

---

[229] Shah / Vargas-Hernandez (2003), 117 ff.

[230] See e.g. Torrance (1962), Torrance (1964), Jansson / Smith (1991)

[231] Often, only conceptual level and embodiment are distinguished leading to n=2

Quality

Quality is an attribute that is usually hard to estimate in early, conceptual stages. At the same time, a large number of developed design methods aim at improving those early steps. This causes a problem and justifies using a methodological approach to cope with the task.

A large number of methods exist to quantify the quality of a product or its idea. Lists and comparisons of such methods can be found in product engineering literature dealing with decision making as well as the German VDI guidelines dedicated to evaluation and selection of ideas and solutions.[232] However, experience shows that the more abstract an idea or a set of ideas is, the harder it is to exactly quantify certain criteria (e.g. price or weight). Therefore, when choosing a method, researchers should be careful and not get lured into choosing the "most precise" method.

$$M_{\text{qual}} = \sum_{j=1}^{m} f_j \sum_{k=1}^{n} S_{jk} p_k \Big/ n \sum_{j=1}^{m} f_j \qquad (5)$$

M_{qual}: overall quality score for the idea
m: number of key functions of the problem
f_j: weight of function j
n: number of design stages (conceptual, embodiment)
S_{jk}: quality score for function j at stage k
p_k: weight for the stage n

Combined Metric

The original authors have not defined how to combine the four scores and neither will this be done here, but much rather the options and strategies a researcher has will be discussed. Math offers us different possibilities on how to "add" the score. We can simply add the four scalars arithmetically (Option I). If we do so, we assume all four measures to be equally important, which means that we have to make sure for the single scores that they are on comparable scales, otherwise one measure will be overly important compared to the other three. The original authors do not justify or explain this, however, they did mathematically take care of this problem ensuring that M_{qual}, M_{quan}, M_{nov} and M_{var} all range between zero and ten. This is achieved through the following boundary conditions:

$$0 < S_i < 10 \quad\Big|\quad \sum f_j \overset{!}{=} 1 \quad\Big|\quad \sum p_k \overset{!}{=} 1 \qquad (6)$$

We might also intend to emphasize one of the measures. This could be e.g. because we are testing a design support that has been especially intended to produce novel ideas (such as synectics). In this case, we can assign weights to the measures (Option II). If our interest is, for some reason, limited to only one of the factors, we can ignore the other factors and just compare different sets of ideas according to the one factor of interest (Option IV). If we are looking for a score that is more balanced we can add the four measures geometrically. Such might be practical, if we don't want to

---

[232] Ehrlenspiel et al. (2005), Ehrlenspiel (2007), Pahl et al. (2005), Hubka / Eder (1982),VDI (1998)

overemphasize sets of ideas that are really good at one of the four factors, but are weak in the others (Option III). E.g. if we do not want to support methods that produce tons of ideas none of them being novel and all of them being quite similar. Which of the options described here is the most suitable depends on the context of the experiment. The researcher should make his or her choice, and explain why, as part of the documentation.

Option I: $$M = M_{quan} + M_{var} + M_{nov} + M_{qual} \tag{7}$$

Option II: $$M = a_{quan}M_{quan} + a_{var}M_{var} + a_{nov}M_{nov} + a_{qual}M_{qual} \tag{8}$$

Option III: $$M = M_{quan}\ OR\ M_{var}\ OR\ M_{nov}\ OR\ M_{qual} \tag{9}$$

Option IV: $$M = \sqrt{M_{quan}^2 + M_{var}^2 + M_{nov}^2 + M_{qual}^2} \tag{10}$$

Table 28 - Profile of "Metric for assessment of design outcomes"

Metric for assessment of design outcomes	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment ▪ Evaluation of design results produced with/without a certain design support.
Answers questions of the type:	▪ *How good* is a design solution or a set of design solutions produced by team x?
Advantages	▪ Quantified measure for comparing design solutions.
Disadvantages	▪ Only advisable to be used for comparison of design solutions within one set of solutions.
Necessary preparation	▪ Conduct design experiment and collect results.
Follow-up work	▪ none
Related techniques	▪ Consensual creativity assessment
Synonyms	none

4.2.3.2 Evaluating design sketches

Typical objects generated during the design process are sketches. Their evaluation can deliver insights on design strategies, activities, level of maturity of a design and much more. Since the quality and appearance of sketches vary between designers, McGown et al. and Rodgers et al. label categories based on the visible elements in engineering sketches (see Figure 16).[233] Such a classification is necessary if sketching activity is to be compared. McGown et al. *quantify* the *information* contained *in*

---

[233] McGown et al. (1998), 446

sketches to be able to compare the efforts and achievements of different designers.[234] They take into account both complexity c and size s of a sketch (Figure 16) arguing that larger sketches contain more detail and take more effort.[235]

$$I_{ps=c*s} \qquad (11)$$

I_{ps}: Information per sketch
c: complexity factor | $1 \leq c \leq 5$
s: size factor | $1 \leq s \leq 5$

In the original study, the designers were provided with A3-size sketchbooks and instructed to use them. Based on this, the size factor s was assigned as shown in Table 29. If different size sketchbooks are used, the values should be adapted. It must also be note that the metric's origin does not lie in engineering. When applied in engineering design, the complexity levels need to be adapted. In the original example as shown in Figure 16, the scale only focuses on 3D sketches. In engineering, technical drawings or principal sketches in 2D may contain a lot more detail and information. One possibility is to group all results from a sketching observation exercise from 'very simple' sketch (this will equal MCGOWN'S complexity level 1) to 'simple', 'average', 'complex', and 'very complex' (equal to complexity level 5).

Another option is to instruct the participants to use a certain size sheet of paper and be very specific about the type of sketches that are expected from the participants. If all sketches are 2D technical drawings, e.g., they can be compared directly and value needs to be assigned and distinguished.

---

[234] See McGown et al. (1998) for the full details of the study

[235] Author's remark: In times of tablet computers, where we zoom in and out of sketches and documents with a movement of two fingers, it is questionable whether size is the correct measure. As the authors explain themselves, it is the degree of detail that they actually take into account.

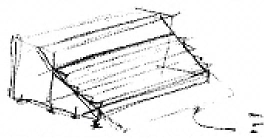
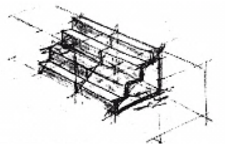
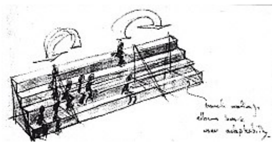
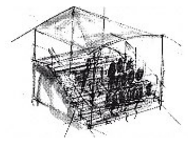
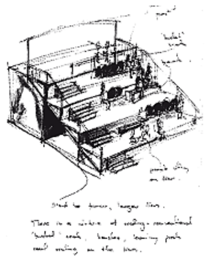
Complexity Level 1 Monochrome line drawing. No shading to suggest 3-D form. No text or numerical annotations are used.	
Complexity Level 2 Monochrome line drawing. There is no shading to suggest 3-D form. But there is use of different thicknesses of line. One or two brief annotations may appear, but no more than 6 or 7 words.	
Complexity Level 3 Monochrome, with rough shading used to give suggestion of 3-D form. The drawing may be annotated to describe certain aspects of the concept. May include dimensions.	
Complexity Level 4 Subtle shading is heavily suggestive of 3-D form. The drawing will almost certainly be annotated. Colour may be used to illustrate certain parts of the concept or arrangement.	
Complexity Level 5 Extensive use of shading to suggest 3-D form. Annotations will be used to ask questions of the idea or explain it. Colour will be heavily used. Generally a very busy drawing – many lines will be used in its construction. The drawing will almost certainly be annotated. Colour may be used to illustrate certain parts of the concept or arrangement.	

Figure 16 - Five levels of complexity in design sketches (Rodgers et al. (2000), 458)

Table 29 - Size factor for Information evaluation of sketches (McGown et al. (1998), 447)

Size of sketch	Assigned value
thumbnail sketch up to 50 mm × 50 mm	1
up to 100 mm × 100 mm	2
up to 150 mm × 150 mm	3
very large—up to an A4 page	4
full page—drawing covers most of the A3 page	5

The metric by McGown et al. produces a way to quantitatively compare sketching activity. While the metric is rather simple and straightforward, its advantage is that with such clear instructions, data reliability is increased.

Table 30 - Profile of "Observation of sketching activities"

Observation of sketching activities	
Field of application within design research	▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies, real-world deployment of design support ▪ Content oriented design research ▪ Design research concerning relation between content and process
answers questions of the type:	▪ Which types of sketching help which types of problem solving? ▪ Does the number / frequency of sketches correlate with design results?
Advantages	▪ Direct data collection, uninfluenced by the observer.
Disadvantages	▪ Limited types of research questions can be addressed with this method ▪ Relation between individual problem solving style and sketching behavior is unknown, therefore, generalization of study results are questionable.
Necessary preparation	▪ Instruct designers ▪ Supply designers with equal size sketchbooks
Follow-up work	▪ Coding of collected sketches ▪ Analysis of collected data
Related techniques	▪ Content analysis
Synonyms	none

4.2.4 Research Methods for Process oriented Design Research

Different opinions exist on the question whether or not design outcomes are suitable for evaluating the quality of a design method. Some argue that it is also important to regard the design process while applying the method. Most likely, different types of methods will call for distinctive ways of evaluating them. The following section lists different approaches to observing design processes.

COLEY ET AL. give an overview on the state of the art of capturing the cognitive behavior of designers. They argue that there has been "a rapid growth in the number of studies into the behavior of designers in recent years, and therefore, it is necessary to provide a critical analysis of this work to identify the most popular techniques currently being utilized to capture cognitive behavior."[236] A similar summary of techniques has been given by STAUFFER ET AL.[237] BENDER has suggested the application of methods from social sciences and also presents an overview.[238] The following section draws together the research techniques described in the treatises of COLEY ET AL., BENDER ET AL. AND STAUFFER ET. AL. Further techniques and explanations are amended based on the study

---

[236] Coley et al. (2007), 311
[237] Stauffer et al. (1991)
[238] Bender et al. (2002)

of the mainly of the “Journal of Engineering Design” and the interdisciplinary Journal “Design Studies."

4.2.4.1 Observation Techniques

A researcher is present in person or by camera. He or she takes notes on what the observed individual(s) is/are doing. As simple as it sounds, researchers should be aware of the different variations of observation techniques in order to document his or her observations for other researchers, properly. STAUFFER ET AL. divide observation into three categories: *'Structured* observations' *observations',* and '*participant observation'.* The latter will be explained in chapter 4.2 under its synonymous title

Ethnography / Ethnographic studies.

Unstructured observation

Unstructured observation is used for explorative types of research. The researcher is not looking for anything specific but wants to find out about the domain. The goal is to identify unusual behavior or actions, curiosities that are worth further exploration. Anything that strikes the researcher as odd is potentially interesting for additional investigation. Therefore, if unstructured observation is applied, it is usually in the very early stages of a research project, or even in preparation of a research project. The outcome of unstructured observations is hard to predict if not unpredictable by nature. It is not based on research question or hypothesis but intents to produce them. A particular problem of such unstructured observation is that due to its exploratory character and unpredictability of results, it has an air of being unscientific. While it is mandatory for developing research questions, funding for such observations is extremely rare. Another problem with unstructured observation is prejudice. Observers must be careful not to be prejudiced with expectations or assumptions. If one expects to observe certain behavior, it is likely that they find it whether it actually exists or not. If e.g. a researcher believes that within a design team, there are tensions between marketing experts and design engineers, any activity supporting the assumption will stick out as 'especially extraordinary'. Therefore, it is advisable to execute such observations in teams of more than one researcher and discuss personal opinions, assumptions and expectations beforehand. Thorough documentation allows the observers and others to revise the data by colleagues or at later points in the research.

Table 31 - Profile of "Observation"

	Unstructured Observation	**Structured Observation**
Field of application within design research	▪ Project Planning and controlling ▪ Empirical research, analysis of real-world design processes ▪ Explorative research, discovering the domain and looking for research questions	▪ Empirical research, analysis of real-world design processes ▪ Looking for specific behavior of designers.
answers questions of the type:	What is extraordinary about the domain?	*Are the designers actually doing activity x?* How often / how long do they conduct activity x?
Advantages	Anything goes	Produces quantifiable data
Disadvantages	▪ Presence of observer (could be a camera) may alter designers behavior ▪ Subjective bias introduced by the observer ▪ Difficult to report. ▪ Lengthy documentation ▪ Hard to plan effort beforehand ▪ Unpredictability of results	▪ Presence of observer (could be a camera) may alter designers behavior ▪ Subjective bias introduced by the observer ▪ Danger of false interpretation of the observed behaviour
Necessary preparation	▪ Set up camera / arrange observatory ▪ Select case /design a task	▪ Set up camera / arrange observatory ▪ Select case /design a task ▪ Selection / maybe testing of a coding scheme
Follow-up work	▪ Structuring of data	▪ Data analysis
Related techniques	▪ Ethnography	▪ Protocol analysis ▪ Coding Schemes
Synonyms	▪ Data collection ▪ Behavioral studies	▪ Data collection ▪ Behavioral studies

Structured observation

If research questions or hypotheses have already been formulated, structured observation is more suitable. The researcher is looking for a specific behavior, actions or phenomena. (S)he is recording whether or not it occurs, how often it occurs, when, by whom, under which conditions and so on. For thoroughly scientific observation, coding schemes can help to reduce bias introduced by the observer(s). If more than one observer take part in the study, coding schemes are highly recommended. Their preparation and application are explained in 0. Structured observation records strictly external behavior, i.e. what the designers actually do. If it is also intended why they do specific things and what they think, other methods must be applied or combined with the observation (typically questioning techniques). E.g., after the researcher observes certain behavior of an individual, an interview or a questionnaire could reveal *why* the designer acted that way.

4.2.4.2 Think aloud Method

The think aloud method is a concurrent data collection method, i.e. the data is collected, while it is being generated, in contrast to retrospective methods where the data is collected after the actions to be analysed took place. A subject (e.g. a designer) is given a task and instructed to verbalize his or her thinking process. No interviewer interrupts the train of thought. A contact person can be available for the subject to ask questions about the task but not for discussion about the problem-solving process. The actions are recorded, transcribed and analysed. Common criticism towards the think aloud method is that the designer might feel uncomfortable and not act naturally if an observer (or camera) is present. They might be distracted by the instruction to verbalize their thoughts, something people do not usually do. STAUFFER ET AL. claim that from their experience, "only one out of more than twenty designers" will actually claim to feel this way.[239] However, it does take a designer longer than under usual circumstances, while the content of the designer's performance remains unaffected.[240] A further limitation is the lack of time for incubation.[241] In real-time data collection, data collection, the researcher has to decide whether the designer should be informed about the problem beforehand to allow for incubation or whether the designer's spontaneous reaction and spontaneous performance is supposed to be part of the observation.[242]

The application of this research method in design science was first reported by EASTMAN who saw designing as something intuitive and applied the method to make those intuitive thought processes transparent.[243] ALBERS AND ALINK used the think aloud method to gain insight on designers' understanding of the concept of "functions".[244] GERO AND TANG did a study in which they compared retrospective and concurrent data collecting methods in design research.[245] They conclude that for process-oriented research, the results are comparable. For further reading VAN SOMEREN ET AL. provide a handbook on the application of the think aloud method. [246] Important groundwork on the topic has been done by ERICSSON AND SIMON.[247] They developed the necessary coding schemes and rigor to turn think aloud method and protocol analysis into a quantitative approach. GERO AND MCNEILL describe protocol analysis specifically for design research.[248]

---

[239] Stauffer et al. (1991), 357

[240] Stauffer et al. (1991), 357, Ericsson / Simon (1993)

[241] It can be shown that designers need time to think about the problem at hand before they come up with creative solutions – this is called incubation.

[242] Compare Stauffer et al. (1991), 357

[243] Eastman (1968)

[244] Alink (2010), Alink et al. (2011), Eckert et al. (2011)

[245] Gero / Tang (2001)

[246] Van Someren et al. (1994)

[247] Ericsson / Simon (1993)

[248] Gero / Mc Neill (1998)

Table 32 - Profile of “Think Aloud Method”

Think Aloud Method	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Human centered, process-oriented design research.
answers questions of the type:	*How* do the subjects approach / solve a problem? What are they thinking?
Advantages	Direct data collection, uninfluenced by the observer.
Disadvantages	▪ Subjects might deliver inaccurate data due to not being used to thinking aloud or because they feel in an exam like situation. ▪ Only applicable with artificial design tasks. Accessing actual industrial processes nearly impossible. ▪ Real-time setting does not allow for incubation, an important factor in creative activities
Necessary preparation	▪ Make the subject feel comfortable with the situation. Make clear that it is not of concern whether or not a good/ideal solution is developed, it is not an exam, a challenge or the likes. ▪ Set up audio recording and or video recording. ▪ Explain task to the subject. ▪ Prepare/select a coding scheme
Follow-up work	▪ Transcription of collected data (estimate effort: 1 Minute of protocol → 10 Minutes transcription) ▪ Analysis of collected data
Related techniques	▪ Retrospective Protocol ▪ Introspection ▪ Prompting ▪ Interviews
Synonyms	▪ Real-time protocol ▪ Concurrent protocol ▪ Protocol analysis ▪ concurrent thinking (method)

4.2.4.3 Introspection

If observer and observed subject are one and the same subject, we speak of introspection. Introspection means “looking into our own minds and reporting what we there discover.”[249] In the context of design research, different names have been assigned to the same type of research. PEDGLEY calls it *practice-led research* while ARCHER calls it *research through design.*[250]

It is the study of one’s own mental processes and originates from both psychology and philosophy. “*Cogito, ergo sum.*" The famous Latin quote by DESCARTES meaning: “I

---

[249] Boring (1953), 170;

[250] Both are mentioned in Pedgley (2007), 463

think, therefore, I am" is a popular example connecting introspection to philosophy. In the past, the philosophical question whether the world as we recognize it with our senses really exist or whether our cognition is just some imaginary state of mind, was pondered long before introspection became popular with a movement within psychology that founded "The new experimental psychology" in 1850-1870.[251]

Introspection is criticized as a scientific method for data collection and not very common in design research, at least, it is rarely explicitly mentioned. Although, no definition is given anywhere as to how long typical introspection studies in design research last, it seems they are more suitable for longitudinal research lasting several weeks, months or even years. Observing oneself for a day is not likely to give any deep insights: Furthermore, the validity of such data is rather questionable.[252]

Nevertheless, introspection is especially interesting for doctoral projects with the researcher being employed within an engineering company. It seems that those PhD- or doctoral constellations are becoming more and more popular. They deliver direct insight into engineering design processes under real-life conditions, which makes them fruitful and valuable for the design research community. Some of these studies – not all of them – can be regarded as introspection, as long as the object under investigation is the design activity of the doctoral candidate himself / herself. It is, however, a difficult research method. It should be applied with great care and good preparation preferably some training: "Its pursuance of course requires that the researcher is also a skilled designer and is prepared to combine the two roles of scholar and designer: something that is known to be intellectually challenging."[253] Consciously observing design activity is extremely demanding because design is something that is believed to happen in large parts unconsciously or subconsciously.

Obviously, in a longitudinal approach, under the "double role" of participant and observer, the researcher must use some type of support for the data collection. PEDGLEY lists the aspects listed in Table 33 to be considered carefully and consciously when deciding on how to collect one's own data. According to PEDGLEY, a *diary* is the only tool that will fulfill the aspects from Table 33 completely after comparing a set of 12 possible data collection methods. With this in mind, introspection becomes very similar to the general application of design diaries and also to ethnographic studies. What distinguishes introspection from these methods is the focus of the research. While the other two methods aim at a holistic perspective from within the design process, introspection is used when the cognitive processes of the designer (0the researcher) within the design project are to be analysed.

Table 33 - Aspects of consideration for data collection through introspection (Pedgley

---

[251] See Boring (1953) for more detail

[252] Just imagine your findings from introspection are based on the data from "a bad day".

[253] Pedgley (2007), 463 referring to Archer (2004) and Hales (1986)

(2007), 469)

Solo effort	Opportunities to employ a second researcher to fulfill a data collector or analyzer role may not exist. Data collection must therefore be executable as a solo effort.
Endurance	Data collection must be compatible with a longitudinal design project, spanning months if not years.
Subject delimitation	Without subject delimitation, literally all aspects of design activity are candidates for capture. This would result either in data overload and researcher fatigue or data dilution, caused by too much breadth and too little depth. Data collection must therefore be carefully directed towards the specialist subject of the research.
Mobility	Data collection must allow designing to be carried out in multiple locations, such as a studio, workshop and home, as is normal for a longitudinal project.

Table 34 - Profile of "Introspection"

Introspection	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Human-centered, Longitudinal projects that aim to generate insights on design activities and processes, in which the data source is potentially the researcher himself/ herself ▪ Doctoral projects with the researcher being employed at a company (where the research is mainly taking place)
answers questions of the type:	▪ Why are decisions made in a certain way? ▪ How are they made? ▪ How does a designer percept certain situations during the design process?
Advantages	▪ Direct data collection, no interference between observer and data source. ▪ Conscious self-observation might lead to improving self-learning and awareness, making researcher eventually a more skilled designer.
Disadvantages	▪ Conscious self-observation can distract from design activity → upsetting the natural rhythms of activity. ▪ Danger of post-event rationalization, when direct data recording is impossible (a diary is written "in the evening", after the events) ▪ Lure of dishonesty (to show oneself) in a good light → Modified behavior or modified self-report.
Necessary preparation	▪ Training with unimportant example project. ▪ Focused definition of research subject to avoid huge amounts of data
Follow-up work	▪ Transcription if video/audio recordings where applied ▪ Reduction of data to focus on research subject ▪ Interpretation of collected data through systematic procedure
Related techniques	▪ Ethnographic studies ▪ Think aloud method ▪ Design Diary
Synonyms	▪ practice-led research ▪ practice-led research

4.2.4.4 Retrospective Protocol

When retrospective protocols are written, subjects are instructed to protocol their design activities from memory. GERO AND TANG claim that since this recollection of memory consists of both short- and long-term-memory, the results may be incorrect or imprecise. "Retrieved data from [long-term memory] may have details omitted or may be generated by reasoning rather than recall. As a result, some researchers utilize videotapes of the design session as cues during retrospection to assist in the recall of the design activity. "[254] However, in their study, they conclude that in process-oriented research, concurrent and retrospective protocol lead to comparable results.

---

[254] Gero / Tang (2001), 284

Table 35 - Profile of "Retrospective Protocol"

Retrospective Protocol	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Human centered design research. ▪ When summary is preferred over details.
answers questions of the type:	▪ *How* do the subjects approach / solve a problem? ▪ What are they thinking? ▪ What happened during the design process?
Advantages	▪ Design activities uninfluenced by observer or unnatural situation. ▪ Real design activities from actual industrial processes are accessible. ▪ Data is already summarized when collected.
Disadvantages	▪ Does not produce many details, rather a summarized version of what happened. ▪ Indirect data collection, subjects might deliver inaccurate data due to: ▪ lost details, they cannot remember ▪ reasoning mixed into reported memory
Necessary preparation	▪ Optionally record design activities on video to assist memory collection during retrospection. ▪ Make the subject feel comfortable with the situation. Make clear that it is not of concern whether or not a good/ideal solution is developed, it is not an exam, a challenge or the likes. ▪ Set up task or identify an actual design process ▪ Prepare/select a coding scheme
Follow-up work	▪ Transcription of collected data (estimate effort: 1 Minute of protocol → 10 Minutes transcription) ▪ Analysis of collected data
Related techniques	▪ Concurrent protocol / think aloud ▪ Introspection ▪ Prompting ▪ Interviews
Synonyms	▪ Backward design protocol ▪ Design notes ▪ Analysis from memory

4.2.4.5 Ethnography / Ethnographic studies

Ethnography is a research method in which the researcher becomes part of what (s)he studies. “The basic tools of ethnography use the researcher’s eyes and ears as the primary modes for data collection.”[255] STAUFFER ET AL. call it “*participant observation.*”[256] Its roots are in anthropology where it was originally used to explore a people’s culture (e.g. MALINOWSKI temporarily lived with and observed the people of the Trobriand Islands in Eastern New Guinea to study their culture and how they trade goods).[257] In the second half of the twentieth century, ethnography was applied in

---

[255] LeCompte / Schensul (1999), 7
[256] Stauffer et al. (1991)
[257] Malinowski (1921)

different settings, when an interest in social processes in (urban or social) subcultures started to develop within sociology. Scientists would disguise as “street workers” to learn about youth poverty or live in a prison, disguised as a prisoner to learn about prison hierarchy and violence.[258] Today, most reported ethnographic studies focus on children and education. A new type of ethnography is evolving, as social behavior online, in forums, discussion groups and such, is becoming a popular filed for scientific investigation. A famous example of an ethnographic study is the *Rosenhan Experiment*: The researchers in this experiment, perfectly healthy people, were admitted to different mental institutions by faking symptoms of auditory hallucinations in order to study how well the staff would distinguish mentally ill from fit patients. After they stopped faking their hallucinations, they were still found to be insane by the personnel of the institutions and forced to stay and take medications. In a second step, a hospital challenged ROSENHAN to provide some “fake patients” and they would detect them. This time, the hospital detected a total number of 41 potential pseudo patients (out of the total of 193 patients at the hospital) that were found not to be mentally ill. To the misfortune of the hospital's reputation, the total number of pseudo-patients that ROSENHAN had actually smuggled into the hospital was zero.[259]

ATKINSON & HAMMERSLEY characterize ethnography as follows:[260] “In practical terms, *ethnography* usually refers to forms of social research having a substantial number of the following features:

- a strong emphasis on exploring the nature of particular social phenomena, rather than setting out to test hypotheses about them
- a tendency to work primarily with "unstructured" data, that is, data that have not been coded at the point of data collection in terms of a closed set of analytic categories
- investigation of a small number of cases, perhaps just one case, in detail
- analysis of data that involves explicit interpretation of the meanings and functions of human actions, the product of which mainly takes the form of verbal descriptions and explanations, with quantification and statistical analysis playing a subordinate role at most”

In a design engineering context, ethnographic studies have been conducted e.g. by BAIRD ET AL.[261] They studied large design teams within engineering projects at Rolls Royce under realistic conditions, including organizational restructuring during the project and a change of methods such as the introduction of new software. MYERS published a “tutorial for conducting ethnographic studies in information systems research” to show its value and limits as a research method for research in information

---

[258] The first such studies have been conducted a century earlier. Engels (1987), the original work to Engels’ study was published in 1845.

[259] The original study was reprinted in Rosenhan (1972), some scientist challenged the study and call it “pseudo-science”, e.g. Spitzer (1975)

[260] Atkinson / Hammersley (1994), 248

[261] Baird et al. (2000)

systems.[262] BUCCIARELLI reports two studies in which the design process within engineering companies was under investigation. “The studies were based on participant-observation techniques: in each case, the firm was approached in the way an ethnographer might approach a foreign culture.”[263]

BJÖRK AND OTTOSSON argue that “to grasp what really happens on a daily basis in a development project, to get the opportunity to reflect upon it, and to understand the complex nature of a development process, it is necessary to conduct insider action research (IAR).”[264] What they call IAR is what sociologists would call an ethnographic study. The researcher participates and interacts with those being observed.

The studies above are all explicitly regarded as ethnographic research by the scientists who published them. What cannot be summarized here but should be mentioned is the large number of studies that take place in doctoral programs where the doctorate candidates are part of an engineering design team within a company. These are often simply called “external doctoral projects” by those sitting "inside" a university office, as this is the more common setup. From an ethnographer’s point of view, those “external doctoral projects” would probably be called “internal doctoral projects”. The researcher is obviously inside the process being studied and those who observe from their university office are the ones in an external position.[265]

Within the engineering design community, such a research setup is often called a case study. While there are many similarities, it seems that the term case study is used without reflection by many. The social sciences have definitions of what a case study is. An ethnographic study is not the same. However, it can be part of a case study. According to MYERS, “the main difference between case study research and ethnographic research is the extent to which the researcher immerses himself or herself in the life of the social group under study. In a case study, the primary source of data is interviews, supplemented by documentary evidence such as annual reports, minutes of meetings and so forth. In ethnography, these data sources are supplemented by data collected through participant observation. Ethnographies usually require the researcher to spend a long period of time in the ‘field’ and emphasize detailed, observational evidence.”[266]

---

[262] Myers (1999)

[263] Bucciarelli (1988), 159

[264] Björk / Ottosson (2007), 195

[265] The author does not intend to raise the discussion, neither does he want to judge who is internal or who is external. It is an interesting thought though. From the author’s point of view, the two are simply different research designs, both have their merits and their limitations and both have delivered progress for the design research community in the past.

[266] Myers (1999), 4 referring to Yin (1994)

Table 36 - Profile of “Ethnography”

Ethnography	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Implementation studies, real-world deployment of design support. ▪ Human-centered, Longitudinal projects that aim to generate insights on design activities and processes, in which the data source is a social group of which the researcher becomes a member, e.g. doctoral projects with the researcher being employed at a company (where the research is mainly taking place).
Answers questions of the type:	▪ Why are decisions between individuals dealt out in a certain way? ▪ How do they agree? ▪ How do the designers interact in certain situations during the design process?
Advantages	Direct data collection as the observer is part of the data source.
Disadvantages	▪ Conscious observation may distract from design activity → upsetting the natural rhythms of activity within the team. ▪ Danger of post-event rationalization, when direct data recording is impossible (a diary is written “in the evening”, after the events). ▪ Lure of dishonesty (to show oneself) in a good light if the team members are aware of the researcher amongst them. ▪ → Modified behavior.
Necessary preparation	▪ Training with unimportant example project. ▪ Focused definition of research subject to avoid huge amounts of data. ▪ Identification and involvement usually of a company willing to participate.
Follow-up work	▪ Transcription if video/audio recordings where applied. ▪ Reduction of data to focus on research subject. ▪ Interpretation of collected data through systematic procedure.
Related techniques	▪ Introspection ▪ Think aloud method ▪ Design Diary
Synonyms	▪ Participant observation ▪ Insider action research

4.2.4.6 Design Journal Analysis

Another popular approach to observe design processes over a *longer period of time* (several weeks up to years) is to instruct designers to write a journal for later evaluation. The origin of the technique is said to have been first utilized by CHARLES DARWIN. To study the growth and influencing factors of babies, he instructed their mothers to keep a diary which he later evaluated.[267]

---

[267] Coley et al. (2007), 318 referring to Darwin (1877)

Table 37 - Profile of "Design Journal"

Design Journal / Diary	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Implementation studies, real-world deployment of design support. ▪ Human-centered, projects that aim to generate insights on design activities and processes, in which the data source is a group of designers. The researcher is not part of the group.
Answers questions of the type:	▪ What activities where conducted to achieve a certain goal? ▪ Who did what and when did they do it?
Advantages	▪ data is collected in real-time, in situ[268] ▪ no specially trained professionals required ▪ no observer present, hence natural activities are not affected ▪ little effort for researcher during data collection, even for large sample size ▪ quantity of data captured, while still large, is more manageable than e.g. video recordings ▪ No transcription necessary
Disadvantages	▪ Danger of incomplete data (imperfect records or unawareness of importance of information). ▪ Designers might protocol events in retrospect rather than as they occur → post-protocol rationalization without researcher knowing ▪ If students are involved, danger of anxiety (does the journal influence my grade?) ▪ Attrition in long term studies: Diary entries diminish in quantity and quality towards the end of the study.
Necessary preparation	▪ Training and feedback during the project (optional) ▪ Motivate participants to keep an accurate journal ▪ Develop guiding questions for the participants to ease effort of journal keeping and focus the records (optional) ▪ Select/prepare a coding scheme for later data analysis
Follow-up work	▪ Structuring the collected data (coding scheme) ▪ Interpreting collected data ▪ Optional: Present to and review with participants the results to ensure plausibility
Related techniques	▪ Introspection ▪ Think aloud method ▪ Ethnographic studies
Synonyms	▪ Diary Method ▪ Analysis of design notes

The approaches vary concerning the designers' freedom when writing their journal. In some studies, checklist-like questionnaires are handed to the designers to be answered on a weekly schedule.[269] Others have a less structured approach. E.g.

---

[268] Meaning in reality and not in a laboratory or classroom setting

[269] Ball et al. (1994)

SOBEK ET AL. asked student designers to write down what they deemed important during a design project for later evaluation.[270] ALBERS AND BRAUN have conducted several studies in which student designers as well as designers in industry were instructed to use document their design activities in the vocabulary of the integrated product engineering model (iPeM). The early studies allowed the participants to simply take notes. Later studies provided a software tool for the documentation allowing for less freedom but easier analysis.[271]

The latter approach is generally less likely to bias the journal entries. It does not fixate the designer on certain subjects through questions. However, it bares the risk of incomplete data. The researcher has no possibility to intervene or motivate the journal writing designer to go into detail. What the researcher deems interesting might be thought unimportant by the designer and hence be left undocumented.

Other advantages of design journal analysis according to SOBEK ET AL. are: "Compared to interviews, retrospective, and depositional methods, the data is collected in real-time, but unlike observational approaches, [journal analysis] does not require specially trained professionals and avoids the possibility of artificially altering [the observed subject's] behavior by having an observer present. Like protocol analysis, the data can be readily quantified using a suitable coding scheme, but it requires little researcher intervention during data collection, and data is collected from processes *in situ* rather than in a laboratory setting. It is also more feasible to collect a relatively large sample size compared to videotaping or other approaches because the quantity of data captured, while still large, is more manageable."[272]

Disadvantages according to SOBEK ET AL.: "Journals may offer an incomplete record of the design process. Where designers either keep imperfect records or are unaware of important information, the journals may fail to capture critical details regarding the development of the design project. [Designers] may 'backfill,' that is, record events in retrospect rather than as they occur, which can lead to omissions of key information. Training and feedback during the project can help [the designers] improve their record keeping skills and discipline, but ultimately, an accurate journal record depends on the designer's commitment to keeping a good journal."[273]

4.2.4.7 Content Analysis

Instead of artificial documents (e.g. journals which the participants are instructed to write), the data-sources can be objects that *naturally occurred* during a project (e.g. sketches, e-mails, minutes, reports, prototypes …). These are analysed through *content analysis*. Originally, content analysis is a technique applied in communication studies. Content such as newspaper reports about a politician can be used to gain

---

[270] Sobek / Jain (2007)
[271] Albers / Braun (2012)
[272] Sobek / Jain (2007), 12
[273] Sobek / Jain (2007), 12

insights on his/her popularity. This is why media based content analysis is used in electoral campaigns. Another popular field for the application is modern consumer research. The Internet is developing into a whole new field for content analysis. How quickly is a video clip on YouTube about a new smartphone passed on through the Internet? What do people in discussion forums have to say about it? How does “the community” on Twitter react?[274]

In content analysis, explicit data is analysed based on a pre-defined coding scheme.[275] The researcher looks for the occurrence of terms or categories of terms, the context in which certain terms/categories are being used, or the source for certain terms/categories. E.g. transcripts of recorded design meetings can be scanned for the use of certain terms such as 'innovation', 'new ideas', “USP” or 're-design', 'adaption', 'variation', 'reuse' and so on. The terms are assigned to categories, e.g. 'Adaption' and 'Innovation'. Now the presence of innovative attitude as opposed to adaptive attitude (compare 4.2.2.5) can be determined on a quantified basis.

Usually, several coders perform the analysis on the same set of data. Their results are compared, and the coding scheme is improved in iterations until the inter-coder-reliability has improved to something better than 80%, meaning that the results of the different coders are almost the same (see chapter 4.3.4 for a full definition of inter-coder-reliability). Alternatively, an established coding scheme from previous studies can be applied.

A famous study using the method of content analysis was conducted by NAISBITT in the nineteen seventies. His team analysed over two million newspaper articles over a period of twelve years identifying 10 Megatrends.[276] Today, it is a well-known and has been used so often in innovation-related projects and everyday language that it has become a buzzword. In 1982, NAISBITT introduced the term predicted the next century to be driven by an information society. At this time, Apple computers were still sold as building kits and made of wood and IBM sold the 200.000th IBM PC worldwide.

The application of content analysis in engineering design projects is not commonly reported. However, many studies and PhD thesis include some type of content analysis. ALBERS ET AL. conducted a study on the change of design as a profession that included a content analysis of the complete lecture courses at two German universities.[277] WIEDNER conducted a content analysis of CAD drawings at a power-tool manufacturing company.[278]

---

[274] E.g. Wu et al. (2011)

[275] See also 0

[276] Naisbitt (1982); The term “megatrend” was unheard of when the study was first published.

[277] acatech (2012)

[278] Wiedner (2013)

Table 38 - Method profile - Content Analysis

Content Analysis	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Human-centered, projects that aim to generate insights on design activities and processes, in which the data source is a group of designers. The researcher is not part of the group. ▪ Design research concerning relation between content and process.
Answers questions of the type:	▪ Who communicates how and when and why with whom during the design process?
Advantages	▪ The collected data is genuine and uninfluenced by the research ▪ No observer present, hence natural activities are not affected ▪ Little effort for researcher during data collection, even for large sample size
Disadvantages	▪ Danger of overwhelming amounts of data. ▪ Time consuming ▪ Iterations necessary to define final coding scheme ▪ No measure to determine the completeness of data.
Necessary preparation	▪ Training and feedback during the project (optional) ▪ Motivate participants to keep an accurate journal ▪ Develop guiding questions for the participants to ease effort of journal keeping and focus the records (optional) ▪ Select/prepare a coding scheme for later data analysis
Follow-up work	▪ Structuring the collected data (coding scheme) ▪ Interpreting collected data ▪ Optional: Present to and review with participants the results to ensure plausibility
Related techniques	▪ Observation of sketching activities ▪ Protocol analysis ▪ Coding schemes
Synonyms	▪ none

4.2.4.8 Observation of sketching Behaviour

It is important to distinguish between the *analysis of sketches* and the *analysis of sketching behavior.* According to COLEY, “Many researchers believe that one area that holds the most interesting cognitive information in design is a designer’s sketching behavior." Its observation has been assigned a somewhat extraordinary role. From a methodological point of view, all aspects of observation of arbitrary design activities apply.[279] The majority of analyses of sketching behavior in the context of design has in the 20th century been conducted in projects that dealt with interior design and

---

[279] Sketching too can be investigated through an unstructured or a structured observation. See chapter for general aspects of observation. 4.2.4.1

architecture. E.g. SUWA AND TVERSKY analysed how students and professional architects interact with their own free hand sketches.[280]

Table 39 - Profile of "Observation of sketching"

Observation of sketching	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies, real-world deployment of design support ▪ Design cognition research ▪ Interaction of designers tacit knowledge and explicit drawings
Answers questions of the type:	▪ What doe designers sketch? ▪ How do they sketch?
Advantages	▪ The created sketches are additional documentation ▪ Combined with an evaluation of the sketches, a holistic picture can be generated.
Disadvantages	▪ Narrow field of research. ▪ No established coding schemes for the analysis of sketching behaviour. ▪ Sketching behaviour is influenced by individual problem solving style. Generalization of data becomes therefore difficult.
Necessary preparation	▪ Arrange a setting in which sketching can be recorded ▪ Develop coding scheme
Follow-up work	▪ Organize data so that recordings an sketches can be assigned to one another ▪ Structure and analyze recorded data
Related techniques	▪ Observation
Synonyms	none

In 2002, TOVEY AND PORTER applied the method within the automotive industry. They did video observations of the sketching activities of post-graduate students and six professional designers.[281] They conclude that if CAD systems are to support concept development, they must acknowledge the importance of sketching activity. SONG AND AGOGINO followed shortly after. They studied student designers in new product development teams and come to the conclusion that “the volume of total sketches, and the number of 3-dimensional sketches has an increasingly positive effect on the design outcome as the design proceeds from preliminary investigation, through conceptual design, to detailed development and testing. Results also show that there is a statistically significant correlation between the total number of individual journal

---

[280] Suwa / Tversky (1997)
[281] Tovey et al. (2003)

sketches created during the design process and an individual student's class grade."[282] A similar study has later been conducted by YANG. She also observed students with the question whether large numbers of sketches lead to better design results. She concludes that the "volume of dimensioned drawings generated during the early-to-middle phases of design were found to correlate with design outcome, suggesting the importance of concrete sketching, timing and milestones in the design process."[283]

4.2.4.9 Case study research in engineering design

Books have been filled with information on case study research.[284] It is a powerful yet demanding method in qualitative research that scientists have been – and probably will be – arguing about. There are aspects for and against the method and whoever applies it should expect criticism from those who disagree the method. In the 1990s, qualitative research has gone out of fashion a bit in the social sciences, and *quantitative case studies* have been suggested. They basically combine different quantitative methods and are used e.g. in medical research and special education.[285] For the sake of focus in this thesis, only qualitative case study research is reviewed. However, any data collected in a case study is open to structuring and quantitative evaluation as described in some of the other research methods in this collection. Distinguishing between the two types is neither necessary nor useful in design research.

Case study research is generally seen as a technique which allows in-depth insight on a single and usually complex case. ABERCROMBIE ET AL. describe: "The detailed examination of a single example of a class of phenomena, a case study cannot provide reliable information about the broader class, but it may be useful in the preliminary stages of an investigation since it provides hypotheses, which may be tested systematically with a larger number of cases."[286]

Concluding from this definition, a single case study would be limited to exploratory research that aims to detect phenomena within a domain that are then subject to additional inspection. Other scientists diminish its value further. CAMPBELL AND STANLEY write: "Such studies have such a total absence of control as to be of almost no scientific value [...] Any appearance of absolute knowledge, or intrinsic knowledge about singular isolated objects, is found to be illusory upon analysis. [...] It seems well-nigh unethical at the present time to allow, as theses or dissertations in education, case studies of this nature (i.e., involving a single group observed at one time only)."[287]

---

[282] Song / Agogino (2004), 1

[283] Yang (2009), 1

[284] E.g. Yin (1994), Stake (1995)

[285] Stake (1995), xi f.

[286] Abercrombie et al. (1984), p. 34, cited in Flyvbjerg (2006), 220

[287] Campbell / Stanley (1966), 6 f., cited in Flyvbjerg (2006), 220

FLYVBJERG on the other hand, argues that these are false accusations and identifies *five misunderstandings* about case study research:[288]

- *"Misunderstanding 1:* General, theoretical (context-independent) knowledge is more valuable than concrete, practical (context-dependent) knowledge.
- *Misunderstanding 2:* One cannot generalize on the basis of an individual case; therefore, the case study cannot contribute to scientific development.
- *Misunderstanding 3:* The case study is most useful for generating hypotheses; that is, in the first stage of a total research process, whereas other methods are more suitable for hypotheses testing and theory building.
- *Misunderstanding 4:* The case study contains a bias toward verification, that is, a tendency to confirm the researcher's preconceived notions.
- *Misunderstanding 5:* It is often difficult to summarize and develop general propositions and theories on the basis of specific case studies."

From a design science perspective, case studies are a powerful way to combine industrial research and even consulting with thorough observation. Well planned, laboratory observations at the university can be combined with real-live engineering. Comparisons can be drawn between students and experienced engineers/designers. Nevertheless, attention should be paid: "Hidden agendas" can bias the results. Designers might be unhappy about cooperating because they feel it costs them extra time and similar pitfalls. On the other extreme, the "scientific attention" can motivate practitioners to "over-cooperate" i.e. they try to help the researcher and engage in activities they would not conduct without the researcher present.

Another difficulty with case study research is that the term has been misused in the past to the extent that it has become a real buzzword. It seems that if none of the "orthodox research methods" apply, scientists call it "a case study". Such misleading terminology must strictly be avoided as it not only diminishes credibility for one's own work but also of thoroughly conducted case studies by colleagues.

---

[288] Flyvbjerg (2006), 221

Table 40- Profile of "Case Study Research"

Case Study	
Field of application within design research	▪ Exploratory research aiming to identify research questions ▪ Empirical research analysis of real-world design processes ▪ Implementation studies, real-world deployment of design support ▪ Investigation of complex situations, when the goal is a holistic picture. ▪ Identification of hypothesis ▪ Falsification of theories ▪ Showing usability / value of a support tool
Answers questions of the type:	▪ What is worth further investigation? ▪ How do the elements in a complex situation connect? ▪ Does a predicted phenomenon really occur?
Advantages	▪ Holistic approach ▪ Works even with very complex situations, unsuited for other types of data collection
Disadvantages	▪ Ongoing argument whether or not it is a valid research method ▪ Effort, due to the necessary application of several research methods
Necessary preparation	▪ Clear definition of the case and the goals of the research to increase credibility
Follow-up work	▪ In case of explorative research, develop hypothesis from data analysis
Related techniques	▪ Ethnography
Synonyms	none

4.2.4.10 Application of Coding Schemes

With coding schemes, it is possible to objectify unstructured data (protocols, transcripts of interviews / recordings). A coding scheme is a *set of rules* for the reduction of raw data material. E.g. couples of designers and managers are told to discuss a given problem. Video recordings are afterwards cut into 10-second frames. For each frame, it is then e.g. noted who had the main part in the talking, and which were the dominating gestures based on a set of 15 body gestures handed to the analysts beforehand. Now, comparisons can be drawn based numbers.

From this example, the main goal of coding schemes becomes clear: To have a transparent, reproducible set of rules for the analysis of raw data. This ensures test-retest reliability (4.3.4.1). If the analysis is repeated with the same set of rules (coding scheme) on the same raw data, the result should be the same. The same applies for parallel test reliability (4.3.4.2). If two analysts are given the same coding scheme and the same data recordings, they should come to equal results. Subjective interpretation through the researcher is reduced. It is beneficial to a research community if a *limited*, established *set* of coding schemes exists. If the same coding scheme is used in different studies, their results become comparable and further insights on the field can be gained from such comparison.

Table 41 - Profile of "Coding Schemes"

Coding Schemes	
Field of application within design research	▪ Empirical research, analysis of real world design processes ▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies, real world deployment of design support
Answers questions of the type:	▪ Are certain patterns recognizable? ▪ Do certain elements in the observed behaviour appear more often than others?
Advantages	▪ Objectifies data ▪ Key to quantification of otherwise qualitative data
Disadvantages	▪ Effort for preparation to ensure high reliability ▪ Few coding schemes available for reuse in design science
Necessary preparation	▪ Identification of the elements in the coding scheme suitable for the task ▪ Refinement of coding scheme until acceptable level of inter and intra operator reliability is achieved
Follow-up work	▪ Data analysis ▪ Graphical representation
Related techniques	▪ Protocol Analysis ▪ Content Analysis
Synonyms	▪ Transcript coding

In engineering design research COLEY emphasizes the works of GERO AND MCNEIL and SUWA ET AL.[289] They have developed coding schemes to support the research efforts towards understanding design thinking from written and verbal protocols. While GERO AND MCNEIL provided a process oriented coding scheme, SUWA ET AL. focus more on content.[290]

4.3 Scientific Methods from other Field of Science

The following subchapter contains a selection of methods that have not been developed explicitly for design science but can be easily adapted for application in design research. Apart from some rather *general techniques*, a subchapter is devoted

---

[289] Gero / Mc Neill (1998), Suwa et al. (1998)

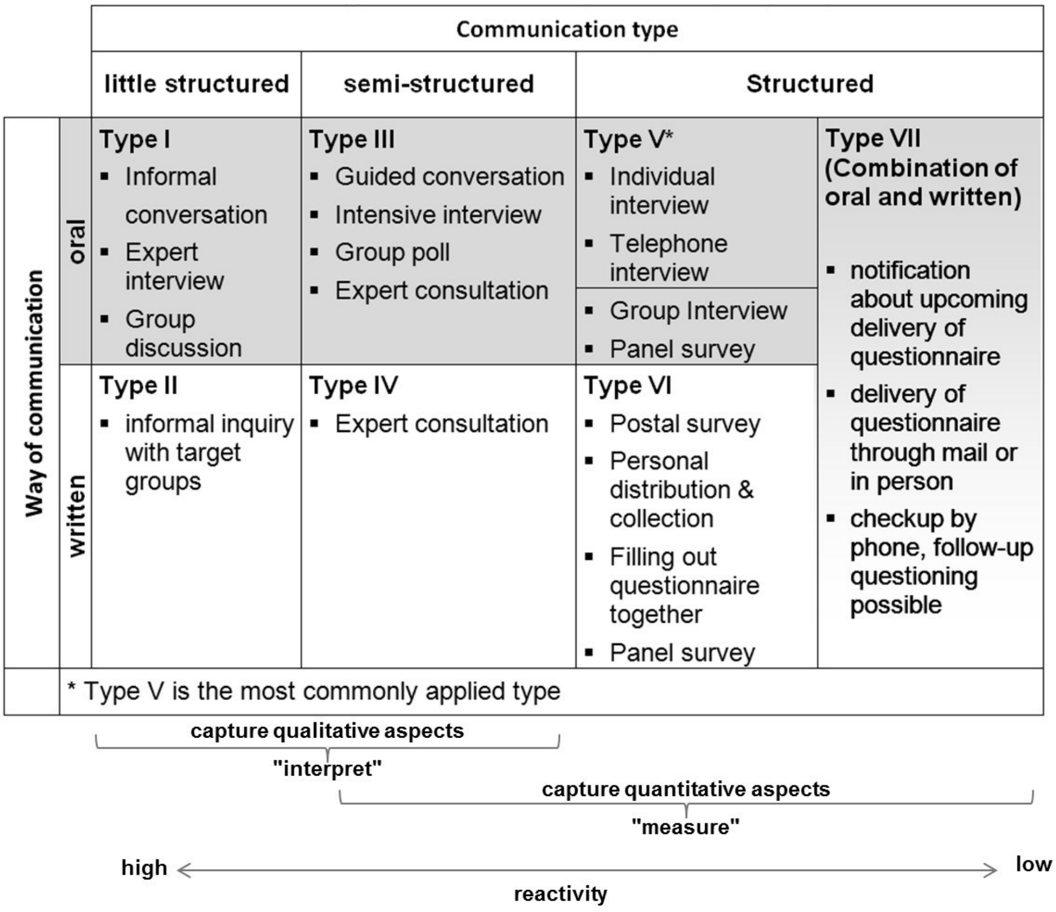
[290] See also Coley et al. (2007)

to *statistical methods*, especially for *small sample size*. BENDER and others have argued in the past that design studies are particularly challenging since it is extremely difficult to generate sample groups that are both large enough to produce statistically significant results and at the same time are composed of realistic representatives of the target population. In plain English: It is a challenge to find an adequate number of real designers in real projects instead of just students in academic projects.

4.3.1 Data Collection Techniques

Somewhere along a research project, data is collected and analysed. Various aspects of data collection within design research are summarized below. They are grouped into content and process oriented approaches in chapter 4.2.3 and 4.2.4. The following is a collection and description of methods for the collection of data that originate from fields of science other than engineering design research and that are not specifically described for the application in design research.

Table 42 - Types of data collection (Atteslander (2008), 123)



		Communication type			
		little structured	**semi-structured**	**Structured**	
Way of communication	**oral**	**Type I** ▪ Informal conversation ▪ Expert interview ▪ Group discussion	**Type III** ▪ Guided conversation ▪ Intensive interview ▪ Group poll ▪ Expert consultation	**Type V*** ▪ Individual interview ▪ Telephone interview ▪ Group Interview ▪ Panel survey	**Type VII (Combination of oral and written)** ▪ notification about upcoming delivery of questionnaire ▪ delivery of questionnaire through mail or in person ▪ checkup by phone, follow-up questioning possible
	written	**Type II** ▪ informal inquiry with target groups	**Type IV** ▪ Expert consultation	**Type VI** ▪ Postal survey ▪ Personal distribution & collection ▪ Filling out questionnaire together ▪ Panel survey	
	* Type V is the most commonly applied type				

ATTESLANDER uses the scheme shown in Table 42 to categorize types of data collection. He distinguishes between oral and written communication, on the one hand, and the degree of structure given by the method on the other.

4.3.1.1 Interview Techniques

One of the most popular techniques for the acquisition of data that involves human actions is interviewing. Chatting out of curiosity and conducting a scientific interview are two very different things. It is advisable for scientists who have not been trained to apply the method(s) with great caution and not without prior preparation. Posing questions is not a trivial matter. Books have been filled on the subject, especially within sociology. The following paragraphs summarize a brief overview designed to raise awareness about the complexity, variety, and difficulty of interview studies, and about potential bias induced by the interviewer.[291] ATTESLANDER gives a set of guidelines on how to pose questions. These apply to all forms of oral data collection techniques. Worksheet 8 in the appendix summarizes the most important aspects in a guideline on how to pose questions, scientifically.

Types of Interviews

Literature on *interview techniques* suggests a range of categories, many overlap.[292] Different authors describe equivalent techniques using different terms. Scientists should therefore be careful when putting a tag on their work. Below is a list of possible categories with a brief description of each. Beware that most interviews do not fit into just one of the categories, neither is it common that an interview is conducted “strictly by the book”. Therefore, usually a mixture of two or more techniques applies. An additional aspect worth mentioning is that interview studies are always – to some extend – biased. It is nearly impossible to standardize the situation, the questions and the emotional situation for both the interviewer and the interviewee. Therefore, one has to be careful to derive and generalize findings from only a small number of interviews.

Narrative Interview:[293] Though initiated by the interviewer through an opening question, from thereon the narrative interview is spontaneous. The interviewer asks back and sums up to keep the interview going. The interviewee has most of the talk time and attention while the interviewer is more restrained. Narrative interviews are a preferred method, when the experience of the interviewee is the topic of interest.

Problem-centered Interview (PCI):[294] PCI is applied as a non-directive dialogue about a problem. The interviewee is informed beforehand and can mentally prepare. The interviewer may use a list of detailed questions but formulates the questions spontaneously to keep the dialogue-like nature of the interview situation. The interviewer has to be very familiar with the topic, so this type of interview might be typical if a researcher is trying to elaborate specifics of a problem (s)he might want to

---

[291] For more comprehensive publications on the subject refer to: Atteslander / Kneubühler (1982) or to Atteslander (2008)

[292] Suggested reading on interview techniques:
German literature: Helfferich (2005), Atteslander / Kneubühler (1982), Atteslander (2008)
English literature: Gubrium / Holstein (2002), Gubrium (2012), Saunders et al. (2012), Kvale (1996), King (2010), Flick (2009)

[293] Suggested reading: Clandinin / Connelly (2000), Riessman (1993), Flick (2009)

[294] Flick (2009), 161 ff., Witzel / Reiter (2012)

address in his or her research. For example: reasons why ideas are not followed through in large engineering companies.

As a variation of problem-centered interviews, the interviewer and interviewee may develop a visualization of the interview, e.g. a mind map.

Episodic Interview:[295] Type of narrative interview where the interviewee talks about individual, past experience, in particular situations (episodes). The interviewer may use a guideline to ensure that all relevant episodes are being addressed. This type of interview is useful, when the researcher is looking for potential improvements. So two situations are imaginable:

- A researcher is trying to find out about bad and good design practices to derive potential research or
- A researcher has equipped the interviewee with some type of design support and questions the interviewee about her experience with its application to gain insight for revision and improvement of the support.

(Semi)-structured interview or semi-guided interview:[296] This is a planned and prepared yet flexible type of interview. The interviewer prepares some guidelines that only get used if it seems necessary. Since this broad description fits most interviews, many scientists in their documentation will stick to it in their publications even if the interview had some specialties to it. One should use a more specific description if possible.

Guided Interview:[297] Interviews in which the interviewer uses a guideline prepared beforehand are called guided interviews. Most commonly the interviewer will note a set of questions to make sure (s)he asks every subject the same questions. In some cases even the exact order of questions might be relevant.

Focused Interview:[298] The interview is prepared by the interviewer. It is an artificial situation where a central stimulus is placed in front of the interviewee. The stimulus can typically be a picture of an object or the object itself. In engineering design research a technical device or a technical drawing is imaginable as well as some type of design support, e.g. a software tool.

Biographical Interview:[299] Interview to determine a person’s past experiences – his or her biography. Job interviews, in which the candidate verbally presents his or her curriculum vitae are an everyday-life example for biographical interviews. However, if experience and social context play a major role a design research project, this format could apply as well.

---

[295] Flick (1997) ,Flick (2009) 185 ff., Atkinson et al. (2000)

[296] Schensul et al. (1999), Flick (2009), Drever / Education (1995)

[297] Keats (1999), 17

[298] Flick (2009), 150 ff.; Keats (1999), 16

[299] Mathis / Jackson (2008)

Ero-epic Interview or Ethnographic Interview:[300] Both interviewer and interviewee participate in an open dialogue. No guidelines are being used; no pressure to answer is put on the interviewee. Suggestive questions are explicitly allowed in this type of interview. Usually, the dialogue is recorded for later transcript.

Discursive Interview:[301] These types of interviews are follow-up interviews. After an initial interview or poll and data analysis, the researcher conducts a further interview, discussing the findings and interpretation aiming to validate findings from the first interview.

Dilemma Interview:[302] Here, the interviewee is confronted with one or more (moral) dilemmas. Used in psychology or sociology to determine an individual's value system. It is not likely to find those types of interviews in engineering design.

Construct Interview:[303] Variation of guided interviews that combines different (psychological) question techniques. The dialogue builds around constructs while the goal is to reveal attitudes/emotions towards certain topics. Constructs can be e.g. pictures, hierarchical positions in a company, people/colleagues and so forth. Assessment centre situations in job interviews make use of such techniques to determine whether an applicant is a team player, alpha leader and so on. They should only be applied with the assistance of trained psychologists. Construct interviews are unlikely to be applied in design research.

---

[300] Schensul et al. (1999), Spradley (1979)

[301] Berndt (2011)

[302] Friebertshäuser (2003)

[303] König (2005)

Table 43 - Profile of "Interview Study”

Interview studies	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies, real world deployment of design support ▪ All fields of design science where peoples’ perception of a situation or a process is of interest. Also when the perception of different groups is of interest (e.g. management vs. designers)
answers questions of the type:	Qualitative interview studies: ▪ How do designers feel about something? ▪ How do they something predict to turn out/develop? ▪ Is there a difference between certain groups in how they describe something/perceive something/ feel about something? Quantitative interview studies: ▪ What is the most common answer to question X? ▪ How many answered Y?
Advantages	▪ Direct data acquisition into peoples thoughts.
Disadvantages	▪ Many sources of bias. ▪ A lot of effort has to be put into each interview: preparation, conduction transcription, interpretation documentation
Necessary preparation	▪ Select and contact interviewees. Prepare a guideline (what do I want to ask?). In some cases, deliver the questions beforehand, so that interviewee can think about the answers.
Follow-up work	▪ Transcription of the interview. ▪ Interpretation ▪ In some cases, discursive interview.
Related techniques	▪ Poll ▪ Survey ▪ Questionnaire
Synonyms	▪ Interrogation ▪ Questioning ▪ Survey

4.3.1.2 Survey

Similar to a guided interview, a survey is used to collect peoples’ perception, opinion, or attitude towards a predefined set of questions, directly. The difference is that no interviewer is present to assist the interviewee. Instead, a questionnaire is prepared and provided to a selected group of people. It is usually used when large numbers of individuals are to be questioned. The results can be analysed quantitatively. If the questioned individuals belong to a selected group, findings about this group can be derived or typical commonalities of members of this group can be revealed. With a rising number of software tools and distribution channels online, surveys are becoming even more popular. It is a very common technique and similar to interview studies. It holds a lot of pitfalls for inexperienced researchers. A lot of literature can be found within marketing research as surveys are one of the most widespread tools to investigate market needs. An important factor is that it is in most cases impossible to repeat the survey as it is to be expected that many participants will refuse to answer a second time. Therefore, once a survey is launched, it has to be perfect. Things that can go wrong are given in Table 44:

Table 44 - Potential weaknesses of surveys

Things that can go wrong	**Countermeasures**
Participants will feel exploited, e.g. if they suspect a commercial motivation.	Explain very clearly what the results are intended to be used for. In commercially motivated surveys, incentives are common. E.g. Each participant automatically takes part in a lottery and has the chance to win something.
When analyzing the data, something unexpected comes up that leads to further questions which have not been asked in the survey.	Pilot study with a reduced number of participants.
The survey is too long and participants do not answer the last few questions honestly but simply finish it as quickly as possible.	Pilot study with colleagues. A few test runs will reveal how long the survey really takes. It is most tempting to include too many questions “just in case”. 10 questions that reveal honest answers are by far more valuable than 30 questions if one cannot be sure about the quality of the answers.
Leading Questions that makes participants feel that a certain answer is expected.	Formulate neutral questions. Include as many positive as negative answers. Keep in mind that participants show a tendency to want to help the researcher.
Things to be taken into account:	**Rule of thumb**
The length of the survey	Aim for about 10 Minutes. More than 20 questions “feel” a lot.
The order of the questions	Difficult questions that need more thinking in the beginning. General information (gender, age, profession and so on) in the end.
If scales are offered, the correct layout of the scale is very important	People tend towards the middle → even number of possibilities forces to decide. Too many possibilities seem more precise but confuse the interviewee (6 to 10 max) If answers are to be quantified, equidistant scales are common If the answers are represented in ranges, put the most likely answer in the middle of the range (example see section 4.3.2 Use the same scale within one survey for all questions if possible

Table 45 - Profile of "Survey"

Survey	
Field of application within design research	▪ Empirical research, analysis of real-world design processes ▪ Experimental studies, evaluation in controlled environment ▪ Implementation studies, real world deployment of design support ▪ All fields of design science where peoples' perception of a situation or a process is of interest. Also when the perception of different groups is of interest (e.g. management vs. designers)
answers questions of the type:	▪ What is the most common answer to question X? ▪ How many answered Y? ▪ Are subgroups within the sample apparent?
Advantages	▪ Direct data acquisition into people's opinions. ▪ Easy access to large sample groups, especially with internet based surveys. ▪ Large number of free online-survey tools available.
Disadvantages	▪ Researcher cannot intervene. ▪ Precise questions have to be formulated. Misunderstandings / misinterpretation through the participants can ruin the complete study.
Necessary preparation	▪ Select and contact participants. ▪ Prepare the survey ▪ Pilot study with small number of participants to determine how long it takes and find possible inconsistencies. ▪ Pilot evaluation to determine possible missing questions.
Follow-up work	▪ Data analysis and data representation (tables graphs)
Related techniques	▪ Poll ▪ Ballot ▪ Interview
Synonyms	▪ Questionnaire

4.3.2 Statistics

Analyzing and interpreting collected data requires some knowledge about statistics and its most important methods. Within statistics, there are two very different types. On the one hand, there is '*descriptive statistics*'.[304] It deals with questions such as: "How can we describe a sample?"; "How can we characterize a measurement?"; "How can we represent collected data?"

---

[304] For definitions of descriptive and inferential statistics, see also Anderson / Finn (1996), 19. ff

On the other hand, there is a branch of statistics that aims at drawing inferences about a population of data from a sample of data taken from this population. This branch is called '*inferential statistics*' For the description of data, *scales* are an important tool. For data analysis and inference, there are a number of different *statistical operations* available.[305]

4.3.2.1 Scales

Nominal scales[306] are part of descriptive statistics. They are used to describe or characterize a sample group. A nominal scale divides characteristics of the group's individuals into different peculiarities without any type of ranking. Example: Imagine a class-picture of 30 students, some have red hair, some blonde and so forth. We can count how many of the students are taller than X, separate by gender and so on.

Ordinal scales / ranking scales[307] are also used to describe a sample group (→ descriptive statistics). An ordinal scale *ranks* the group's individuals according to different characteristics. Example: Customers are asked how satisfied they are with a product {'very satisfied' | 'satisfied' | 'unhappy' | 'very unhappy'}.

An ordinal scale or ranking scale is especially useful to characterize "soft" values or to group measurable characteristics (height can be measured in cm, but if the group members are of similar age and reference values are available, we can divide the group simply into {tall | normal | small} people.

Interval scales[308] are suitable to characterize quantitative data. This way, mathematical operations are applicable (comparison, calculate differences and such). This is necessary for what is commonly called "statistical operations" (median, mean, distribution …). Example: If we take a look at the group picture from the example above again, the students' age or exact height, weight and so on could be put into an interval scale. After that, calculating the average weight would be possible.

4.3.2.2 Statistical operations

The goal of this section is to draw attention to some of the most common statistical operations apart from calculating the average. The basic types of scales described above allow the application of different types of statistical operations as shown in Table 46.

---

[305] For a comprehensive overview, refer to the standard literature, e.g. Anderson / Finn (1996)

[306] Anderson / Finn (1996), 33 f.

[307] Ibid.

[308] Ibid.

Table 46 – Possible operation with different types of scales

Statistical operation **Scale**	Mean	Median	Mode
Interval	✓	✓	✓
Ordinal	⊘	✓	✓
Nominal	⊘	⊘	✓

Mean (coll. "Average")[309]

The mean value *M* – in everyday language it is called 'average' – is calculated as the sum of all measurements *x(i)* in the sample divided by the number of measurements *n*.

$$M = \frac{1}{n}\sum_{i=0}^{n} x(i) \tag{12}$$

The mean value is easy to calculate and therefore, popular. However, it is more valuable in homogeneous groups than in mixed groups. Example: If we compare two competing basketball teams, calculating their mean height will give us an idea which team has a physical advantage. Let's say one teams mean height is 2,01 m and the other team's mean height is 2,06 m. Basketball players are usually somewhat tall and a basketball team is in this aspect a rather homogeneous group, so this value is quite suitable to describe the two samples. If we took a picture of the same team together with a group of kids (e.g. their fans). The average height might maybe be 1,77 m. What does this value tell us?

Median[310]

The median divides a group into two equal halves referring to one particular attribute. The data must be available at least in an ordinal scale. In the sample group in Table 47 a), the median is "Bob", since 3 designers (Fritz, Danny and John) have less experience, and 3 designers (Martha, Danielle and Sue) have more experience. In Table 47 b), the median is "4-6 years" since the same numbers of individuals are more experienced and less experienced. Compared to the mean, the median is less sensitive to extreme values. Especially for small sample sizes, which are rather typical for design research, this can play an important role.

---

[309] Anderson / Finn (1996), 77; Dodge (2008), 336

[310] Anderson / Finn (1996), 74; Dodge (2008), 346

Table 47 - Example sample; Design experience measured in weeks (a) and on an interval scale (b)

a)

Name	Working experience in engineering design in weeks
John	220 weeks
Bob	300 weeks
Martha	302 weeks
Danny	3 weeks
Sue	897 weeks
Danielle	560 weeks
Fritz	3 weeks

b)

Years of Working experience in engineering design	Individuals	Count
<2	Fritz; Danny	2
2-4	-	
4-6	Bob; Martha; John	3
6-8	-	
8-10	Sue; Danielle	2

Mode[311]

The mode is the value within a sample that occurs most often. The sample, therefore, must be described at least in a nominal scale. This value is more useful if there is only a limited set of peculiarities. Example: We take an imaginary group of designers from a large company and see how long their working experience in engineering design is very precisely (Table 47). The mode in a) is '3 weeks'. However, it is rather questionable whether this value is useful. It suggests that the company employs a lot of rookies. If we put the values in an interval scale like in table b), a different picture is indicated: the mode is 4-6 years. The work force of the company appears to be more experienced, which is actually the case.

4.3.2.3 Hypothesis Development and Hypothesis Testing

Building hypothesis is something not all design scientists are necessarily familiar with. As a consequence, it is sometimes mistaken for simply stating something of which one does not know whether it is true or false. While this is actually not wrong, it is also only part of the story. What is quite striking, as well, is the fact that some design researchers seem to feel obliged to formulate a hypothesis, even if their research and research design does not necessarily have to be hypothesis driven. The result can look something like this: “Hypothesis: My method helps.” This is not a well formulated hypothesis which becomes comes obvious when the researcher tries to test it.

---

[311] Anderson / Finn (1996), 70; Dodge (2008), 351

Sociology and statistics, management research, and marketing research offer a lot of literature aiming to assist hypothesis building.[312] The following section is included here to present the necessary vocabulary and some of those criteria that help to formulate high-quality hypothesis and to help decide whether a hypothesis driven approach makes sense at all.[313]

Null Hypothesis and Alternative Hypothesis[314]

In engineering design research and in the validation of design support, a most common scenario would be that one has developed some type of design support and now wants to test its effects on designers. This can be done by stating a null hypothesis and an alternative hypothesis and then testing the *null hypothesis significance level* with a procedure called “Null Hypothesis Significance Testing (NHST)”.[315] The null hypothesis (H_0) would usually be that the method has *no influence*. If dealing with a parameter (μ) that is known for a population, the null hypothesis would be that the sample group is no different. In mathematical terms:

$H_0: \mu=x$

If this statement were true, a comparison between designers that use the method and any other designers should show no significant difference.

The *alternative hypothesis* (H_1) can be formulated in different variations. The basic assumption would in all cases be that *there is an influence*, hence a significant difference could be observed in a comparison. $H_1: \mu \neq x$

However, if a certain tendency is expected and supposed to be shown, the following statements could also serve as an alternative hypothesis. $H_1: \mu<x$ *or* $H_1: \mu>x$

Directional and non-directional hypotheses[316]

Hypotheses can be distinguished as directional hypotheses and non-directional hypotheses. In the first type, it is clearly stated whether the analysed factor in the sample is larger or smaller than in the population. The latter type does not differentiate. It only claims that there will be a difference. Directional hypotheses are tested with a one-tailed test design. Non-directional hypotheses are tested with a two-tailed test design (see also chapter 4.3.2.4).

Thesis Statement and antithesis[317]

Similar to a hypothesis is a thesis statement of which it is unknown whether it is true or not. Someone who proposes a thesis statement usually claims it to be true. The counterproposition would be called an antithesis. Both statements do not have to fulfill specific scientific criteria and should not be used in a scientific publication.

---

[312] Eisenhardt (1989), Anderson / Finn (1996)

[313] This is not a chapter on hypothesis building in science. I will focus on the most important terms and point the reader to some suggested literature for optional further reading.

[314] Dodge (2008), 249 f., 388

[315] see chapter 4.3.2.4

[316] Grinnell / Unrau (2010), 41

[317] Breach (2009), 23 f.

Trivial fact statement

One has to be careful not to mix the assumption of a fact such as “Prices for steel have risen over the past 6 months!” Such a statement, even if it is unclear whether it is true or not, is not a hypothesis. It can simply be looked up and answered true or false without the necessity to apply any statistical techniques.

4.3.2.4 Statistical Hypothesis Testing

A large amount of research is based on the attempt to falsify or accept a hypothesis.[318] The common procedure is to collect data and to analyze whether or not the data supports the assumption that a hypothesis is true or false. We then decide whether we accept or reject the hypothesis.

Significance in this context describes how (un)likely the experimental result has occurred by coincidence. It is usually represented by the p-value (sometimes called “achieved significance level”). If *p* is small, it is very unlikely that the results produced with a sample are different from the stated null hypothesis by coincidence. Although various methods exist to do statistical hypothesis testing, they all end up calculating the p-value and is it always interpreted the same way: Small p means that the sample is significantly different from the population and H_1 can be accepted. What exactly a small p-value is, is not clearly defined. However, many scientists have accepted the convention shown in Table 48:[319]

Table 48 - Convention for the interpretation of p-values[320]

p-value	The observed difference may be called :
$p > 0{,}1$	not significant
$p \leq 0{,}1$	marginally significant
$p \leq 0{,}05$	significant
$p \leq 0{,}01$	highly significant

Example:[321]

A design scientist develops a problem solving method and claims that his method helps to find more solutions for a certain type of problems. He tests a group of 40 designers (Group A) and comes to the result that without his method, the designers need 40 minutes (mean value) to solve the problem intuitively. He then tests a sample group of 33 designers (Group B) that apply the method. Those designers take 34 minutes in average. Is this due to the method, or could it just be coincidence?

None of the 73 participants knew the problem or its solution in advance. The experiments were conducted under comparable conditions. The null hypothesis and alternative hypothesis in such a case could be formulated as:

---

[318] On the condition that a research hypothesis has been formulated

[319] One has to be careful though. In the recent past, a debate has erupted about this convention Gigerenzer (2004), Gigerenzer (1998)

[320] More on p-values and significance, see, e.g.: Dodge (2008), 434f.; Anderson / Finn (1996) 392

[321] The values are made up. See Table 49 for example data.

H0: *There is no influence on the time needed to solve the example problem.* $\mu_A=\mu_B$

H1: *Designers using the method take equal or less time than those without it.*

$\mu_B \leq \mu_A$

This is a directional hypothesis. It is therefore tested with a one-tailed test (two-tailed would also include the option that designers take more time if they use the method). No direct information about the population "designers" is available. All we have is a sample ($n_A=40$) of that population. If several thousand tests with designers had been made, we would be very confident that the mean value of those tests represents the actual time average designers need to solve the problem without the tested method. The question is: How likely would a group of designers not using the method take the same time as or even less than our sample group B? $p(\mu_A \leq \mu_B)$

The test statistic T to compare the means of two samples (Anderson / Finn (1996), 475):

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} \qquad (13)$$

T : test statistic
$\bar{x}$: mean value sample x
$\bar{y}$: mean value sample y
s_x; s_y : standard deviation of sample x and sample y
n_x; n_y : sample sizes of sample x and sample y

Insert the example values:

$$T = \frac{2171{,}53 - 2059{,}97}{\sqrt{\frac{285{,}54^2}{40} + \frac{190{,}73^2}{33}}} = \mathbf{1{,}99}$$

For Sample sizes larger than 30, we can assume standard deviation. This means we are allowed to look up the p-Value in a table that contains the values for normal distribution (9.3.1).[322] *T* is just below 2. This means that if *H₀* were true, there is only a (100% - 97,13%) = 2,87 % probability that the 33 designers were so fast although the method had no influence. Since this is not very likely, we can discard H0 and accept *H1*. We are confident there is a statistically significant indicator that the method reduces the time needed to solve the example problem.

---

[322] Standard Distribution Tables can be found in any standard statistics book, e.g. Anderson / Finn (1996) or online, e.g. http://web.as.uky.edu/statistics/users/ascho4/STA200files/summer%202011/normal%20table.pdf;

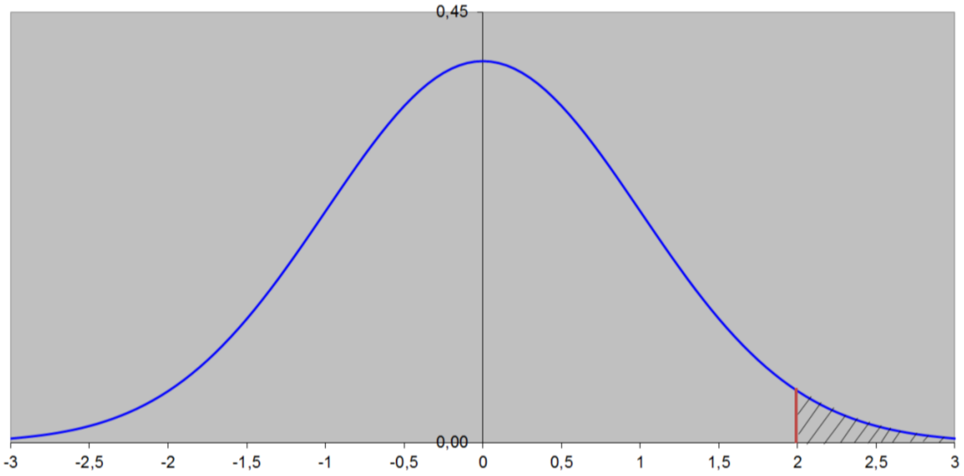


Figure 17 - Test statistic for Standard Normal Distribution, (source: own illustration, generated in MS Excel)

Table 49 – Hypothetical Example Data

Without Support [seconds]			
A1	*2156*	A21	*2384*
A2	*2127*	A22	*2149*
A3	*2263*	A23	*2312*
A4	*2628*	A24	*2226*
A5	*2179*	A25	*1922*
A6	*1330*	A26	*2268*
A7	*1989*	A27	*2156*
A8	*2272*	A28	*2397*
A9	*2397*	A29	*1951*
A10	*2455*	A30	*2437*
A11	*2217*	A31	*2084*
A12	*2666*	A32	*1663*
A13	*2525*	A33	*2063*
A14	*2421*	A34	*1516*
A15	*2290*	A35	*2002*
A16	*1831*	A36	*1873*
A17	*2410*	A37	*2201*
A18	*2486*	A38	*1648*
A19	*2164*	A39	*2262*
A20	*2299*	A40	*2242*
Standard Deviation	**285,54**		
Mean	**2171,53**		
Sample Size	**40**		

With support [seconds]			
B1	*2050*	B21	*2179*
B2	*1750*	B22	*2064*
B3	*1982*	B23	*2040*
B4	*2188*	B24	*1900*
B5	*2187*	B25	*1892*
B6	*1695*	B26	*1935*
B7	*2021*	B27	*2066*
B8	*1996*	B28	*2397*
B9	*2094*	B29	*2346*
B10	*2373*	B30	*1791*
B11	*2074*	B31	*2635*
B12	*2217*	B32	*1893*
B13	*2090*	B33	*2178*
B14	*1962*		
B15	*2121*		
B16	*2102*		
B17	*1921*		
B18	*1953*		
B19	*1895*		
B20	*1992*		
Standard Deviation	**190,73**		
Mean	**2059,97**		
Sample Size	**33**		

4.3.2.5 The Central Limit Theorem and Law of large Numbers

As engineers have their laws of thermodynamics, statisticians have the central limit theorem: "*If the sample size is large, the distribution of the sample mean of n different observations is well approximated by a normal distribution.*"[323] The central limit

---

[323] Anderson / Finn (1996), 330

theorem is easily confused with the law of large numbers: *"If the sample size is large, the probability is high that the sample mean is close to the mean of the parent population."*[324]

The law of large numbers only states that if you want to find out the mean of variable *x* of a population, a large sample will help you get closer to the actual mean of the population.[325]

The central limit theorem states that if you divide that large sample into several smaller samples and put the mean of each sample on a graph, that mean will be distributed normally around the population's real mean. Many statistical operations are mathematically derived from that statement. It is the central law that allows making assumptions about a population without necessarily studying each individual of that population.

4.3.2.6 Sample Size

How many individuals have to be studied to be confident one has generated reliable findings about the population of interest? Will sample size *n* be large enough?

The larger a sample the better, that is, if time and money are not important. Realistically, design research is restricted by constraints like available participants, limited time and funding. Under these aspects, extremely large sample sizes appear wasteful, and it is important to weigh:

- Which sample size is large enough to produce reliable results?
- From which sample size on, would a further increase hardly improve the results any further?

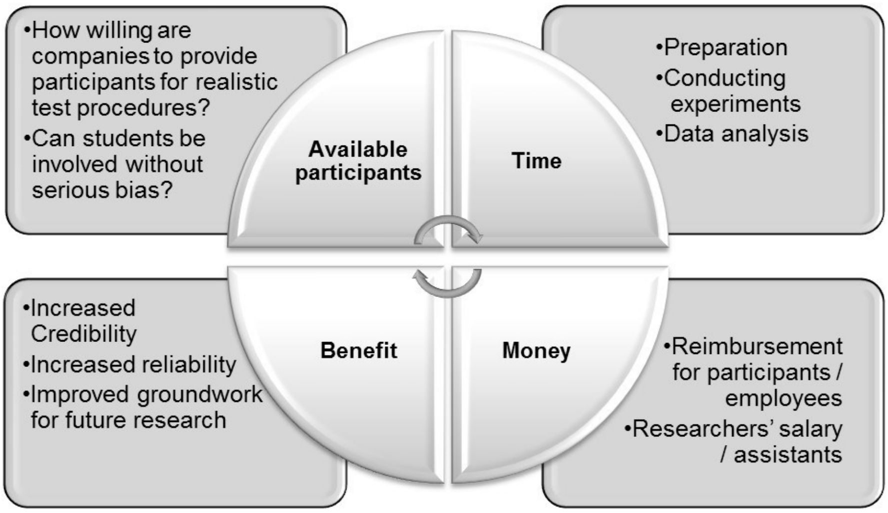


Figure 18 - Finding the ideal sample size under realistic conditions in design research (own illustration)

---

[324] Anderson / Finn (1996), 328

[325] The law of large numbers is generally accounted to Jacob Bernoulli who delivered the first mathematical proof of it. Further reading: http://www.encyclopediaofmath.org/index.php/Law_of_large_numbers, 10/06/2013, 16:17

Figure 18 shows that it is impossible to give one ideal sample size. A scientist has to decide under the given circumstances, so the sample size is a result of the applicable conditions. Under the aspect of hypothesis testing, GOSSET developed the so called "Student's Distribution" as he was considering how to achieve statistically sound results even if only small sample sizes are available.[326] The Student's Distribution or often T-distribution is a family of distributions depending on the sample size *n*. They are symmetrical and bell shaped similar to the normal distribution and – in agreement with the central limit theorem – for $n \rightarrow \infty$ become the normal distribution. Student's Distribution takes into account that for smaller sample sizes, one cannot be as confident about any findings from the data as for large samples. Hence, for smaller *n* it becomes flat compared to a normal distribution, i.e. the probability for extreme values is higher. The area under the bell-shaped curve to the right of a given value in the right-hand tail is larger than the area to the right of that same value under a normal distribution curve. This will be important to keep in mind for the following chapters in which a selection of typical test cases is presented.

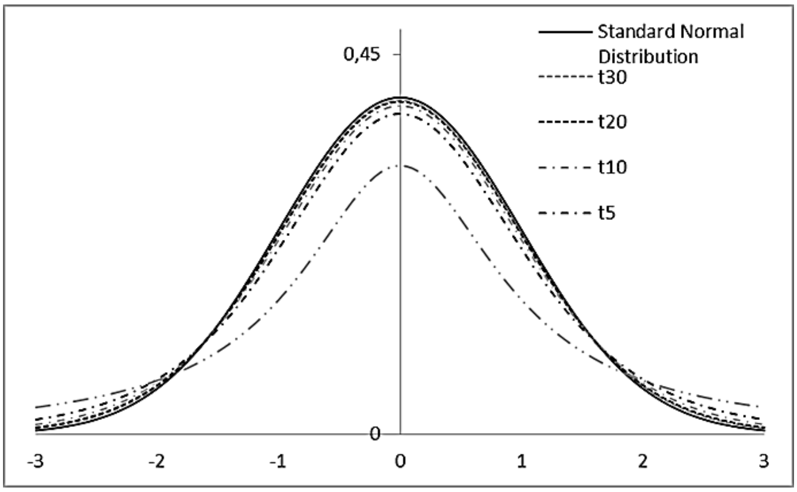


Figure 19 - Normal Distribution and Student's t-Distribution for 1, 5, 10, 20 and 30 degrees or freedom (own MS Excel illustration)

The exact shape is determined by the degrees of freedom *t=n-1*. Figure 19 shows that with growing sample size, Student's Distribution approximates the standard normal distribution. For *t=30*, Student's Distribution is hardly distinguishable from a normal distribution. This is the reason why many authors assume normal distribution for sample sizes larger than 30. Some claim that *n=20* to be a large enough sample. In some mathematical operations there might be an advantage in assuming normal distribution. However, if a table with Student's Distribution is available (e.g. 9.3.2) one might as well just look there. The results differ, but the effort is the same. With a computer at hand, MS Excel provides all the necessary functions as well.

---

[326] According to Dodge (2008), 234 f., Gosset wanted to stay unknown because his employer – the Guinness brewery – did not support scientific publishing. He therefore submitted his paper under the pseudonym "Student". The today famous Student's Distribution was published in: Student (1908).

Apart from the sample size, the appropriateness of the chosen samples is also important. Some questions for self-check are given in Worksheet 10, section B.

4.3.2.7 Common Test Cases[327]

Hypothesis testing deals with the question: “How different is an observed value from the expected value if *H_0* was true?” or in mathematical terms: “How distant is the observed value from the expected value if *H_0* was true?” The *distance* is represented by the test statistic *t*. As explained in chapter 4.3.2.3, one starts with the assumption that the null hypothesis is true. In the case of a design support, we assume that although a design support has been introduced, there is no influence on the designers. We basically try to prove that it doesn’t work. Only if we are unable to prove that *H_0* is true, we have no choice, but to reject it.

If the expected value for H_0 was μ_0 and the observed value is $\bar{x}$, that difference can be calculated from the ‘distance’ between the two: $\bar{x}$-μ_0. The distance alone does not include any information about the absolute values of μ and μ_0. Neither is any information about the quality of the observation included, which makes it impossible to compare different observations. In order to standardize the ‘distance’, statisticians divide it by the *standard error ($s_{\bar{x}}$)* which again includes the *standard deviation* (*s*) and the *sample size (n)*. The obtained value is called a *test statistic (t)*. Definitions for these terms can be found in standard literature, such as the Cambridge Dictionary of Statistics.[328] A comprehensive online resource can be found at www.encyclopediaofmath.org.

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \qquad (14)$$

standard error:[329]

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (15)$$

standard deviation:[330]

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \qquad (16)$$

test statistic:[331]

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \qquad (17)$$

---

[327] Note that all the procedures in this chapter are common and standard procedure in statistical data analysis. The methods are described in a large number of different textbooks and are not the author’s creation. The chapter summarizes those methods, the author found potentially useful for application in design research.

[328] Everitt / Skrondal (2010)

[329] Anderson / Finn (1996), 414

[330] Anderson / Finn (1996), 113

[331] Anderson / Finn (1996), 414

t: test statistic

$s_{\bar{x}}$: *standard error of measured sample*

s: standard deviation in the measured sample

n: sample size

In certain cases, the standard deviation might be known exactly from very large samples in past studies.[332] In that case, a standard normal distribution can be assumed, and s is replaced by σ. In design studies however, this will rarely be the case.

$$z = \frac{\bar{\mathrm{x}} - \mu_0}{\sigma_{\bar{\mathrm{x}}}} \quad (18)$$ [333]

with $$\sigma_{\bar{\mathrm{x}}} = \frac{\sigma}{\sqrt{n}} \quad (19)$$ [334]

z: test statistic for standard normal distribution

$\sigma_{\bar{x}}$: standard error of measured sample

σ: standard deviation in the population the sample was taken from

The basic test statistic $t = \frac{\bar{\mathrm{x}} - \mu_0}{s/\sqrt{n}}$ can be applied to a vast number of cases. It is a measure for how far the sample mean is located from the parent population's mean. In the example (page 115 ff.), the test statistic is *t=1,99*. The area under the curve to the right of *t* equals 2,87% of the total area under the standard distribution curve.[335] This means only 2,87% of the values that belong to the parent population will be further away from the mean than 1,99σ. The corresponding surface percentiles to the values for *t* can be looked up in tables. As a rule of thumb, for samples of *n=30* and larger, standard distribution can be assumed and the corresponding tables can be used. For sample sizes with *n<30*, use Student's Distribution for n-1 degrees of freedom. Sometimes, $\bar{x}$ *and* μ_0 can be obtained directly, sometimes they have to be calculated from other obtainable values. The following paragraphs contain a selection of test cases that can be useful in design science. They will not contain the mathematical derivations or underlying models. Any comprehensive statistics textbook will include and explain step by step, where the formulas come from.[336]

---

[332] This is not uncommon in medical studies that are repeated over and over again or continuously updated, e.g. the birth weight of babies, average maximum age in industrial countries,…

[333] Anderson / Finn (1996), 397

[334] Dodge (2008), 508

[335] The vertical line in Figure 17 is 1.99σ to the right of the mean

[336] E.g. Anderson / Finn (1996), Rumsey (2011) for (English), Winker (2006), Zucchini et al. (2009) (German)

Testing hypotheses about a mean[337]

A common thing to do with obtained data is to calculate the mean. Usually, this will be the case, when data is collected from a sample group and one wants to find out whether that group differs in some characteristic from its parent population. E.g. designers are given some type of design support, and it is of interest if they perform better than “regular designers”. If we know from previous studies which performance to expect from designers without a support, we can compare it with that of our sample group and have to decide whether or not it is significantly different. If we describe the performance as data on an interval scale, we can also calculate the mean and the distribution.[338] At this point, we need to differentiate. The following instruction will explain both cases:[339]

- Do we really know about the parent population, i.e., can we be sure that what we know about “normal designers” is reliable enough to assume that it is true for any random sample of “normal designers”? In that case, we are testing a sample against a population.
- Or, and this will very likely be the more common scenario in design science, are we actually setting up *two groups,* one of which tests the design support, and the other one acting as the control group without the support? In this case, the rules for testing two samples against one another apply.

Step one: Build your hypothesis

H_0 is that the design support has no influence hence the mean of the sample group’s performance ($\bar{x}$) will equal the mean of the comparison group. If we compare against a parent population that mean will be indicated as μ_0. If testing two samples against one another, we will differentiate it as x_0.

Step two: Obtain the mean values

$\bar{x}$: Mean value of the observed sample
μ_0: Mean of population
x_0: Mean of control group

Step three: Get the standard deviation(s)

If a sample is compared against a parent population, the standard deviation of the parent population must be known. This is indicated as σ. If testing two samples against one another, they are differentiated as s_1 for the tested sample and s_0 for the control group. We have to calculate s_1 and s_2 from our obtained sample data (16).[340]

Step four: Calculate the test statistic

If a sample is *compared against a parent population.*[341]	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	*$\bar{x}$: sample mean* *μ_0: population mean* *n: sample size* *s: standard deviation in the measured sample*

---

[341] Anderson / Finn (1996), 414

If *testing two sample against one another,* the test statistic is:[342]	$t = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}}}$	$\bar{x}_1$: *tested sample mean* $\bar{x}_0$: *control group's mean* n_0: *sample size of control group* n_1: *size tested sample group* *s: standard deviation in the measured sample*

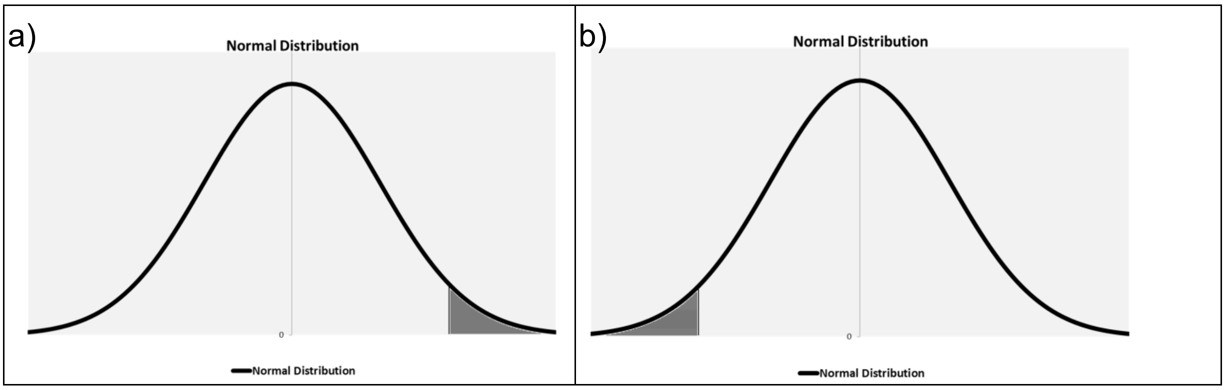
Step five: comparing t to a known distribution

The question that now has to be answered is: Which percentage of data is more extreme than *t*? The answer to this question is usually obtained from a table (or software which has the table integrated). This procedure is the same, no matter whether comparing two samples or sample vs. population. Before proceeding, two other cases need to be distinguished though:

First, we need to clarify if the sample size *n* is large enough to assume standard distribution of data; is $n \geq 30$?[343] This decides which table is to be used to determine the probability of data for values more extreme than *t*, even if H_0 was true.

- Yes ($n \geq 30$) → Use Standard normal distribution (Appendix 9.3.1)
- No ($n < 30$) → Work with Student's Distribution (Appendix 9.3.2)

Second, it needs to be clarified if it is a one-tailed test, or a two-tailed test. This depends on the hypothesis statement. If a certain direction can be expected, it is always advisable to include this information, formulate a directional hypothesis and to do a one-tailed test, as it increases the level of significance (in the example given on page 115 ff., a one tailed test was chosen). One-tailed hypotheses tests check whether an observed mean is significantly larger Figure 20 (a) or smaller Figure 20 (b) than the population mean, but not either one. Only the surface to the right Figure 20 (a) or to the left Figure 20 (b) is relevant.



---

[338] This is a prerequisite for testing means. If the data is not available on an interval scale, other tests must be used (see 4.3.2.1).

[339] I will not go into the mathematical reasoning as to why the two scenarios are being dealt with differently. Textbooks on statistics deal with this, e.g. Anderson / Finn (1996).

[340] MS Excel, SPSS, MATLAB and other Software tools all provide a command that returns the standard deviation of a vector.

[341] Anderson / Finn (1996), 414

[342] Anderson / Finn (1996), 475

[343] Some authors claim 20 to be a large enough sample, others 25. Yet others will use Student's Distribution even with sample sizes larger than 150.

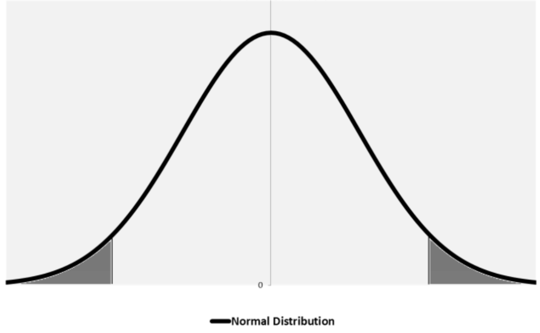
c) 	One-tailed tests: How likely is a value larger than the mean (a) or smaller than the mean (b). Two-tailed tests: Are two means significantly different? This can be found on the right as well as on the left extreme of the distribution (c). This decreases the level of significance or, at a given significance level, demands for more extreme observed means in order to reject the null hypothesis.

Figure 20 - Visualization of one-tailed an two-tailed test data (own MS Excel-illustration)[344]

Two-tailed tests, check for significantly different means which can be found on the right as well as on the left extreme of the distribution Figure 20 (c). This doubles the relevant area (the probability of such extreme data even if H_0 where true) under the curve and the level of significance or, at a given significance level, demands for more extreme observed means in order to reject the null hypothesis. The significance level can then be obtained from the correct table as listed in Table 50. In the example (page 115, ff.) a two-tailed test would have been applied if the researcher only wanted to test whether the design support changes the average time, designers take to solve a problem.

Hypotheses tests about a mean as described above rely on normally distributed or nearly normally distributed data. Before going into the test procedure, it is advisable to have a look at the data and make sure that it at least looks something similar to a bell-shaped shaped curve.[345] If the data does not look normally distributed, one should choose other hypotheses tests[346].

The significance level sometimes is used confusingly. This is due to some authors referring to the confidence with which the data indicates that the H_0 is true, others use the term to emphasize the confidence with which they believe H_1 to be true. So a 5% significance level for the null hypothesis is the same as 95 % confidence in its rejection. In this thesis, the term significance level will be used in reference to H_0. So a 1% confidence level is 'higher' in the terms of being more obvious than a 10 % level of significance. All instructions and described hypothesis tests will test the significance of H_0.

---

[344] Graphs of the standard normal distribution can be found in any statistics textbook, e.g. Anderson / Finn (1996), 288. Under the Wiki Commons Archive (http://commons.wikimedia.org) 190 graphs are listed (10/06/2013, 18:41) under the title "normal distribution". It goes back to Gauß (1777-1855) and is often also called Gauss-Distribution.

[345] Tests to check whether or not a set of data is normally distributed:.e.g. Anderson / Finn (1996)

[346] E.g. test for median, or proportion

Table 50 - Case dependent hypotheses tests

			Table to use	**How to read the table[347]**	**MS Excel**
One-tailed	$\bar{x}<\mu_0$	$n\geq30$	Standard normal distribution 9.3.1	Calculate *z*, find the nearest slightly <u>larger or equal</u> value in the table (e.g. for z=-2,534 go to line -2,5 and over to column 0,03). The value in that cell is the significance level.	To obtain the significance level of a test statistic z: English: = NORMSDIST(*z*) German: =STANDNORMVERT(*z*)
		$n<30$	Student's Distribution 9.3.2	Calculate *t,* go to row *n-1* (the degrees of freedom). Find the nearest value slightly <u>larger or equal</u> to t in this row and read the column heading α. This is the significance level.	To obtain the significance level of a test statistic t: English: =TDIST(ABS(t),n-1;1) German: =TVERT(ABS(t);n-1;1)
	$\bar{x}>\mu_0$	$n\geq30$	Standard Normal Distribution 9.3.1	Calculate *z*, find the nearest value slightly <u>smaller or equal</u> to z in the table (e.g. for *z= -2,534* go to line -2,5 and over to column 0,03). The value in that cell is your significance level.	To obtain the significance level of a test statistic z: English: = 1-NORMSDIST(*z*) German: =1-STANDNORMVERT(*z*)
		$n<30$	Student's Distribution 9.3.2	Calculate *t,* go to row *n-1* (the degrees of freedom). Find the nearest value slightly <u>smaller or equal</u> in this row and read the column heading α: (1- α) is the significance level.	To obtain the significance level of a test statistic t: English: =TDIST(t,n-1;1) German: =TVERT(t;n-1;1)
Two-tailed	$\bar{x}\neq\mu_0$	$n\geq30$	Standard Normal Distribution 9.3.1	Calculate z, find the nearest slightly <u>smaller or equal</u> value in the table (e.g. for z=+1,673 go to line 1,6 and over to column 0,07). 2x the value in that cell is the significance level.	To obtain the significance level of a test statistic z: English: = 2x(1-NORMSDIST(ABS(*z*))) German: =2x(1-STANDNORMVERT(ABS(*z*)))
		$n<30$	Student's Distribution 9.3.2	Calculate *t,* go to row *n-1* (the degrees of freedom). Find the nearest value slightly <u>smaller or equal</u> to t in this row and read the column heading α: **<u>2α</u>** is the significance level.	English =TDIST(t,n-1,2) German =TVERT(t;n-1;2)

Testing Hypotheses about Proportion[348]

Hypotheses about a proportion can also be common test cases. Here, hypotheses testing also works with data that is only available in nominal scales (e.g. answers from

---

[347] The tables in Appendix 9.3.1 and 9.3.2 give the area under the distribution curve to the left of the significance point. The instructions are for tables of that kind. Some statistics textbooks give the right tail instead; some only list the positive half (since the distribution curve is symmetrical). Different rules apply for those!

[348] See e.g., Anderson / Finn (1996), or Lewis-Beck (1995)
A tutorial can also be found online: http://stattrek.com/hypothesis-test/proportion.aspx, 10/1/2013, 18:15

a multiple-choice survey). In design research, an exemplary scenario would be that we observe designers and do continuous surveys on whether they find the defined processes within their company transparent enough. A certain proportion of them are unhappy with design-process-transparency. We then introduce a support to a sample group. The support has been developed in order to improve design-process-transparency.[349] After some time of application we ask the sample group as well and would like to know whether or not the proportion of designers that are unhappy is significantly smaller than in the parent population.

Again, a second scenario is imaginable. We take two groups of designers and equip one of the groups with the support. The control group is left without it, and we would like to find out if the proportion of designers that are unhappy is the same in both groups or not.

The test procedure is very similar to testing a hypothesis about a mean. We will eventually end up calculating a test statistic, depending on the case differentiation:

- *Testing a sample against a population*
- *Testing two samples against one another*?

The basic formula for test statistics based on sampled data: $t = \frac{\bar{\mathrm{x}} - \mu_0}{s/\sqrt{n}}$ or in cases of normally distributed data: $z = \frac{\bar{\mathrm{x}} - \mu_0}{\sigma_{\bar{\mathrm{x}}}}$ will be transformed. Instead of the mean value we compare to the population's mean, we now compare two proportions, so $\bar{x}$ is replaced by $\bar{p}$ and μ_0 is replaced by p_0. We get:

Test statistic for testing a sample's proportion against a population:[350]

$$z = \frac{\bar{\mathrm{p}} - \mathrm{p}_0}{\sigma_{\bar{\mathrm{p}}}} \tag{20}$$

The standard error $\sigma_{\bar{p}}$ cannot be calculated from the sample, since the data is qualitative. However, hypothesis about proportion can mathematically be interpreted as an event occurring with a certain probability. The rules of binomial distribution apply.[351] The proportion $\bar{p}$ of a sample (sample size n) is the same as the result of a Bernoulli experiment with n repetitions and the probability $\bar{p}$ of one of the two possible outcomes to occur for each round. In plain English: If we ask designers whether they are satisfied with the process-transparency or not, the event "yes" has a certain probability. If we ask n designers and x say "yes", the proportion of satisfied designers is $\bar{p}=x/n$ as is the probability of a future designer to answer yes if we ask him/her. In

---

[349] E.g. Albers et al. (2011a)

[350] Anderson / Finn (1996), 427

[351] E.g. Anderson / Finn (1996), 316 ff.

this case, the mean is $\bar{p}$. The proportion $\bar{p}=x/n$ is approximately normally distributed with mean $E(n,x)=\bar{p}$.

The standard error[352]	$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$	*with:* $q = 1 - p$	(21)

So to calculate the test statistic for testing hypotheses about a sample's proportion compared to a population, we can use:

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad (22)$$

with $\bar{p}$: *observed proportion after intervention*
p_0: proportion of population (before intervention)
n: sample size (after intervention)

If two samples are being compared with one another, the following changes have to be made to (22). The two proportions are named $\hat{p}_1$ for the tested sample and $\hat{p}_0$ for the control group. The standard error is the same as the standard error of the differences:

$$\sigma_{\hat{p}_1-\hat{p}_0} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

If H_0 where true, the proportions would be the same, which is what is being tested here, so the following simplification can be made: $p_1=p_2=\hat{p}$

The test statistic for testing hypotheses about two samples' proportion' is then:[353]

$$z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_0})}} \qquad (23)$$

with n_1: sample size test group
n_0: sample size control group
$\hat{p}$: *Proportion of individuals with characteristic in both groups* $\hat{p} = (a+b)/(n_1+n_2)$
$\hat{p}_1$: *Proportion of individuals with characteristic in test group* $\hat{p}_1=a/n_1$
$\hat{p}_0$: *Proportion of individuals with characteristic in control group group* $\hat{p}_0 = b/n_0$

Table 51 - Classification of two samples according to presence of a characteristic

---

[352] Anderson / Finn (1996), 414
[353] Cp. Anderson / Finn (1996), 479

(yes: present; no: not present), cp. Anderson / Finn (1996), 479

	Test group	**Control group**	**Both groups**
Yes	***a***	***b***	*a+b*
No	***c***	***d***	*c+d*
	$n1=a+c$	$n_0=b+d$	$n=n1+n_0=a+b+c+d$
proportions	$\hat{p}_1=a/n_1$	$\hat{p}_0=b/n_0$	$\hat{p}=(a+b)/n$

Testing hypotheses about a median[354]

The following test procedure is also documented as 'The Sign Test'. The median is less sensitive to extreme data (outliers), even more so in cases of small sample sizes.[355] *This makes it especially interesting for design research*. For larger sample sizes, skewed data is an indicator to apply median-based rather than mean-based operations as well (see Figure 21).

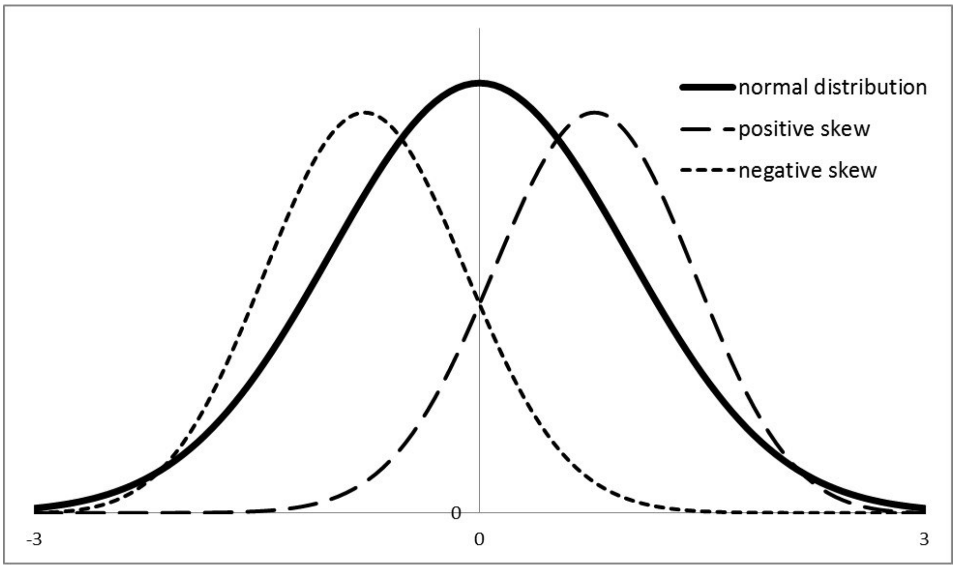


Figure 21 - Skewed data distribution (own MS-Excel illustration)

The median *M* is the number that divides the data into two halves, so drawing one random sample, the probability of getting a result greater than M is ½, the probability of drawing a result smaller than M is also ½. In other terms, the proportion of results larger than *M* is ½ and vice versa: $p(x<M)=0,5= p(x>M)$.

Possible H_0 and H_1 for hypothesis test are shown in Table 52:

---

[354] Sprent / Smeeton (2010), Anderson / Finn (1996), or Lewis-Beck (1995)

[355] Educational studies with classroom settings of about 20 students therefore often operate with median instead of mean values.

Table 52 - Null and alternative hypotheses testing a median

		Comparing a sample to a population	**Comparing two samples**
Null Hypothesis	H_0:	$\overline{M}=M_0$,	$M_1=M_0$,
Alternative Hypothesis (directional)	H_1:	$\overline{M}< M_0$	$M_1< M_0$
	H_1:	$\overline{M}> M_0$	$M_1> M_0$
Alternative Hypothesis (non-directional)	H_1:	$\overline{M}\neq M_0$	$M_1\neq M_0$
		$\overline{M}$: sample's median M_0: Populations median	M_1: test sample's median M_0: Median of the control group

To turn this into a test of proportion, we can take the sample observation and count the number y of results smaller than M_0.[356] If H_0 where true, the proportion y/n would be ½. The test statistics derived from (22):

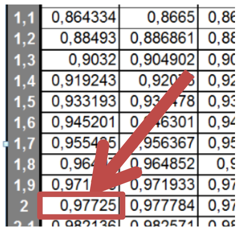
Comparing a sample to a population:[357]

$$z=\frac{\mathrm{y/n}-1/2}{\sqrt{\frac{1/2\times1/2}{n}}}=2\sqrt{n}(\frac{y}{n}-\frac{1}{2}) \tag{24}$$

with *y: number of results smaller than M_0*
n: sample size

Special case:[358]

For a two-tailed test and large sample size ($n>30$, standard normal distribution can be assumed) at a significance level of 4.55 % (which is ~5 %), the null hypothesis would be rejected if, $|z|>2$.[359] This allows for a very quick and easy test.



1,1	0,864334	0,8665	0,86
1,2	0,88493	0,886861	0,88
1,3	0,9032	0,904902	0,90
1,4	0,919243	[illegible]	0,92
1,5	0,933193	[illegible]	0,93
1,6	0,945201	[illegible]	0,94
1,7	[illegible]	[illegible]	0,95
1,8	[illegible]	0,964852	0,9
1,9	[illegible]	0,971933	0,97
2	0,97725	0,977784	0,97
2,1	0,982136	0,982571	0,98

$$z=\left|2\sqrt{n}\left(\frac{y}{n}-\frac{1}{2}\right)\right|>2$$

$$|\frac{y}{n}-\frac{1}{2}|>\frac{1}{\sqrt{n}}$$

If the difference between the medians exceeds $\frac{1}{\sqrt{n}}$, the difference is significant!

Testing the median of a sample against a control group

In a scenario where a sample is tested compared to a control group, the same assumptions apply, and the test is turned into a test for proportion from (23):[360]

---

[356] See page 119 f.

[357] Anderson / Finn (1996), 430

[358] Ibid.

[359] For $z=2$ the table returns *0.97725*, in a two-tailed test, the significance level is: $\alpha=2x(1-p)$, in this case: *2x(1-0,97725)=0,0455*. In a one-tailed test, the significance level would even be at 2,275%

[360] Cp. also Cp. Anderson / Finn (1996), 479

$$z = \frac{\hat{\mathrm{p}}_1 - \hat{\mathrm{p}}_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_0}\right)}} = \frac{\hat{\mathrm{p}}_1 - \hat{\mathrm{p}}_0}{\sqrt{\frac{1}{2} \times \frac{1}{2\left(\frac{1}{n_1}+\frac{1}{n_0}\right)}}}$$

Comparing two samples;

$$z = 2\sqrt{\frac{n_1 n_0}{n_1 + n_0}}\left(\hat{\mathrm{p}}_1 - \hat{\mathrm{p}}_0\right) \tag{25}$$

with $\hat{p}=1-\hat{p}=1/2$

$\hat{p}_1$: *Proportion of individuals with characteristic in test group* $\hat{p}_1=a/n_1$

$\hat{p}_0$: *Proportion of individuals with characteristic in control group* $\hat{p}_0 = b/n_0$

n_1: sample size test group

n_0: sample size control group

a: number of individuals with the tested characteristic in test group

b: number of individuals with the tested characteristic in control group

Testing Hypotheses of Equality of Proportion[361]

If pairs of individuals are observed but qualitative data is generated, the above-described *t-test for correlated samples* cannot be applied. However, in cases of dichotomous data, we can test whether the proportions of the matched pairs have changed. Such data can be organized in a turnover table.[362]

Example: If we take the example from the hypothesis test for proportion (p. 125) and make sure that both times, we question the same group of individuals, we can summarize the data as shown in Table 53.[363]

---

[361] E.g. Anderson / Finn (1996), Anderson (2013)

[362] Dichotomy is the separation between two groups (male-female, yes-no, satisfied-dissatisfied, …); See e.g. Anderson / Finn (1996), 182

[363] Alternatively, if we are sure, that both groups are comparable, we can do the test with two different groups. However, in that case, there is an increased risk to introduce bias into the test by changing the individuals, and a comparison of two independent sample proportions might be the more appropriate test procedure (page 119 f.)

Table 53 - Turnover table to organize data for paired hypotheses tests[364]

<table>
<tr><th></th><th></th><th colspan="2">Second survey</th></tr>
<tr><th></th><th></th><th>satisfied</th><th>not satisfied</th></tr>
<tr><td rowspan="2">First survey</td><td>satisfied</td><td>a</td><td>b</td></tr>
<tr><td>not satisfied</td><td>c</td><td>d</td></tr>
<tr><td colspan="4">a: is the count of designers who answered “satisfied” in the first and second survey.
b: is the count of designers who answered “satisfied” in the first but “not satisfied” in the second survey.
c: is the count of designers who answered “not satisfied” in the first but “satisfied” in the second survey.
d: is the count of designers who answered “not satisfied” in the first and second survey.</td></tr>
<tr><td colspan="4">b and c are the designers, who changed their mind between the first and the second time they were asked.
(b+c) is the total number of “mind changers”, <i>n</i>.</td></tr>
</table>

H_0 is that nothing has changed, so the proportion of those who have changed from “satisfied” to “not satisfied” must equal the proportion of those who have changed their minds vice versa. In other words if we disregard those who haven’t changed at all, the proportion of those who have changed in one way, compared to those who have changed in general is one half.

Mathematically, the null and alternative hypotheses would be:

Null Hypothesis	H_0:	$c/(b+c) = ½$
Alternative Hypothesis (directional)	H_1: H_1:	$c/(b+c) < ½$ $c/(b+c) > ½$
Alternative Hypothesis (non-directional)	H_1:	$c/(b+c) \neq ½$

With the test statistic for a proportion (22):

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{c}{b+c} - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2}\times\frac{1}{2}}{b+c}}} = \frac{\frac{2c}{b+c} - 1}{\sqrt{\frac{1}{b+c}}} = \sqrt{b+c}\left(\frac{2c}{b+c} - 1\right)$$

The test statistic for equality of proportions is calculated as:

$$z = \sqrt{b+c}\left(\frac{c-b)}{b+c}\right) \qquad (26)$$

with

p̄: observed proportion of mind changers in one direction amongst all mind changers

p_0: expected proportion of mind changers in one direction amongst all mind changers if H_0 where true *(=½)*

n: number of all mind changers *(b+c)*

---

[364] Cp. Anderson / Finn (1996), 437

According to ANDERSON and FINN, the minimum number of mind changers may not be lower than 25, otherwise the mathematical assumption of normally distributed binomial data cannot be made.[365]

Paired Measurements[366]

In certain experimental setups, paired measurements are taken, i.e. the exact same sample group is observed twice, before and after an event (such as exposure to a new design support). Similarly, if two observations of each individual are impossible without severe bias or not possible at all, *matched samples* can be used. Individuals with comparable characteristics are put into groups and randomly, one subject from each pair is exposed to whatever is to be tested. In design science, a typical scenario would be to construct pairs of designers with similar background, experience and creative intelligence and have one designer from each pair test a design support. The other designer of each pair has to solve the task without the support. In such cases, the test is also sometimes called "*matched t-test*" or "*t-test for correlated samples*".[367] Mathematically, the test-procedure is based on the null hypothesis *that the mean difference between the pairs is zero.* Hence, all data from the two groups must be available on the same interval scale. *H_0 and H_1* would be:

Null Hypothesis	H_0:	$\mu_d=0$
Alternative Hypothesis (directional)	H_1: H_1:	$\mu_d < 0$ $\mu_d > 0$
Alternative Hypothesis (non-directional)	H_1:	$\mu_d \neq 0$

Determine the distance between of the mean difference from zero:

$$t = \frac{\bar{\mathrm{d}} - \mu_\mathrm{d}}{s}$$

with

$$s = \frac{s_{\bar{d}}}{\sqrt{n}}$$

$\bar{d}$: mean difference of the pairs
μ_d: 0
n: sample size → The number of pairs, not individuals!
$s_{\bar{d}}$: standard deviation of the differences of all pairs

The test statistic for paired measurements is:[368]

$$t = \frac{\bar{\mathrm{d}}}{s_{\bar{\mathrm{d}}}}\sqrt{n} \qquad (27)$$

Example:

Some design scientists claim that ideation is improved if the subjects designers are exposed to the topic and a certain time of incubation is allowed for subconscious ideas to emerge and develop. Hence, it is advisable to inform the participants of a

---

[365] Anderson / Finn (1996), 438

[366] Peat et al. (2009), Cp. Anderson / Finn (1996), 431 ff.

[367] Anderson / Finn (1996), 432 f.

[368] Cp. Anderson / Finn (1996), 435

brainstorming, brainwriting, or other such methods, a day or two before the actual session.

We now want to test if there is significant proof for this assumption using a paired measurement. We group the students into two comparably strong groups, ‘Team I’ and ‘Team 0’. A sample of 60 design students from the same class is chosen randomly. They are put in order according to their score in the design exam they had taken shortly before the test and grouped in pairs. The first of each pair is assigned a random. If the number is even he or she goes into Team 0, if the number is odd, he or she goes into Team 1. The second of each pair goes into the opposite Team. This way, we avoid selection-bias and at the same time make sure, we match comparably strong individuals into pairs.

The students from Team 1 are given the design task 24 hours before the experiment, while the students from Team 0 are only confronted with the example problem right before the experiment. Each student is given one hour to generate and scetch as many solutions to the problem as he or she can generate. We count the number of ideas each student hands in and compare whether students from Team I (after 24 hours of incubation) tend to turn in more ideas than students from Team 0 (no incubation).

The null hypothesis is that the incubation has no influence on the number of ideas generated. If this was the case, the mean of the differences must be zero. H_0: μ_d=0

The alternative hypothesis is that incubation has a positive influence, i.e. H_1: μ_d>0
From the example data in Table 54 we obtain all necessary values to calculate the test statistic.
The mean difference of the pairs is: $\bar{d}$=1,8; the standard deviation of the differences of all pairs: s$\bar{d}$ = 4,24; the number of pairs is n= 30; the test statistic is:

$$t = \frac{\bar{\mathrm{d}}}{s_{\bar{\mathrm{d}}}}\sqrt{n} = \frac{1{,}8}{4{,}24}\sqrt{30} = 2{,}33$$

With a sample size of 30, we have to refer to Students t distribution with 29 degrees of freedom. For a one-tailed test, the null hypothesis is rejected at the 2,5% level, if t>2,045. With t=2,33, this is the case. However, at the 1% significance level, we could not reject H0, as t would have to be greater than 2,462, which is not the case.

So, interpreting the data, we have good reason to believe that a 24-hour incubation period prior to creativity sessions does indeed have a positive influence on the number of ideas generated by the participants.

Table 54 - Hypthetical example data

Team 1 with incubation			**Team 0 no incubation**			
Name	**Exam result**	**Number of Ideas**	**Name**	**Exam result**	**Number of Ideas**	**Difference**
Prince	100	13	Tamara	100	11	2
Peggy	99	8	Noemi	99	10	-2
Buck	97	11	Dirk	99	11	0
Joey	97	14	Susann	97	6	8
Javier	96	16	Brandon	97	9	7
Barbie	96	16	Aldo	95	11	5
Erinn	95	9	Ralph	94	13	-4
Lashell	91	12	Micaela	93	14	-2
Maryln	91	17	Monnie	90	11	6
Donnette	87	11	Denis	87	6	5
Rosanna	86	11	Bernardo	86	8	3
Columbus	85	15	Les	84	9	6
Malissa	83	10	Clifton	82	6	4
Andrea	80	9	Emelina	81	16	-7
Jesus	80	15	Erasmo	80	6	9
Hisako	79	17	Desire	79	8	9
Sharri	79	13	Zack	79	14	-1
Lisabeth	78	16	Lincoln	76	15	1
Donald	74	8	Julio	75	12	-4
Louvenia	74	11	Normand	73	13	-2
Tobias	72	9	Cathey	72	10	-1
Moises	71	14	Horacio	71	9	5
Kay	70	14	Van	70	10	4
Jamison	69	14	India	69	11	3
Alberto	69	7	Mirian	69	7	0
Mariko	67	14	Earleen	67	13	1
Andy	67	8	Lina	67	12	-4
Shante	63	15	Fern	65	10	5
Christiana	62	12	Ernest	62	14	-2
Gerda	60	14	Omer	60	14	0
				Mean difference		**1,80**
				Standard Deviation		**4,24**

Signed Rank Test for matched Samples[369]

If two samples are matched, i.e. if the same individuals are observed at two different occasions, or if pairs of comparable individuals are being observed, a *p value* can be calculated. This is possible, even with data on an *interval scale,* which does not have to be normal. In such occasions, a *signed-rank test* is used.

It is based on comparing the location of two samples by looking at the *median of the differences*. This is typical for scenarios where individuals are tested before and after a medical or psychological treatment or two different teaching methods are applied to comparably talented students. In design research, the typical scenario would be to compare design outcomes of two groups composed of comparably talented designers. In one of the groups, the designers are equipped with a certain design support.

---

[369] Cp. Anderson / Finn (1996), 481 ff. or Lee et al. (2000), 762 ff.

Step 1 – Hypothesis formulation

Null Hypothesis	*Both samples are in the same location, the mean of rank-differences is zero.*	H_0: $m=0$
Alternative Hypothesis (directional)	*The effect in the first group is larger/smaller than in the second group.*	H_1: $m < 0$ H_1: $m > 0$
Alternative Hypothesis (non-directional)	*The effect in the first group differs from that in the second group.*	H_1: $m \neq 0$

Step 2 – Calculation of the differences

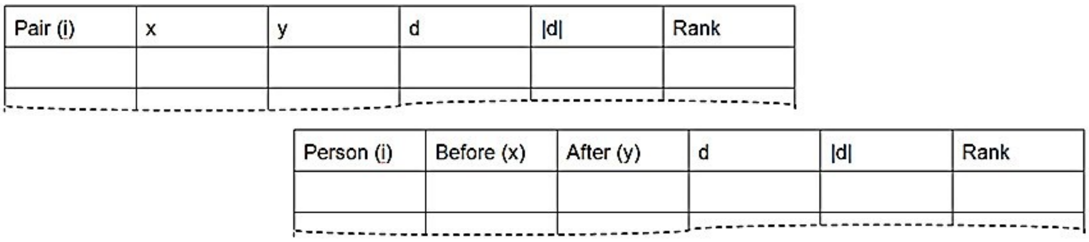
For each pair, the difference, including **algebraic sign,** is calculated,

$$d_i = x_i - y_i$$

d_i: difference
x_i: Observed values from first group
y_i: observed values from second group

Step 3 – Order by rank

Next, the pairs are ordered by absolute value of their differences |*d*| from smallest to largest. Each pair is assigned a rank starting from 1 for the smallest difference. In cases of a tie (two or more identical absolute differences), the average rank of the tied differences is assigned (e.g. if two differences occupied ranks 7 and 8, they are both assigned rank 7,5). Figure 22 shows two suggestions on how to arrange the data purposefully for paired measurements, depending on whether matched pairs or individuals before and after an event are being observed.



Pair (i)	x	y	d	|d|	Rank

Person (i)	Before (x)	After (y)	d	|d|	Rank

Figure 22 - Suggestions how to organize data for signed rank sum tests

Step 4 – Building the rank-sum

Add all ranks that have a positive difference (d>0) to determine the rank-sum *RS*.[370]

$$RS = \sum Rank_j(d > 0) \tag{28}$$

Step 5 – Determining *p*

RS is compared with all possible rank-sums for permutations of differences leading to the same *absolute differences*.[371] Table 55 shows the permutations for an example of

---

[370] Cp. Lee et al. (2000), 763

[371] Look at the example or any of the examples found online to easier understand what is meant by that.

three absolute differences. The number of permutations of +/- in a set of n differences is 2^n.[372]

In Table 55, the absolute differences are *a*, *b* and *c*. $2^3=8$ permutations are possible. Hypothetically, in a setup with 15 pairs of designers, $2^{15}=32768$ permutations must be handled, which makes it impossible to do it with pen and paper. There are MS Excel-Tools available, MATLAB routines and SPPS code to conduct the test as well as other tests that return a similar result (e.g. Wilcoxon Rank-Sum-Test, p. 136 f.).

The percentage of rank sums of the theoretically possible permutations that are equal or higher than the obtained rank-sum from Step 4, is *p*, which is interpreted as in all other hypothesis tests presented here. A low value for p (e.g. <5%) indicates that the results are significant and H_0 cannot be accepted.

If *RS=5*, two out of eight permutations would be equal or larger in their rank-sum, so *p would be 25%,* in an example with three pairs. It is impossible to get a meaningful result with three pairs as the lowest possible value for p would still be 12.5%. This would always lead to accepting H_0. At least 4 pairs are necessary to purposefully apply the rank-sum-test!

Table 55 - Permutations of differences leading to the same *absolute differences for three pairs*

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Rank**
a	-a	a	a	-a	-a	a	-a	1
b	b	-b	b	-b	b	-b	-b	2
c	c	c	-c	c	-c	-c	-c	3
RS_1= 1+2+3 =6	RS_2= 2+3 =5	RS_3= 1+3 =4	RS_4= 1+2 =3	RS_5=3	RS_6=2	RS_7=1	RS_8=0	

Wilcoxon Signed-Rank-Sum-Test for matched Pairs[373]

An improved version of the signed-rank test is *Wilcoxon's signed-rank test.* It also works with rank sums, however, it does not use permutations, which makes it a lot easier to use when the sample size is something larger than 5 (which is actually still a terribly small sample size for any statistician).

As in the signed-rank test above, the data should be arranged similar to Figure 22. Differences and absolute differences are calculated and arranged by rank of the absolute differences (Step1 through 3). However, if for a pair the difference is 0, the

---

[372] It is basically a coin toss. N attempts lead to a possible distribution of heads and tails of 2^n.

[373] Wilcoxon (1945), Lee et al. (2000), 762 ff.

pair is excluded from the set. This leads to a reduction of applicable sample size.[374] Ties between pairs are treated as in a regular signed-rank test, their mean rank is assigned to all of them.

Step 4' – Calculation of a test statistic W[375]

The values of the signed ranks that belong to positive differences are summed up. The same is done for the ranks of negative differences. The smaller of the two results is chosen as the test statistic *W*.[376]

$$W_+ = \sum R_i \,(d_i > 0)$$
$$W_- = \sum R_i \,(d_i < 0)$$
$$W = \min(W_+, W_-)$$

d_i: Difference of Pair i
R_i: Rank of pair i

(29)

W can be compared to a critical value in the corresponding table (Appendix: 9.3.3)

H_0 is rejected if $W<W_{critical}$. For large sample sizes, *W* approximates a normal distribution. According to Sachs, for $n>25$, a z-Value can be calculated and used as in other hypothesis tests with a standard normal distribution table for the critical z-values (9.3.1).[377]

$$z = \frac{|W - \frac{n(n+1)}{4}|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

W: $W=\min(W_+, W_-)$
n: sample sizes (after elimination of pairs with x-y=0)

(30)

Example:

As an example, the same scenario as for the matched t-test is used. The hypothetical data is taken from Table 54. Step 1 and step 2 have already been done in Table 54. Table 56 shows the pairs ordered by rank, i.e. from smallest absolute difference to largest (step 3). Note that compared to Table 54, the pairs with d=0 have been excluded. The sample size is reduced to 27. Next, The Rank-Sums are calculated as described in 4'.

W_+=279

W_-=98,5

Therefore, **W=98,5**.

Next, the table from Appendix 9.3.3 is used to retrieve the critical value for W which may not be exceeded in order to reject H_0.

---

[374] If a large number of pairs compared to the sample size is being excluded, it is an indicator to accept the null hypothesis.

[375] Note that some textbooks use T instead. The test is then often called Wilcoxon T Test.

[376] Lee et al. (2000), 763

[377] Sachs (2004)

For 25 pairs[378] and a significance level of *0,05* for a one tailed test, the critical value is 100. Since 98,5 is smaller, we are allowed to reject H_0 and conclude that there is a positive influence.

Table 56 - Example data ordered by absolute difference (data taken from Table 54)

Person 1_i	x	Person 2_i	y	d	|d|	Rank	Ranks of pos. D	Ranks of neg. d
Sharri	13	Zack	14	-1	1	2,5	0	2,5
Lisabeth	16	Lincoln	15	1	1	2,5	2,5	0
Tobias	9	Cathey	10	-1	1	2,5	0	2,5
Mariko	14	Earleen	13	1	1	2,5	2,5	0
Prince	13	Tamara	11	2	2	7	7	0
Peggy	8	Noemi	10	-2	2	7	0	7
Lashell	12	Micaela	14	-2	2	7	0	7
Louvenia	11	Normand	13	-2	2	7	0	7
Christiana	12	Ernest	14	-2	2	7	0	7
Rosanna	11	Bernardo	8	3	3	10,5	10,5	0
Jamison	14	India	11	3	3	10,5	10,5	0
Erinn	9	Ralph	13	-4	4	14	0	14
Malissa	10	Clifton	6	4	4	14	14	0
Donald	8	Julio	12	-4	4	14	0	14
Kay	14	Van	10	4	4	14	14	0
Andy	8	Lina	12	-4	4	14	0	14
Barbie	16	Aldo	11	5	5	18,5	18,5	0
Donnette	11	Denis	6	5	5	18,5	18,5	0
Moises	14	Horacio	9	5	5	18,5	18,5	0
Shante	15	Fern	10	5	5	18,5	18,5	0
Maryln	17	Monnie	11	6	6	21	21	0
Columbus	15	Les	9	6	6	21,5	21,5	0
Javier	16	Brandon	9	7	7	23,5	23,5	0
Andrea	9	Emelina	16	-7	7	23,5	0	23,5
Joey	14	Susann	6	8	8	25	25	0
Jesus	15	Erasmo	6	9	9	26,5	26,5	0
Hisako	17	Desire	8	9	9	26,5	26,5	0
						Sum	279	98,5

4.3.2.8 Statistical Errors in Hypothesis Testing

When we do a hypothesis test based on a sample from which we try to make conclusions about a whole population, e.g. observing a group of designers from one company and then assuming that the observed is true for all designers from that

---

[378] The tabularized data for the test is only available for 25 pairs. Since we actually have 27 pairs in the example, the interpretation of the data based on 25 pairs is very concervative.

company, there is always a possibility of making an error. It is possible that by coincidence, the observed sample does not represent the population (e.g. the sample group is unusually fast solving a problem). There are two types of errors as shown in Table 57. A *type I error* occurs, when we believe to detect an influence, although there is no influence. In statistical terms: We reject the H_0 although it is true. A *type II error* occurs when we do not detect an influence although there really is an influence. We accept H_0 although it is false.

Table 57 - Possible errors in drawing statistical conclusions (Anderson / Finn (1996), 406)

		Conclusion	
		Accept null hypothesis	**Reject null hypothesis**
True 'State of nature'	**Null hypothesis is true**	correct	Type I error
	Alternative hypothesis is true	Type II error	Correct

In the example in 4.3.2.4, there is a chance that the 40 designers we observed without a design support are unusually slow and do not represent designers in general. If that is the case, the actual mean for the time necessary to solve the example problem would be shorter. The delta between the means (in our example about 112 s) for designers with and without the method is smaller and there might actually be no influence of the design support on the time needed to solve a problem. In that case, we would have rejected H_0 although it were true – a type I error would have occurred. Generally, it can be said that large sample size reduces the risk of such errors. Systematic errors can be reduced by randomization thus avoiding unconsciously selecting a sample group that does not represent the population distribution.

The ideal case would be to have access to a very large number of "representatives" of the population that we want to conclude about. We could randomly choose a still large sample group from those representatives. When conducting research in design science, this ideal can hardly be achieved. The number of available companies or their designers is limited. Large sample sizes are seldom available. It is likely that companies will choose which engineers/designers they provide for interviews or experiments, so there is an increased chance of systematic error we have to deal with.

4.3.3 Validity

LINN and GRONLUND posed five important cautions when using the term *validity* in relation to testing and assessment.[379]

---

[379] Linn / Gronlund (1999), p. 49, cited in Treffinger et al. (2002), p.31

- "Validity refers to the appropriateness of the interpretation of the results of an assessment procedure for a given group of individuals, not to the procedure itself
- Validity is a matter of degree; it does not exist on an all-or-none basis. . . .
- Validity is always specific to some particular use or interpretation. . . .
- Validity is a unitary concept [based on various kinds of evidence].
- Validity involves an overall evaluative judgment. It requires an evaluation of the degree to which interpretations and uses of assessment results are justified by supporting evidence and in terms of the consequences of those interpretations and uses."

For further reading on reliability, its subcategories and how it can be assessed, refer to CARMINES AND ZELLER.[380]

BINZ AND KELLER have collected criteria for credibility and validity of design support that are summarized as questions for self-check in Worksheet 2. More questions for self-check are available in Worksheet 10 (sections E,D, and F).

TREFFINGER ET AL. point out some considerations on the terms ' and ' in the context of instruments that are supposed to assess, measure or support creativity. While those considerations are being made by scientists who dedicated their work to the field of phycology and the education of gifted students, it seems well transferable to the field of design research. Here too, we can observe designers with different extends of talent. We find many methods and instruments that are supposed to support creativity and encourage designers' talents for designing. Hence, the considerations are reproduced here:

"Although we often say, almost glibly that any instruments we use in identification must be 'valid and reliable', we need to use those terms with considerable caution. The terms '*validity*' and '*reliability*' represent important principles in testing and measurement, but they are not as absolute and fixed as some people seem to assume. In addition, in any domain of giftedness or talent, there will be many variations of productivity and accomplishment over time." [381]

They summarize that regarding instruments used to measure an effect or achieve an intended result, the question for validity and reliability cannot be answered as "yes it is" or "no it is not" valid/reliable. "Determining validity and reliability are on-going processes." That process must take into account for which subjects reliability and validity are to be checked, and under which conditions. The key question thus is not 'Is this valid?' but much rather "Given the evidence available, for what, in what respects, for whom, and under what conditions are my findings valid and reliable?"[382]

---

[380] Carmines / Zeller (1979)

[381] Treffinger et al. (2002), who cited from Treffinger et al. (2001), pp.3 f.

[382] Loc. cit.

TREFFINGER ET AL. also discuss the term ‘evidence’. They point out that what a researcher calls evidence can be based on assumptions. Statistical operations are often based on the assumption that what is being measured is a “stable trait” in a population. Complex human behavior – such as designing – may include effects that are not stable but depend on experience. Assumptions such as normal distribution or at least a symmetrical distribution within a population should be made carefully. At the same time, most procedures for hypothesis testing are based on such assumptions. Finally, TREFFINGER ET AL. warn about false generalization. In their research on gifted children, a finding and the applied measurement methods might be valid and reliable for a certain age but might not be for other ages. In design science, similar caution is advised when measuring effects with students and assuming them to be transferable to designers in companies. More subtle but just as dangerous: Experimental results with designers from small and medium-size companies might be highly valid and reliable. However, one must be very careful to assume that the observed effects are the same for larger companies that might pose a completely different environment for the designers. These different aspects pointed out by TREFFINGER ET AL. have been described as different categories of *validity* and *reliability*.

4.3.3.1 Content Validity[383]

Content validity (also called “logical validity”) refers to how well the design of a measurement set up by the researcher includes all aspects and influencing factors of the intended observation. In simple words, ask yourself before an experiment: “Will I really observe everything that belongs to the problem?” The following two examples, one for good and one for low content validity will make this clearer. Finding an example of good content validity is much harder. Any construct will have factors that limit the content validity. Reasons for this can be found in STACHOWIAK’S model theory.[384] Any experimental setup is a model of a limited part of reality that is to be observed. Models always reduce reality to a relevant selection. In other words, content validity describes how well a researcher is capable of including the *relevant* part in an experimental setup or an observation.

An example of low content validity:

A researcher wants to find out if a certain process model has a positive influence on the success of a design team. He observers how well the different teams implement the process-model and assess the outputs of the team. He discovers that the teams which ignored the model did worse than the teams that did, in fact, use it, and he reasons that the model makes design teams more successful. What the researcher did not track in this example was how well the individuals in each team cooperated. In

---

[383] Kridel (2010), 924

[384] Stachowiak (1973b) describes three features of models: ‘Mapping Feature’, Reduction Feature’, and ‘Pragmatic Feature’. More on Stachowiak’s model theory in the context of design science can be found at: Albers / Meboldt (2007a),Meboldt (2008), Oerding (2009)

some of the teams that did not implement the model there were rivalries and animosities while the other teams got along well and enjoyed the chance to try out the new model. They stayed longer hours but did not find their work stressful.

An example of good content validity:

Student design teams are given the task to propose different conceptual solutions for a problem. One of the proposals shall be selected to be developed further. The goal is to have the students work and perform under “real-live live conditions” although they are at the university. In order to increase content validity, an industrial partner is acquired that gives out the design task - a real problem that really needs to be solved for the company. For milestone-presentations, high-level representatives of the company are present and decide which of the concepts are to be further developed. On top of those measures, the company has to reimburse the university for the involved scientists, and in exchange gets to keep all engineering results developed during the project. The company, therefore, is put in a customer-like role increasing the pressure on the students, who are fully aware of the constellation, to a more realistic level.

4.3.3.2 Face Validity[385]

Similar to 'content validity', 'face validity' refers to the degree to which a measurement really catches, what it is supposed to measure. Face validity is used, when content validity cannot be shown. It is based on the subjective rating of experts, so it is somewhat less valuable.[386] In other words, if you have no way of showing that you are establishing high content validity, at least get someone else’s opinion who is familiar with the subject. If that person / those people say that the setup seems good, then you can claim that you have achieved face validity. The more experts’ opinions you get and the more established these experts are in the relevant field, the more reliable your experiment’s face validity.

Example:

A researcher wants to find out if stimuli coming from the interior design of a room have an influence on creative performance. A group of students that all know each other are invited to participate in the test. The students are put in two teams. The researcher has no means to test the participant’s creative intelligence, so he asks the students if they deemed the two teams equally creative. After exchanging some of the participants between the team, all students agree that the teams are “fair” in respect to creative skills. While this way of reducing bias is not based on a systematic method, it still beats ignoring the potential bias that would result from unequal teams. Also, scientists working in the field of cognitive psychology like e.g. GIGERENZER have revealed interesting findings about the astonishingly high quality of “gut feelings." After all, face validity is better than nothing.

---

[385] Kridel (2010), 924

[386] Mind that expert-ratings in many studies return results comparable with those resulting from other metrics. Expert ratings are not as bad as it sounds.

4.3.3.3 Criterion Validity[387]

Often, researchers cannot directly measure what is at the core of one's interest. The data might not be accessible due to different reasons, e.g. ethics in the social sciences, physical limitations of what can be measured in the natural sciences or a large time shift between cause and effect very typical for management science). The only workaround in those cases is to measure indirect variables believed to be connected to the criterion of interest.

In design science, many product development processes have been suggested. They ultimately aim at better market success of the companies applying the processes.[388] This, however, can hardly be measured. First of all, product development and market success occur at very distant points in time.[389] Secondly, a criterion like market success is subject to a complex network of factors. Backtracking success to the implementation of a certain process model, excluding all other, external factors is nearly impossible. However, factors could be observed, believed to have a positive influence on market success. E.g. the difference between planned and actual development time can be observed. It is believed that good project management has a positive influence on market success and that a good correlation between planning and execution (plan-actual-delta) is a measure of good project management. The more evidence is established within the scientific community that the plan-actual-delta is a good measure for project management quality, the higher the criterion quality for plan-actual-delta for measuring project management quality becomes. Accordingly, the more evidence from previous studies is available that project management quality is a suitable variable to measure market success of a company, the higher the criterion validity.

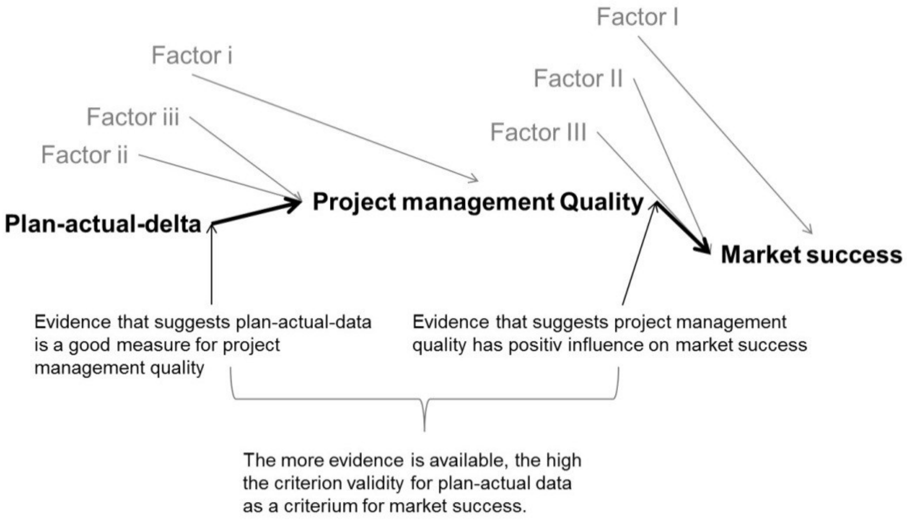


Figure 23 – Chain of evidence to establish criterion validity (own illustration)

---

[387] Kridel (2010), 923

[388] E.g. Cooper (1999) deals with such question. Ernst (2002) gives a comprehensive overview

[389] In the automotive industry, development projects alone take several years, and ultimate market success can only be judged after considerable time in market.

Criterion validity can be further distinguished as concurrent and predictive validity.

Concurrent Validity:[390]

Concurrent validity describes the degree to which a tested variable is related to another, established measurement that could be alternatively applied. In the example above, this would refer to the plan-actual-delta as a measure for project management quality in contrast to an alternative way to asses project management quality, e.g. expert assessment through an external consulting team. Both could be done in parallel and the results be compared. If they match to a large extend, we assign the two tests a good concurrent validity.

Predictive Validity:[391]

When we are expecting a future performance based on data we are acquiring now, predictive validity refers to the degree to which our current measurement allows us to make predictions of the future. In the example, the second part of the criterion validity can be discussed under the aspect of predictive validity. Project management quality is measured to make predictions of market success. If past studies have shown companies that have good project management also have superior market success, there is evidence suggesting high predictive validity.[392] The better such studies can show that there is a direct influence, and the more companies the studies are based on, the stronger the case for a good predictive validity.

Predictive validity is difficult to establish and an easy target for critics of new methods or studies. It takes time to establish predictive validity since several independent studies are necessary to build it. At the same time, scientists are reluctant to use methods that have not yet been proven to have a high predictive validity, making it harder to improve it.

4.3.3.4 Construct Validity[393]

Construct validity refers to how well a test measures what it is supposed to – the intended hypothetical construct. Intelligence tests, for example, are constantly criticized for low construct validity. What makes people so sure that a low score in an IQ-tests really means the person is not intelligent? The more evidence is available, the higher the construct validity. It is therefore, a popular topic in psychology to argue about. A common way to establish construct validity is to conduct experiments and calculate a correlation coefficient. E.g. some scientists believe that general intelligence and creative intelligence are closely linked (compare 4.2.2.5). If we took a large sample of designers that have undergone an IQ-test, and later asses their creativity, we can compare those two results and calculate the correlation coefficient.

---

[390] Kridel (2010), 923

[391] Ibid.

[392] The example is of course oversimplified. In order to really establish predictive validity we would also have to show that there is significantly less companies with good market success although they have evidently bad project management.

[393] Kridel (2010), 924

On the question: “What is a good validity coefficient?” CRONBACH stated: “The best you can get." When measuring the validity as a correlation coefficient, it is uncommon to rise above 0.6.[394]

In sociology and the related statistics, 0.6 is usually an acceptable correlation coefficient. In technical systems in engineering, the same mathematical construct is used to compare simulation models with experimental results. Here, 0.8 and higher are common minimum requirements. MAIER achieved correlation coefficients between 0.9 and 1.0

4.3.4 Reliability

“The extent to which [measurements] are repeatable” is called reliability. “Any random influence which tends to make measurements different from occasion to occasion is a source of measurement error.”[395] In other words, reliability describes the consistency of several measurements under constant conditions.

Reliability is, like validity, nothing a test or study has or does not have. It is built up, and it is a researcher’s obligation to build the case, presenting the factors that increase and decrease the reliability of the presented work. While reliability alone does not ensure validity (one can consistently measure the wrong effect with similar results), limited reliability does result in limited validity. Reliability is established in different subcategories.[396] Some questions for self-check are given in Worksheet 10, section C.

4.3.4.1 Test-retest Reliability[397]

Test-retest reliability – sometimes also referred to as repeatability – describes to which extend a measurement can be repeated *by the same observer*, *of one and the same item*, *under the same conditions*.

In design research this is a challenge since a large part of design science is inseparably connected to some type of problem solving. A problem, however, cannot be solved twice under the same conditions. This has been argued in chapter 2.2.3.

4.3.4.2 Parallel Test Reliability[398]

This subcategory of reliability is known under several synonyms: ‘Equivalent-Forms Reliability’, ‘Equivalent-Test Reliability’, ‘Alternate-Forms Reliability’, ‘Alternate Test Reliability’, and ‘Parallel Forms Reliability’. It describes the degree to which two independent tests on the same subject lead to the same results, when they are conducted *by the same observer, with the same participants, under identical conditions, with two different items with the same intent*.

---

[394] Maier (2011), 121

[395] Nunnally et al. (1967), 206

[396] For further reading on reliability, its subcategories and how it can be assessed, refer to Carmines / Zeller (1979)

[397] Kridel (2010), 739

[398] Ibid.

Parallel test reliability is important whenever there is a risk that test takers / participants might remember some of the responses / solutions they made during a previous session. An alternative form or test must therefore be developed. A big problem is that it is extremely difficulty constructing two forms that are essentially equivalent.[399] Proof of how difficult good parallel test reliability is achieved can be gathered after any university exam. Many students will claim that the last years' exams were much easier. They are claiming to have encountered a low degree of parallel test reliability.[400] However, the professors and the assistant researchers go through a lot of effort trying to make the exams comparably difficult every year. They compare test score statistics afterwards to check whether they have achieved a good level of parallel test reliability and so on. In other words, year after year, they are establishing their case for high parallel test reliability.

For design science, BENDER'S 'task design' and SCHRODA'S 'task evaluation' have been presented in chapter 4.2.1. These methods are designed to help increase parallel test reliability for design tasks. Expert Consensual Assessment (4.2.2.5) can also be used to decide if two (or more) design tasks are comparably difficult, i.e. if they show a high level of parallel test reliability.

4.3.4.3 Inter-rater reliability[401]

Inter-rater reliability is also known as inter-rater consistency, inter-assessor reliability, inter-rater agreement, or concordance. It addresses how consistent a test or an observation is made: *by different observers, of one and the same item / situation, under the same conditions*.

Whenever data collection is divided between several researchers, inter-rater reliability has to be checked critically. In interview studies with many participants or coding of data such as video or audio recordings this is quite common. In such cases, the observers are given a set of rules on how to judge. Ideally, the results should be identical and completely independent from who makes the judgment. If low inter-rater reliability is found, it means that the test, including all support given to the raters, should be revised to increase its inter-rater reliability. Questions might be rephrased to reduce ambiguity, scales for quantifying personal judgment might be revised or extra explanations might be added. Inter-rater reliability tests are also useful to check the quality of a questionnaire to see if interpretations of the questions leave too much room for ambiguity.

---

[399] Proof for this statement can be gathered after any university exam. Many students will claim that the last years' exams were much easier.

[400] Note that they can actually not make this claim since those few that did indeed take both exams under comparable conditions must have failed the first one and might be better prepared the second time. The rest has practiced last year's exam at home, usually with a best practice solution available, so identical conditions do not apply here at all.

[401] Cp. Johnson et al. (2008), 170 ff.

4.3.4.4 Intra-rater reliability[402]

In contrast to inter-rater reliability, intra-rater reliability (also intra-rater consistency) describes how consistent results of a test or an observation turn out when it is repeated *by one single observer, of one and the same item / situation, under the same conditions, at different times.*

It is important, when longitudinal studies are conducted, where a researcher judges similar situations at different points in time. Ambiguity of the judgments has to be minimal so that the different observations can truly be connected to the observed situation. Example: In an ethnographic study, the researcher takes notes, every Friday on how well the members of a design team cooperated, and how much they got accomplished in the past week. He uses a checklist for his Friday-task. Before the study, he rated a hypothetical situation several times to check the intra-rater reliability of the checklist.

4.3.5 Validation and Verification

Closely related to the concepts of validity and reliability are the terms “validation” and “verification." A whole philosophical branch of science – epistemology – has been and is still debating on how humankind creates knowledge, which knowledge can and cannot be taken as “truth” or “true knowledge.” Since this thesis is not directed at that philosophical part of science, it shall not be dealt with in detail here. The next section will present a selection of research paradigms that can be considered as the result of that philosophical debate. Some of the more current philosophers are KANT, POPPER, KUHN and FEYERABEND among other authors.

Mainly in the fields of engineering and the natural sciences, there are various definitions of the two terms validation and verification that are sometimes mixed or even used as synonyms, although a strict differentiation can be made. Within this thesis, the terms shall be differentiated in accordance with the German VDI guideline that describes verification as the formal evaluation between a system's properties and its specifications.[403] Therefore, verification is the process that assesses whether a system has been built and engineered correctly. Validation, on the other hand, is the process that assesses whether the right system has been built. In other words: „Does it meet the customers’ expectations?“ ALBERS ET AL. point out that this can only be done if the requirements are made explicit in the system of objectives (see also 5.1.1). Only then can a comparison between the system of objectives and the system of objects be conducted.[404]

---

[402] Ibid.

[403] Verein Deutscher Ingenieure (2004)

[404] Albers et al. (2009)

4.3.6 Summing up

”Empirical research must fulfill some minimum requirements in order to produce valid results. In sociology, this is a commonly accepted rule and well established in practice. In design science, however, these criteria are often disregarded.”[405]

This whole thesis is an attempt to improve that situation and make those well-established rules easily available for design scientists. Readers should not mistake this for a comprehensive overview over statistics and sociology. It is not a textbook much less a library. The author has chosen certain methods trying to provide the tools that address common situations in design research. There will always be research situations that none of the methods presented here can address. However, with a basic understanding from the explanations here, finding suitable methods and terminology should be easier.

The goal was to provide a solid database as part of the framework, design scientists can choose from, according to their needs and situation. Before the actual framework is presented, an extra chapter was included dealing with the different research paradigms. Research paradigms are the reason that some of the research methods come into fashion and go out of fashion over time. Some methods are the subject of ongoing arguments.

4.4 Research Paradigms

Looking at the historical past of science one might make a surprising discovery. What is considered good scientific practice, changes over time and differs throughout the scientific communities. A thorough study today must not necessarily be appreciated tomorrow. There have been times, when measurement and observation were found to be only for those “fools” incapable of reason. 3000 years ago, the superior way of conducting science was to use reason instead of sensory inquiry ago. But even without going back to the old Greeks, observing the current research in engineering design, it seems that different opinions exist. At the 2012 Design Conference e.g., a podium discussion was held on the topic: *”Design Research should be about developing new products, technologies and services, not theories, models and methods.”*[406]

The following paragraphs will take the reader through the evolution of modern epistemology by presenting a series of different paradigms.[407] As this is an issue of philosophy, there have been countless attempts to describe different notions and attitudes of science. It is not the author's intention to judge, which of them are correct, neither is it possible to list all of them.[408] The goal here, is to make the point that the

---

[405] Bender (2004)

[406] There is a video available at: http://mod.carnet.hr/index.php?q=watch&id=1843; 25/2/2013, 15:12

[407] The classification of research paradigm is subject to paradigm-shifts as well. A full list can never be given. For further reading on the topic, see also: Guba (1990), Habermas (1972), Reason / Rowan (1981), Rowan (1981), Kuhn (1996)

[408] For further reading on epistemology, the following sources might be of value: Argyris (1980), Bunge (1983), Kuhn (1977), Kuhn (1996), Weimer (1979)

definition(s) of what is scientific and what is not, are somehow temporary.[409] At the end of the day, the question whether some study is scientific or not can only be answered under the consideration of its authors' research paradigm. The correct question would then be: *Which scientific school/worldview does the study belong to and did it fulfill that school's rules?* The following paragraphs will present some of those schools and their typical characteristics. Before writing your next paper, ask yourself: Which school do I regard myself as part of?

4.4.1 Rationalism

> *"Rationalism is the theory that says that reason in itself is the source of all knowledge, superior and independent of sense perception"*
>
> Albert Einstein (1879-1955)

Some of the most important roots of science evolved from the ancient Greeks. At the same time, they believed in many myths, gods, and monsters. They were dominated in their thoughts believing that some invisible power determined their actions and future, never thinking to put those beliefs to a challenge. Naturally, their scientific thinking was dominated in a similar way. In the time between 500 BC until the 16th century, it was believed that knowledge comes from man's mind. "Knowledge, all knowledge, about gods and dragons and how the world works come from your mind, your reasoning and only your reasoning."[410] A scientist who believes in this kind of philosophical view is called a rationalist.[411] (S)he does not build hypothesis and theory to put to examination but uses logical reasoning to prove his/her point. The underlying philosophical view of knowledge is a foundationalist view like that of Aristotle "According to this view knowledge of the world rests on a foundation of indubitable beliefs from which further propositions can be inferred to produce a superstructure of known truths […]."[412]

LEVENSPIEL gives an example of a typical rationalist's point of view:[413] In order to figure out, how many teeth a lion has, a rationalist would not try to catch a lion in order to simply count. He relies on reason, hence would argue that there are 28 teeth. There are four parts of the lion's mouth (upper right and left, lower right and left). So it must be a number that can be divided by four. Seven is a magical number and *4x7=28*.

A famous rationalist was Aristotle, who developed a whole set of mechanical laws without a single experiment. It is also the rationalist view of the world that is responsible

---

[409] This statement just revealed me as a believer of the Relativistic / Holistic / Social School of Epistemology (4.4.7)

[410] Levenspiel (2007), 1.2

[411] It was Descartes, who put a name to this philosophical view in 1641. More than 2000 years after Socrates and his students had started it. Descartes (1931), see also Pedersen et al. (2000)

[412] Honderich (1995), cited in Pedersen et al. (2000), 11

[413] Levenspiel (2007), 1.2

for the many different models of the universe that existed until the 16th century the most famous in the western world being that the earth is a disc with everything else circling around it.[414]

4.4.2 Authoritarianism

During the era of rationalism, in the western world, the church adopted some of the "laws of nature" as they had been reasoned by Aristotle and decided to consider them as laws of god that could not be questioned. For several hundred years, no scientist would openly question those laws or experiment in order to see whether they are right or wrong. Doing so would have been considered heresy and force the church to purify the scientist from those "evil ideas".[415] The few who dared to question the church's believes of those times where e.g. Galileo Galilei and Copernicus, the latter being smart enough to keep his thoughts a secret right until he was about to die, to avoid the church's purification. Galilei was put to trial and found guilty by the inquisition of the Church of Rome until the pope recalled the sentence in 1992 – 350 years later. However, around the time of the famous trial against Galileo, several important inventions were made that together with the braveness of Galilei let to a new paradigm.

4.4.3 Empiricism

Through inventions such as the mechanical clock, the thermometer and optical lenses, scholars became capable of making precise measurements and observations. This started the scientific age with an era of empiricism.[416] Empiricists believe that "*All knowledge about the world comes from measurements of the real world.*"[417] Finally, this new way of thinking led to question e.g. Aristotle's laws of mechanics: "Aristotle claims that 'an iron ball of 100 pounds falling from a height of one hundred cubits reaches the ground before a one-pound ball has fallen a single cubit', I say that they arrive at the same time. You find, on making the experiment that the larger outstrips the smaller by two finger-breadths. Now you would not hide BEHIND these two finger-breadths the ninety-nine cubits of Aristotle, nor would you mention my small error and at the same time pass over in silence his very large one?" [418]

Empiricists were the first scientists to make predictions on how – under a certain condition – some controlled part of the world would behave and then observe whether or not the prediction was correct, so with empiricism, the constructs of "theory" and "experiment" were born, leading to the number-one guideline of empiricists: "If your

---

[414] Levenspiel (2007) gives an overview over the many alternative models of the earth and the universe.

[415] This sometime involved burning the scientist.

[416] In contrast to rationalism, empiricism was given its name right at the time it occurred. John Locke is said to have introduced the term around 1690; see Pedersen et al. (2000)

[417] Levenspiel (2007), 1.7

[418] From a speech by Galilei cited by Levenspiel (2007), 1.7

experimental result does not fit your theory, check your experiment. If it is OK then look at your theory and find its flaw – because theory must agree with your experiment."[419]

4.4.4 Positivism

WITTGENSTEIN introduced the notion of *positivism*.[420] A positivist is a scientist who – similar to a foundationalist – believes that one should infer from the known truths to additional statements, and verify those statements laying the foundation for further inference. Therefore, any statement that cannot be formalized in a way that allows for analytical or empirical investigation is of no value for a positivist, as he cannot rely, hence not build on that statement.

4.4.5 Foundationalism

Foundationalists believe that there is an entity of truths that describe the world. In their view, one can only build on the known truths to proceed in generating new truths. Science is strictly about uncovering those truths and not about interpretation. True objectivity hence exists, and any approach to uncover the objective truth must ensure that it is itself objective. These scientists, consequently, tend to formalize and often share a reductionist view of the world. Naturally, those scientists call for absolute rigor and quantitative validation.[421]

4.4.6 Reductionism

"Methodological reductionists postulate that the properties of the whole are the sum of the properties of the parts. Hence, analysis of the parts is sufficient to gain knowledge about the whole."[422]

This means, in order to verify a complex set of postulates/hypothesis, a reductionist will try to break things down into smaller pieces, verifying each of those pieces individually. The whole set of postulates is true if all pieces could be proven true. The set is false if one part of it can be falsified. This approach allows for very systematic action in science. However, it ignores what in systems theory is called emergence. Effects that occur from the interrelations of the individual subsystems.

Reductionism is closely linked with *foundationalism* and *formalism*. All three schools share the assumption that:

"1) *truths (knowledge) are innate and absolute*,
2) *only rational knowledge is valid*, and
3) *objectivity exists*."[423]

---

[419] Levenspiel (2007), 1.9

[420] Wittgenstein (1921)

[421] Due to their very formal approach towards science, they are also called "Formalists". See Pedersen et al. (2000), 2

[422] Pedersen et al. (2000), 2

[423] Pedersen et al. (2000), 2

As shown in chapter 2.3, design science overlaps in parts with psychology and parts of the social sciences, hence including subjective statements. This is why in design science, especially reductionist and formalist thinking become quite problematic and many design scientists probably will not see themselves as reductionist formalist scientists.[424] At the same time, reductionist- formalist scientist will “frown upon” design scientists, not crediting their research a lot of value. So the question is: Is there a niche in epistemology for those that include subjectivity into their research – as e.g. social scientists and design researchers often do?

4.4.7 The Relativistic / Holistic / Social School of Epistemology

As PEDERSEN ET AL. point out, not everybody agreed with the thought that there is a foundationalist set of innate and absolute truth: KANT, e.g. differentiated between truth that can be experienced and truth that is “added by the mind.”[425] HEGEL went further an disagreed completely with the concept of innate (given) truths. He regarded truth as a process. It doesn’t just exist, but develops. This argument led to HEGEL’S concept of *thesis – antithesis – synthesis*. So in his view, truth was not just added by the mind but what is added is the result of a thought *process* or a *dialogue that includes conflict and contradiction*.[426] “In his view knowledge is socially, culturally, and historically dependent, hence, there are no neutral foundations of knowledge, and entirely objective verification of knowledge claims is not possible.”[427]

THOMAS KUHN supports this attitude. “He argues that in any given epoch scientists work within and against the background of an unquestioned theory or set of beliefs (a paradigm). […]When the ruling paradigm cannot provide adequate explanations to scientific problems under investigation, [then] this inadequacy makes way for new paradigms.”[428]

Other philosophers have added to the relativistic / holistic / social school of epistemology, but its basic concept is that truth is a combination of things given, and things derived from the given by the mind through thorough discussion and according to certain rules. These rules are called a paradigm and the rules change over time as they depend on social and scientific context.

4.4.8 Summing up

A number of paradigms have been presented. It is not trivial – maybe even impossible – to be exact about who belongs to which paradigm. Their definitions overlap. However, there are some streamlines recognizable as is their evolution over time.

---

[424] Mind that in modern times most scientists will believe in foundationalism – most often even the concept of what is scientific is used exchangeable with what is a foundationalist view of the world.

[425] Kant (1933)

[426] Hegel (1959)

[427] Pedersen et al. (2000), 3

[428] Loc. cit., cp. also Kuhn (1996) and Kuhn (1977).

Figure 24 shows a trend from the "old" reductionist/formalist/foundationalist towards what SEEPERSAD ET AL. call the holistic/social/relativist school of epistemology.

Since in "old-school school epistemology," quantitative validation and absolute scientific rigor were of the highest value, the picture might lead to the assumption that in this "new school of thought," quantitative validation is unimportant, and we are going into times of purely qualitative validation. This is not the case! The relativistic holistic/social/relativist includes qualitative research. It does not replace quantitative validation. Neither has scientific rigor lost its importance for science. So for a modern researcher, the world has become actually more complex as (s)he has to decide which way of validation makes sense and therefore, should be applied.

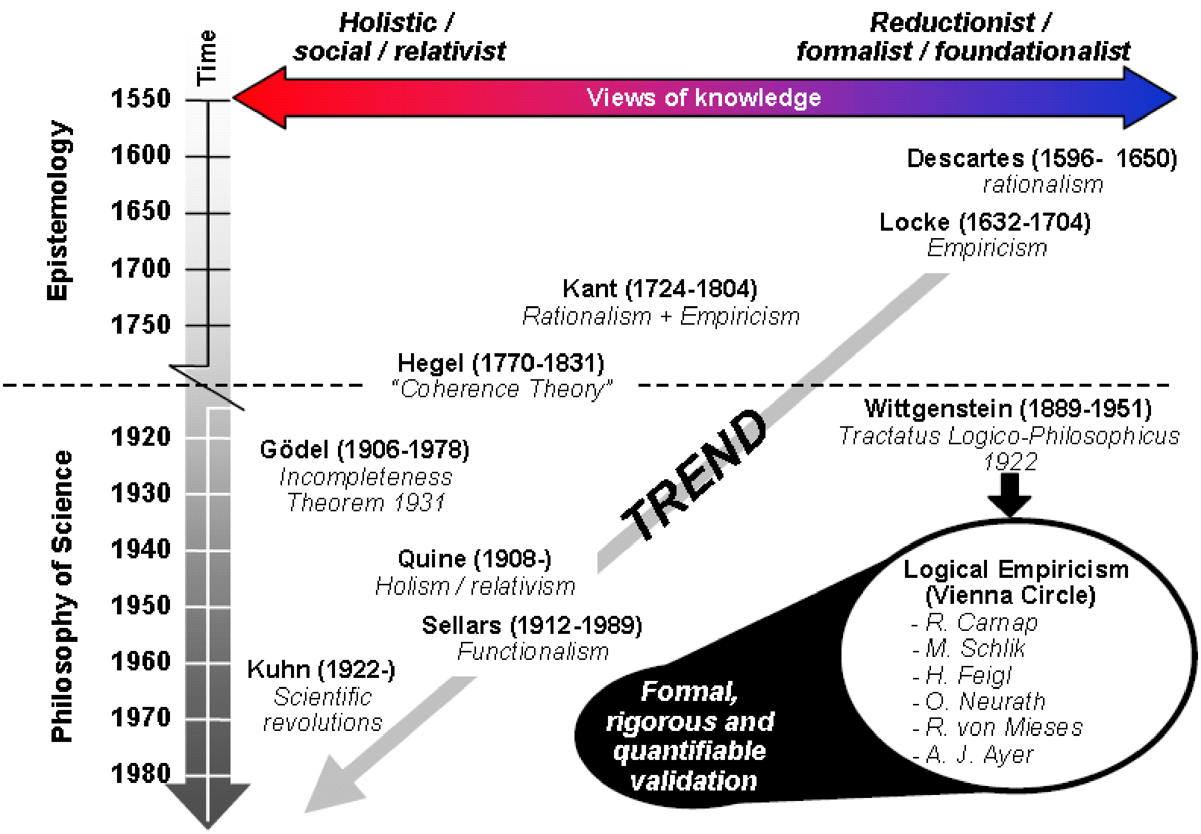


Figure 24 - The evolution of thought, Seepersad et al. (2006), 7

5 Supporting the Development of Heuristic Design Support

The overall goal of this research is to support design support developers in their efforts to create *reliable, credible and valid heuristic design support.* In the previous chapters, it was pointed out that validity and reliability are the result of a process. Valid and reliable outcomes are therefore being determined from the very first step in a development project. Consequently, the work presented here is not about validation in the understanding of a single, isolated activity.[429] It is about a framework that can be applied to structure and describe any design support development project aiming at creating heuristic design support. The goal is to expand the operation system of design science with:

- a preselected set of suitable methods and practices for the development of heuristic design support,
- together with information about the required resources for their application to support project management of design support development projects,
- instructions for their correct application,
- indications to relevant, comprehensive literature, and
- strategies for the selection of a particular set of methods from the pool.

Chapter 4 contains the preselected set of available methods. The methods are provided together with profiles in the form of tables, which contain the information about advantages/disadvantages, necessary resources, and similar methods. Suggestions on further reading and instructions for the application have also been given. This sets the sub-focus of the framework to be developed in this chapter on *integrating the information and providing necessary working sheets and selection strategies*. The guiding question will therefore be: *How can one chose and combine appropriate methods and activities in order to develop design support with relevance, credibility, and a high level of reliability and validity?*

The framework is based on the assumption that, in engineering design research, certain aspects cannot be fully proven. It is impossible and more important, it is not useful either. Much more, developing design support is about building a case for credibility. Hence, the framework aims at convincing:

---

[429] Note that iPeM does include „Validation“ as one of the activities of product engineering. However, this is not a contradiction. The single activity is only to be found in the static part of the model. Looking at the dynamic part of iPeM, e.g. in reference processes, validation is either modeled as a continuous activity running alongside all other activities or it is modeled repeatedly in between other activities. E.g. Braun et al. (2013a)

- *potential users* that a support will help them in certain situations.
- *fellow scientists* that a thorough path has been followed and made transparent for them to reconstruct, interpret, argue about and build upon.

However, following this assumption, great care must be taken not to misuse it as an easy excuse to leave out some uncomfortable steps of validation! Much *more effort* and careful argumentation are necessary to build credibility if an effect *cannot be shown*. This is also the opinion of PEDERSEN ET AL.:

”We define scientific knowledge within the field of engineering design as socially justifiable belief according to the Relativistic School of Epistemology. We do so due to the open nature of design method synthesis, where new knowledge is associated with heuristics and non-precise representations*, thus knowledge validation becomes a process of building confidence in its usefulness with respect to a purpose."*

With PEDERSEN ET AL. and BLESSING ET AL. in mind, the framework should help build a case for validity and reliability alongside a process comparable to product engineering processes. In chapter 4.1 available models, approaches and frameworks with similar targets have been presented. Among them:

- DRM
- The Spiral of applied Research
- Cantamessa’s Model
- Foundation for the Development of Design Methods
- The Validation Square

These are the principal alternatives available which serve – in some way or another – the purpose of building credibility for a design support.[430] Any of these models can be used in the development of design support. If applied correctly, they will contribute to increased credibility of the produced results, compared to an unstructured approach.

Why develop a new Framework?

Since there are some approaches available already, the question comes to mind, why is a new framework suggested instead of applying one of the existing frameworks? The advantages and disadvantages have been presented in chapter 4.1. All the above frameworks have a rather consecutive appearance except for the spiral of applied research, which again is designed to structure large-scale research programs or the efforts of a whole group and is less likely to be helpful for a single doctoral project or similarly limited research projects. It has been shown in engineering design that sequential process models have a disadvantage. They are not very flexible in reacting to changing conditions or targets, and they usually address either the designer’s perspective on a project – he/she is mostly concerned about what has to be done next

---

[430] The other contents presented in chapter 4.1 either deal with more specific and isolated questions within design support development, or they address some thoughts on a more abstract level, such as general advice for the development of design support.

– or the management perspective, dealing with planning and controlling issues, such as “Are we proceeding according to schedule.”

ALBERS ET AL. have developed an alternative approach to modelling engineering design processes: the integrated Product Engineering Model – iPeM. Its main purpose and underlying mental concept is to provide an approach that allows to address both perspectives: The management and the engineering perspective, thus provide a model that is just as useful for planning and controlling an engineering project as it is for navigating through the engineering design process. They key feature of iPeM is that it is an activity-based approach rather than a stage or phase based approach. The question “What needs to be done” is dealt with primarily. A phase model that shows how long this might take and when which activity should be conducted can be derived in a second step. Thus far, several studies have shown that this concept is purposeful and an improvement to engineering design processes.[431] This is the reason why it was decided to use iPeM as the underlying concept for the design support development framework.

Transferring this concept to design support development – adapting the iPeM approach – promises some particular benefits over the existing models:

- If the adaption is successful, the approach will help in structuring and planning research projects representing the management perspective. In other words, it can be used to structure applications for funding and similar activities prior to the actual research.
- The same model will assist the designers' perspective, e.g. a doctoral student who has to decide how to proceed with the research at a certain point within a project.
- In addition, the researcher is also offered concrete methods to choose from, as the methods are directly linked to the activities. So, not only the decision for the next activity is supported, but also conducting that activity.

While all available frameworks do address one or more aspects of the mentioned advantages, none of them addresses all those aspects at once.

Finally, the question remains whether or not other engineering frameworks could be used as the underlying concept instead of iPeM. Prominent models to consider would be the German VDI guideline, the V-model, or typical quality assurance models from the world of software development, such as CMM (Capability Maturity Model) or SCRUM. Again, the argument is that none of the existing process models combine the management and the engineering perspective. That case has been made in the development of iPeM itself.[432] If these models do not have the sought-after advantage within their original purpose, it is assumed that they will not provide the desired assistance in design support development either, unless a lot of extra effort is put into

---

[431] Albers / Braun (2011), Braun et al. (2013a), Albers / Braun (2012), Braun et al. (2013b)

[432] Albers / Braun (2011), Braun et al. (2013a), Albers / Braun (2012), Braun et al. (2013b)

them. This is not to claim that iPeM is the only possible framework to build upon. On the contrary, it would be beneficial for the design science community if other models were adapted, so they could be compared to one another and to the iPeM based approach in particular. However, after considering all the above-mentioned alternatives, the author of this thesis has decided that iPeM is – in his view – the most promising and convincing basis for a new framework.

5.1 Comparing Design Support Development and Product Engineering

In design research, a common perception is that engineering design and the development of design support can both be regarded as creative design processes themselves. E.g., BLESSING ET AL. note: “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.”[433]

This framework has its origin in product engineering. Therefore, the parallels between design support development and product engineering are discussed in the following paragraphs. *Designing*, *creativity,* and *problem solving* are closely related subjects.

Product engineering as a sociotechnical system has been described by ROPOHL who introduced the ‘ZHO-Model’[434]. In this approach, product engineering is described with three interrelated Systems. They are the ‘System of Objectives’, the ‘Operation System’, and the ‘System of Objects’. The System of Objects which in Ropohl’s view contains the results of engineers’ activities (machines, tools, facilities, …). The engineers’ activities, which generate these artifacts, are summarized in the Operation System. The targets which the engineers orientate themselves on, are collected in the System of Objectives.

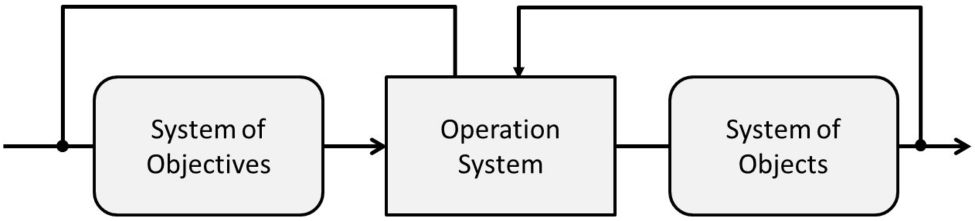


Figure 25 – The control cycle of engineers’ activities (cp. Ropohl (1975), 33)

ALBERS AND MEBOLD, AND later ALBERS AND BRAUN, built on ROPOHL’S Theory of Systems. In the development of the integrated Product Engineering Model iPeM, they describe product engineering as the transformation of a System of Objectives into a corresponding System of Objects. This is done by the Operation System.[435] Aspects of ROPOHL’S model can also be found in several other modelling approaches for design

---

[433] Blessing / Chakrabarti (2009), 178

[434] The German abbreviation ZHO stands for “Zielsystem” (Engl. System of Objectives), Handlungssystem (Engl. Operation System), and Objektsystem (Eng. System of Objects)

[435] Albers / Meboldt (2007a)See Albers / Braun (2011), Braun et al. (2013a)

processes[436]. The systems theory perspective is so fundamental for ALBERS' view on product engineering that it became the second of his five hypotheses on product engineering:

> *„Based on systems theory, product engineering can be described as the transfer of an (initially vague) System of Objectives into a concrete system of objects by an operation system.“*[437]

If we also regard the development of design support as a creative development process, we can transfer ALBERS' iPeM to design support development. The differences between product development and design support development must then be discussed on a more detailed level. In ALBERS' perspective on product engineering, the next, more detailed level can be found within the elements of the system triple: System of Objectives; Operation System; System of Objects.

MEBOLDT presents the most comprehensive definitions for those terms in a product engineering context.[438] ALBERS and BRAUN ET AL. as well as ALBERS ET AL. have later further specified the definitions as more experience with the iPeM was gained.[439] However, taking a closer look, all those definitions aim at and are anchored in product engineering. They are too concrete to be applied to the development of design support without modification. The following paragraphs will briefly review the definitions and identify the parallels and differences thus leading to the *proposal* of a set of definitions tailored to the context of developing design support, as a specific subcategory of product engineering.[440]

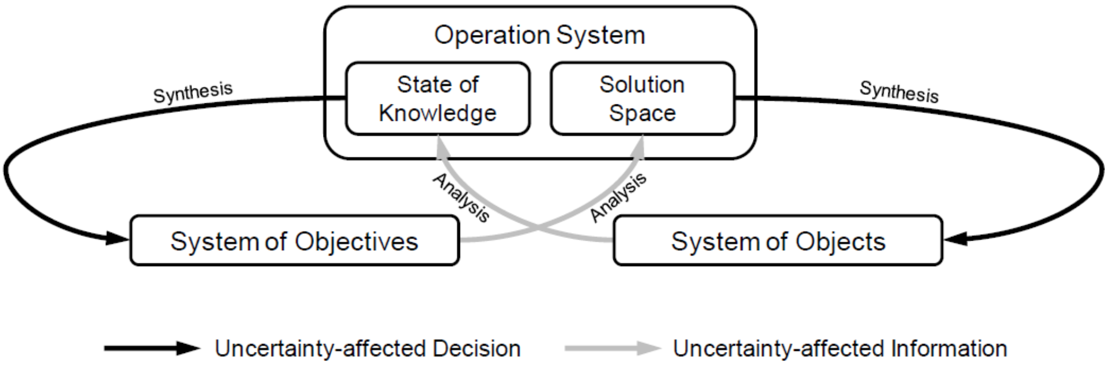


Figure 26 - System Triple of Product Engineering (Albers et al. (2011b), 2)

---

[436] Braun and Albers name e.g. Lindemann, Ehrlenspiel and Negele Braun et al. (2013a). A very comprehensive overview of the development of Ropohl's approach and alternative versions thereof, is given in Lohmeyer (2013), 21 ff.

[437] 2nd hypothesis, Albers (2010b), 4

[438] Meboldt (2008), 95ff.

[439] Albers (2010b), Albers et al. (2011b), Braun et al. (2013a)

[440] This means that the definitions shall not be replaced or compete with Meboldt's baseline definitions. Much rather they have to be compatible with those baseline definitions!

5.1.1 System of Objectives

MEBOLDT gives a definition for System of Objectives in the context of product development, based on ROPOHL, according to which the System of Objectives is initially vague and gets concretized during the engineering project. It *describes all relevant objectives along with constraints, dependencies and interrelations. It contains only information, which must be made explicit and documented in a way that it is traceable and reasonable.*[441]

Later, the understanding of Systems of Objectives was further specified by ALBERS and by BRAUN. *"The System of Objectives represents the set of objectives, their relationships and constraints to depict an intended future condition for developing the right product. The System of Objectives is developed throughout the whole engineering process."*[442]

The later definitions do not claim that the System of Objectives is complete in the end of a development project. This represents the view that the final state of the system of objective within a development project might as well be the starting point of a new project, hence the same content will represent an incomplete system of objectives. Therefore, the term "complete" is directly connected to the corresponding development project (its System of Objects and Operation System).

The key features of Systems of Objectives, according to the definitions are:

- The system of objectives is initially vague.
- It evolves during the design process as it is concretized by the Operation System.
- It contains not only objectives but also includes further information necessary for realization.
- It includes all interrelations of the information, including conflicts, hence provides justification for all decisions.
- All information must be explicitly documented.

Definition for design support development

Specifying this concept from product engineering to the development of design support, the System of Objectives can also be seen as initially vague. It is concretized as the research proceeds and knowledge is generated, thus containing not only the objectives of the design support but also constraints about the applicability of the support. A definition for the System of Objectives in this altered context could therefore be:

---

[441] The full definition is found in: Meboldt (2008), 158

[442] Albers et al. (2012b), 3

The System of Objectives in design support development[443]

*The system of objectives describes **all relevant objectives**, their **constraints**, **dependencies** and **interrelations** (e.g. conflicts) and justification of a design support. The system of objectives contains the **explicit documentation** for its realization. Its elements must be **traceable and reasonable**. It contains only **information, no physical objects**, thus it becomes the repository of reliable knowledge and planning of design support development.*

*Starting from an **initially vague** system of objectives, a **more complete** system of objectives is developed as the research proceeds. The vague system of objectives is being **concretized** and expanded **by** the **operation system**. The system of objectives must be **checked for consistency** time and again. If inconsistencies arise, its elements need to be revised.*

*As the design support is being **utilized**, **experience** and further **knowledge** about it are **generated**. These can be **fed back into the system of objectives**. Hence, the system of objectives is **never complete**. There is **always** room for **improvement**.*

The development of the system of objectives is a core aspect of the design support development.

5.1.2 Operation System

In ROPOHL'S understanding, an Operation Systems contains everything necessary for an operation to occur. With operations he means activities that are executed by the operation system, as transformations of information, matter, and energy.[444]

In MEBOLDT'S definition of the Operation System, it is described as a sociotechnical system. It interacts both with the system of objectives and the system of objects, as it analyses and synthesizes them in an iterative manner. All resources necessary for this are also part of the Operation System.[445]

Later, ALBERS and BRAUN distinguish activities from resources:

"*The Operation System is a sociotechnical system that contains structured activities, methods and processes. Additionally, it contains the involved people and required resources. The Operation System analyzes and synthesizes the Systems of Objectives and of Objects in an explorative, iterative and co-evolutionary process*"[446]

Those definitions are not explicitly specific to product engineering, thus they can be transferred to the description of the Operation System in design support development. However, the key features are:

---

[443] Based on Meboldt (2008), 158; Albers et al. (2012b), 3; Braun et al. (2013a), 4

[444] Cp. Ropohl (1975), Ropohl (2009), 93-117

[445] For full definition see Meboldt (2008), 159. He derived it from a collection of previous definitions found in literature. Those are summed up in Meboldt (2008), 96 f.

[446] Braun et al. (2013a), 4

- The Operation System is *sociotechnical*
- *contains structured activities, methods and processes*
- *contains people and resources*
- *analyzes and synthesizes the Systems of Objectives and of Objects*

The Operation System in design support development[447]
The Operation Systems in design support development is a scientific and sociotechnical system. It is composed of structured activities, methods and processes. Additionally, it contains the involved people and the required resources. Involved people are the design researchers and designers. The Operation System analyzes and synthesizes the System of Objectives and the System of Objects.

However similar to the original definitions, one needs to be aware that it has originated from a context very much specific to product engineering, according to MEBOLDT: “The Integrated Product Development Model” […] acts as a foundation for a product development process that models logic and language for research and practice.”[448] Transferring the definition it to a different context cannot be done without further clarification and interpretation of the definition in that altered context.

5.1.2.1 Interpretation in the Context of Design Support Development

Design support is a possible outcome of design research. Here too, the Operation System is a sociotechnical system. It contains both the researchers that actively analyze and synthesize the System of Objectives and the System of Objects, as well as the users of the design support. The users are the participants of experimental or empirical studies. They represent the later users of the design support, so they take the role of the customer. From them, information for the System of Objectives has to be extracted, which can be done either through controlled experiments or through empirical observation. The design researchers take the role, product developers have in the Operation System of product engineering. Equipped with technical and scientific tools, they synthesize the System of Objects, in this case, the design support.

The activities of the Operation System design support development and of product development still overlap in vast parts. After all, the development of new, innovative technical solutions is science![449] While both the development of design support and product engineering can be regarded as creative problem solving processes, they differ in the larger-scale dimension of the product life cycle. In iPeM, ALBERS calls that

---

[447] Based on Meboldt (2008), 159; Albers et al. (2012b), 3; Braun et al. (2013a), 4

[448] Meboldt (2008); note that the “integrated Product Development Model was later renamed to “integrated Product Engineering Model”.

[449] Regard e.g. the design of lightweight systems such as the ceramic clutch developed in the CRC 489. Such achievements are impossible without scientific inquiry. At the same time the scientific inquiry would not take place if it wasn’t for society’s demand for new technical solutions.

dimension "The activities of product engineering"[450] After initial project planning, the development of a product starts with the detection of a profile, going through various stages of maturity such as its production and market launch and finally, its decomposition/the analysis thereof (Figure 27).

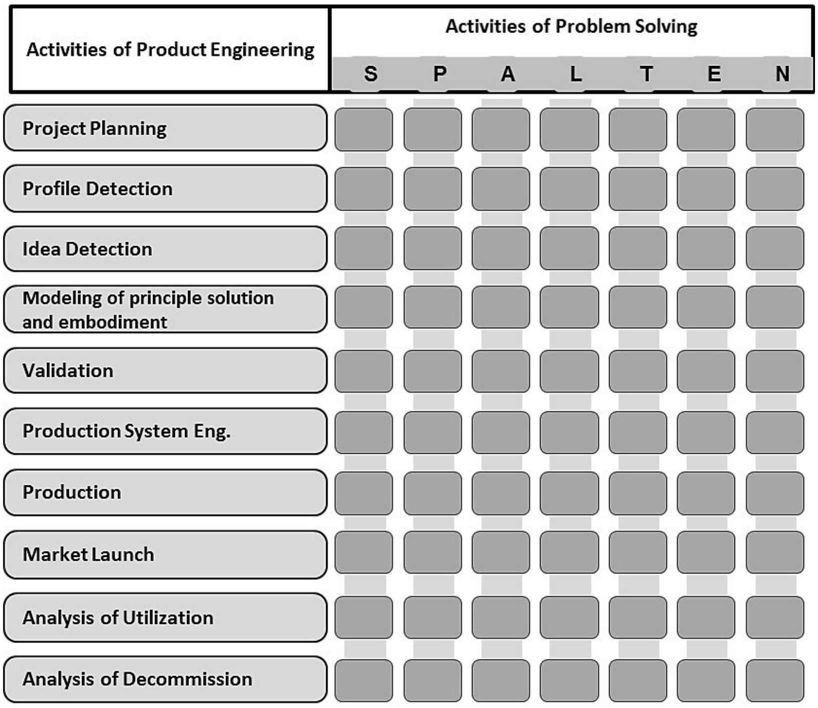


Figure 27 - iPeM's activity matrix, Albers / Braun (2011)

Design support follows a different path. The activities of product engineering cannot be applied. They need to be adapted and specified for design support development. This will be one of the steps expanding the operation system of design support development in chapter 5.2.1.

5.1.2.2 Problem solving Activities of Design Support Development

The development of design support is a creative design process which includes activities of problem solving, just like engineering design processes.[451] "A problem is a deviation between the arbitrarily little known initial state (Actual State) and the desired arbitrarily vague final state (Target State), linked with the partially unknown path from the Actual to the Target State."[452]

Passing each of the different stages of maturity described in the activities of design support development in the previous section can be interpreted as such a problem. At any point in the process, the desired final stage of maturity is the next level of maturity of the design support. How to get there is unclear, resources have yet to be acquired and so on. Hence, it is only consequent to describe the process of working one's way towards the next level of maturity as a problem-solving process.

---

[450] Albers (2010b), Albers / Braun (2011)

[451] Compare e.g. Albers et al. (2011a), or Albers / Braun (2011)

[452] Albers et al. (2005), 2

Different problem solving processes have been suggested in literature.[453] One well suitable problem solving cycle is ALBERS' *SPALTEN*. It is a universally applicable procedure suited both for planned as well as unexpectedly occurring problems.[454] SPALTEN divides problem-solving into seven basic, reoccurring activities (see Table 58). It is fractal, i.e. any of the seven activities can be again described as a problem-solving process which once again can be approached using SPALTEN, and so on.[455]

The SPALTEN problem solving cycle has a rather universal character. Previous application of SPALTEN have shown that.[456]

seems well suited to be the problem-solving process in a framework that addresses the development of design support specifically.

Mind that the choice of modelling approach does not change the basic idea and logic of such a framework. If a researcher feels more comfortable with a different problem-solving approach, he/she has more experience with, it is possible to substitute the SPALTEN logic. However, with the vision of one well established framework in the future of design research, it appears preferable that researchers start to agree on one problem-solving process. This will increase traceability, comparability and ultimately lead to improved acceptance of the outcomes of the operation system – in other words, it will increase the acceptance of design support (compare chapter 2.1).

Table 58 - Problem-solving activities SPALTEN (compare Albers et al. (2005))

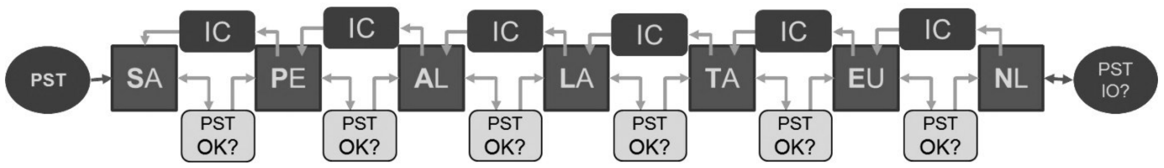
---

[453] Schregenberger (1980); VDI (1993); Albers et al. (2002); Albers et al. (2005); Pahl et al. (2005); Ehrlenspiel (2007); Lindemann (2009)

[454] See also: Albers et al. (2002), Albers et al. (2005)

[455] For more details on the fractal nature of SPALTEN see: Albers et al. (2010)

[456] Albers et al. (2005), Albers / Meboldt (2007b), Braun et al. (2013a), Braun et al. (2013b)



	Problem solving activity	Description
SA	Situation analysis	Gather relevant information
PE	Problem containment	Focus on the relevant problem and related information.
AL	Detection of alternative solutions	Apply intuitive or discursive methods to generate a variety of possible solutions.
LA	Selection of solutions	Assess different solutions from A, then select the best alternative.
TA	Analysis of consequences	Investigate possible risks and opportunities supposing you would implement the selected alternative.
EU	Deciding and implementing	Final decision for the alternative and consequent implementation.
NL	Recapitulation and learning	Analysis of the process and storage of gained knowledge.
PST	Problem Solving Team	Team can exist of one single researcher or a group of researchers. It is not fixed throughout the project but may change.
PST OK?	Is Problem Solving Team still O.K.?	Before and after each step, the PST has to be challenged: Was it the correct team for the last step? Should it be repeated with a different team? Is the current PST the correct team for the step ahead?
IC	Information Check	Check whether the information from previous step which led to decision/action has meanwhile changed.

5.1.3 System of Objects

ROPOHL’S *original understanding of Systems of Objects is that they contain only artificial ‘things’, created by human beings – artifacts. They transform, transport, and store matter, energy, and information.*[457]

Meboldt suggests a definition in which he further specifies that the artifacts can be both tangible and intangible outcomes and explicitly states that there must be corresponding information about the elements in the System of Objects that can be found in the System of Objectives. He also specifies intermediary results in product engineering as elements of the systems of objects.[458]

ALBERS and BRAUN later specified MEBOLDT’S definition:

”The system of objects comprises developed artifacts. It includes not only the final product, but also intermediate steps or results on its way to finalization, such as

---

[457] Ropohl (1975), Ropohl (2009), 117-134

[458] The full definition can be found in: Meboldt (2008), 159. He derived it from a collection of previous definitions found in literature. Those are summed up in Meboldt (2008), 96

documents or prototypes. The system of objects is (ideally) finalised, when its state corresponds to conditions described in the system of objectives."[459]

5.1.3.1 Interpretation in the context of design support development

Transferring ALBERS definition to design research and more specific to the development of heuristic design support, again both tangible and intangible outcomes of the activities conducted by the operation system are imaginable.

Tangible objects are e.g. printed questionnaires, developed and used by a researcher during the development of a design support. *Intangible outcomes* are all kinds of knowledge about design, methods and procedures about how to do design (better), and so on. Within the field of heuristic design support, the by far larger portion of objects will be intangible.

If the development of design support is conducted in a scientific manner, it must be traceable and meaningful. *Hence, each object must have its corresponding system of objectives.* Although a corresponding system of objectives does not guarantee scientifically sound design support, any design support that is suggested without a corresponding system of objectives is guaranteed to be anything but scientific.

Whether or not the three roles MEBOLDT assigns to systems *of objects* (resources; scientific objects and outcomes of the operation system) may be transferred or have to be reduced depends very much on one's personal research paradigm. If ones fundamental believe is that the only acknowledgeable outcomes of scientific activities is documented knowledge, *scientific objects* and *outcomes of the operation system* become one and the same thing. If design science according to one's personal paradigm also produces artifacts (tangible and intangible), the differentiation between scientific objects and outcomes of the operation system can be abided.[460] In any case, scientists produce objects that they use as resources as part of the operation system.[461] At the end of a design support development project, the system *of objects* becomes the design support. This puts a strong emphasis on the correct and thorough documentation as part of the system of objects. For the researcher as part of the *executing resources in the operation system*, the activity of documenting becomes one of the key activities under this perspective.[462]

All other objects, intermediate results, documents, experimental results and so on remain part of the system of objects and can be retrieved at any point in the future, e.g. when a variation or improved version of the support is to be developed.

The last part of MEBOLDT'S definition states that intermediary results are systems of objects of sub-systems of objectives. This is true for design support development as

---

[459] Albers et al. (2012b), 3

[460] The competing research paradigms are discussed in chapter 4.4

[461] E.g. a coding scheme derived from a preliminary study (see also chapter 4.2)

[462] This is quite compatible with the stereotypical picture of scientists sitting at their desks, writing publications.

well. Here to, until the final design support is completed, these intermediary results serve as resources (e.g. the results of a pilot study or preliminary versions of a questionnaire, …) or they produce knowledge for the operation system (e.g. results from an experiment showing that assumptions about success factors were wrong and need to be revised).

5.1.3.2 Definition for Design Support Development

A definition of the "system of objects" specifically tailored to design support development would therefore be:

The System* of Objects *in design support development

Systems of objects in the development of design support are artifacts. They can be tangible and intangible outcomes of the operation system. The objectives of a system of objects must be described in the corresponding system of objectives. Otherwise, the actions of the operation system would be unscientific. In scientific development of design support, each object has to have a corresponding system of objectives. Upon successful completion of a development project, the system of objects is consistent with the documentation of the design support.

Intermediary results remain in the system of objects*. Until the final documentation of the support is completed, the systems* of objects *serve as resources, or they produce knowledge for the operation system.*[463] *After the final documentation, they may serve as resources of the operation system for future systems of objectives.*

5.2 Expanding the Operation System

ALBERS' iPeM has shown to be a useful approach to support product development.[464] The framework suggested for the modelling and navigation through design support development will be based on iPeM, tailored to the specifics of design support development. Its resemblance with iPeM has several advantages: Design researchers familiar with the current literature on design processes will most likely be familiar with iPeM, hence the suggested framework will be easy to grasp and can be applied instantly.[465] Believing that the development of design support is nothing but a very specific development project, it is only consequent to build on existing models and frameworks that already exist for development projects. It was shown in the literature review that not doing so would actually decrease the credibility of this work (see chapter 2.1) and add to the already considerable criticism against design research as a science.

---

[463] This definition is not supposed to be a new definition but is derived from Meboldt (2008), 159, and Albers et al. (2012b)

[464] Albers / Braun (2012); Braun et al. (2013a)

[465] For readers unfamiliar with iPeM: Albers (2010b); Albers / Meboldt (2007a); Albers / Braun (2011); Albers et al. (2011a); Meboldt (2008)

5.2.1 Activities of Design Support Development

It was announced in chapter 5.1.2.1 that the activities of product development as Albers has defined them for iPeM need to be modified in order to suit the specifics of design support development. MARXEN AND ALBERS have presented a model that describes how design support passes through different levels of maturity.[466] From the further development of the model, the Activities of design support development are derived. The model is based on contemporary research methodology in design science, most of all BLESSING'S AND CHAKRABARTI'S DRM (4.1.1) and CANTAMESSA'S empirical findings (4.1.3). Analysis and discussion of the model have shown that:

- it is compatible to existing research methodology such as DRM (4.1.1), REICH'S Layered Model (4.1.4) or the Validation Square (4.1.5)
- it includes research related to education studies[467], which contemporary attempts to describe design support development lack (see also Table 60).
- it links CANTAMESSA'S five categories and puts them in a logical order.

Consequently, CANTAMESSA'S five categories will be incorporated into the framework, substituting the activities of product engineering for this specific field within design - design support development (Table 59).

Figure 28 shows ALBERS' AND MARXEN'S model of how design support matures through its different stages using CANTAMESSA'S five categories (cp. Table 59). There are two possible initializations (I) resulting from empirical observation of real-world design processes. The identification of *difficulties in practice* can lead design researchers to try to develop support in order to reduce or eliminate those difficulties (II). Alternatively, design scientists might find a designer/team to be very successfully at solving certain problems. The developers of design support try to describe, understand and generalize what makes the observed designer(s) so successful. They suggest a design support that resembles the designer's actions in order to offer support to other designers (II). Both origins for the design support are of empirical nature since they result from observations of reality. At the same time, the real-world origin ensures the development of *relevant* design support![468]

---

[466] Marxen / Albers (2012)

[467] Birkhofer and Jänsch have dedicated comprehensive research efforts arguing that design methods need to be teachable, e.g. Jänsch / Birkhofer (2004). Albers states that if it cannot be taught, it is not a support, e.g. Albers et al. (2006)

[468] Relevance being one of the key objectives for the development has been emphasized throughout this thesis and in literature: It is a key feature of DS I in Blessing / Chakrabarti (2009) but also for Reich (1994), 7; Pedersen et al. (2000), 4; Keller / Binz (2009), 2-205; Lienert / Raatz (1998), 29ff.; Bender (2003), 401 ff.

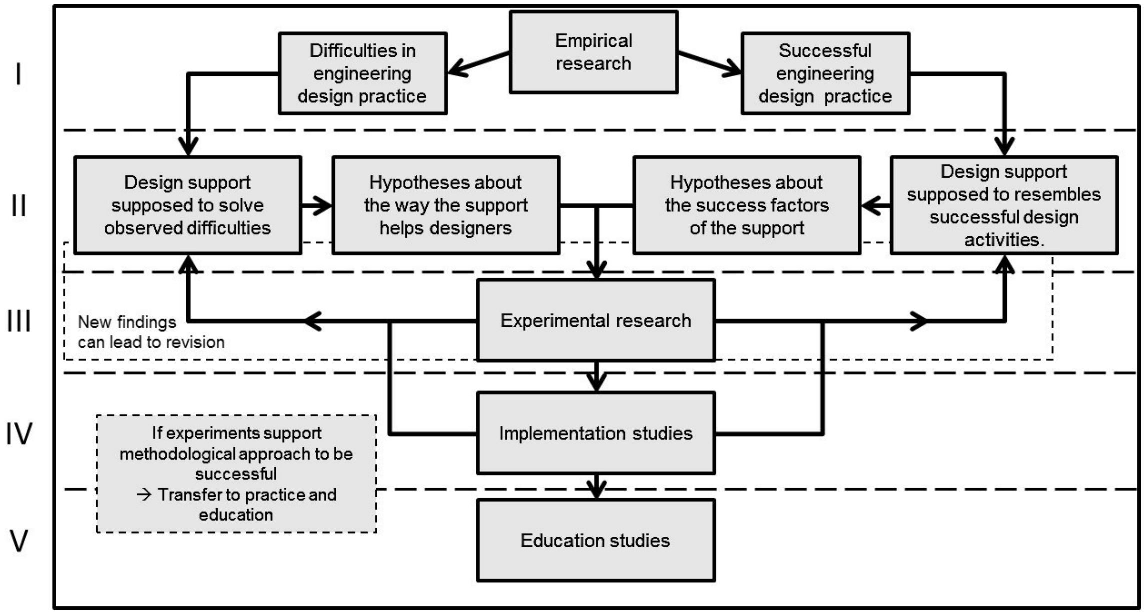


Figure 28 - Development path of design support, own illustration

In the next step, the support needs to be evaluated (III), which can ultimately only be done by experiment. Theoretically, one could skip the step and proceed right away to industrial implementation. However, it would be hard to find willing partners, nearly impossible to show that the support works as intended and in the likely case of the support failing to perform right away, the scientist would damage the reputation of the support and design science as a whole. Therefore, it must be avoided at all means to skip experimental evaluation. If the results support the hypotheses and assumptions made in (II), the next level of maturity of the design support can be tackled. Its evaluation is taken from laboratory-like conditions to real-life industrial situations (IV). Only if the assumptions and hypotheses can be upheld under these realistic conditions, the support can mature further and finally, it is useful to analyze how it can be transferred to educational programs (V). In reality, the step by step maturation from I to V is rarely encountered. The more common case will be revisions and repetitions along the process – hence the backwards loops in the model. Experimental research (III) may reveal inconsistencies. In this case, the validity of the chosen experiment needs to be checked. If the inconsistencies cannot be eliminated through improvement of the experimental setup, the underlying assumptions and hypotheses need to be revised. Even if the experimental stage is passed successfully, implementation into an industrial environment (IV) may prove impossible leading to backwards iterations. This does not diminish the value of the research. Iterations are common in any scientific endeavor. Why should it be different in the development of design support?

Table 59 - Five Categories of design research (Cantamessa (2003)

I	Empirical research, in which researchers analyze real-world design processes.
II	Development of new tools and methods for supporting the design process or elements of it.
III	Experimental research, in which researchers purposely set up design processes in controlled environment
IV	Implementation studies, in which researchers discuss the real-world deployment of innovative methods and tools.
V	Other, which includes papers dedicated to theory and education.

In summary, the left side of the model shows a *demand-driven development*. It can be compared to what epistemologists call *deductive reasoning*: A theoretical approach is developed, making assumptions about what might help to solve the problem and then try to confirm this by experiment and observation. The *right side* shows some typical parallels with *inductive reasoning*, where a successful pattern is *generalized* to a theory that is assumed to be true as long as it is not proven wrong.

With ALBERS' & MARXEN'S model, it is possible to describe both problem-driven development of design support as well as success-driven development of design support. Revisions and backwards iterations can be mapped. The often neglected final step of transferring the new design support to education is included as well. It is necessary in order to equip potential, future users of the support with the necessary knowledge about its application. The comparison with DRM as shown in Table 60 shows the compatibility between the two models.[469] In total, it can be concluded that the five categories well fit to be used to cluster activities of design support development.

Comparing CANTAMESSA'S five categories, DRM or ALBERS' and MARXEN'S model with the activities of product engineering in the iPeM, one aspect is still missing in all three approaches that iPeM has: A defined starting point. In the development of iPeM, ALBERS ET AL. have revised and improved it over time. While the early publications put as the first activity of product engineering[470] "*Profile detection,"* it became later apparent that initialization of development projects can be better modeled as a set of activities summarized as "*Project Planning." Planning*".[471] A clear advantage is that setting up the *initial System of Objectives* can be modeled independently from a product profile, allowing e.g. for strategic goals to be the trigger for a development project.

Transferring this idea to design support development the question is: "What is it that leads to the start of a design research project?" "Why does a researcher start to make observations?" In order to be able to model the initial activities that kick of a research

---

[469] For a more comprehensive discussion of the compatibility of the model with DRM: Marxen / Albers (2012)

[470] At the time the activities of product engineering were still called „Macro activities"

[471] Albers (2010b)

project, it seems advisable to include activities of project planning in the activities of design support development in section 5.1.2.2.

A further activity that has developed over time in the iPeM is “Analysis of Utilization”. Regarding Product Engineering as one process for each single product, proved to be a very limited view. In fact, a large part of product engineering projects are those, in which a next generation of a product is being developed. Examples can be drawn from various industries, e.g. cellular phones or the car industry.[472] What makes Analysis of Utilization so important is that it is the key to modelling where and when insights about the users' natural interactions with the product and market acceptance are gained. Feeding these insights back into the system of objectives, this system of objectives can then be the starting point of the next generations system of objectives, triggering a new development project in the “Project Planning and Controlling” activities. Integrating both Project Planning and Controlling as well as Analysis of Utilization is necessary in the development of design support, as it enables the Operation System to close the loop and also describe continuous improvement of design support as well as adaption of existing design methods.

Based on this reasoning, the activities of design support development can finally be derived:

Activities of design support development

- Project Planning and Controlling
- Empirical Research, Observation of real-world Design Processes.
- Embodiment of Design Support
- Experimental Studies, Evaluation in controlled Environment
- Implementation Studies, real-world Deployment of Design Support.
- Transfer Studies dedicated to Industry and Education.
- Analysis of Utilization

---

[472] When the author started working on his thesis in 2007, the first iPhone was launched. In 2013, the 5th generation and 8th model will be launched. Car manufacturers commonly launch a new model every three years by now.

Table 60 - Comparing DRM and the proposed model

Demand driven	Success driven	Research Clarification	Descriptive Study I	Prescriptive Study (PS)	Descriptive Study II
Project planning		(X)			
Empirical research	Empirical research	X			
Difficulties in engineering design practice	Successful engineering design practice	X			
Design support supposed to solve observed difficulties	Design support supposed to resembles successful design activities.		X		
Hypotheses about the way the support helps designers	Hypotheses about the success factors of the support		X	X	
Experimental research				X	X
Implementation studies					X
Education studies					
Analysis of Utilization					(X)

Table 60 draws a comparison between the activities of design support development and what is addressed in DRM. Project planning and Controlling is only partly addressed there as the final step of Research Clarification.[473]

What DRM does not address at all is the transfer of the design support into education. In the Descriptive Study II, success evaluation is addressed as the ultimate evaluation stage. Its outcomes are strictly linked to the "Measurable Success Criteria" defined early in a DRM-guided project.[474] Success evaluation tests whether the design support really achieves what it promised in a real-world environment, e.g. shorter time-to-market. These insights are only accessible in long term observations. They are extremely difficult to achieve and suffer from strong bias as surrounding conditions change over time. What DRM does not offer is the integration of simple yet valuable insights about potential improvements that can be gained during or after the application evaluation in real-world design processes.

5.2.2 Activities Matrix for Design Support Development

The activity matrix for design support development consists of two dimensions: The activities of design support development as derived in the previous section and the problem-solving activities.

---

[473] Blessing / Chakrabarti (2009), 67

[474] Blessing / Chakrabarti (2009), 184 ff.

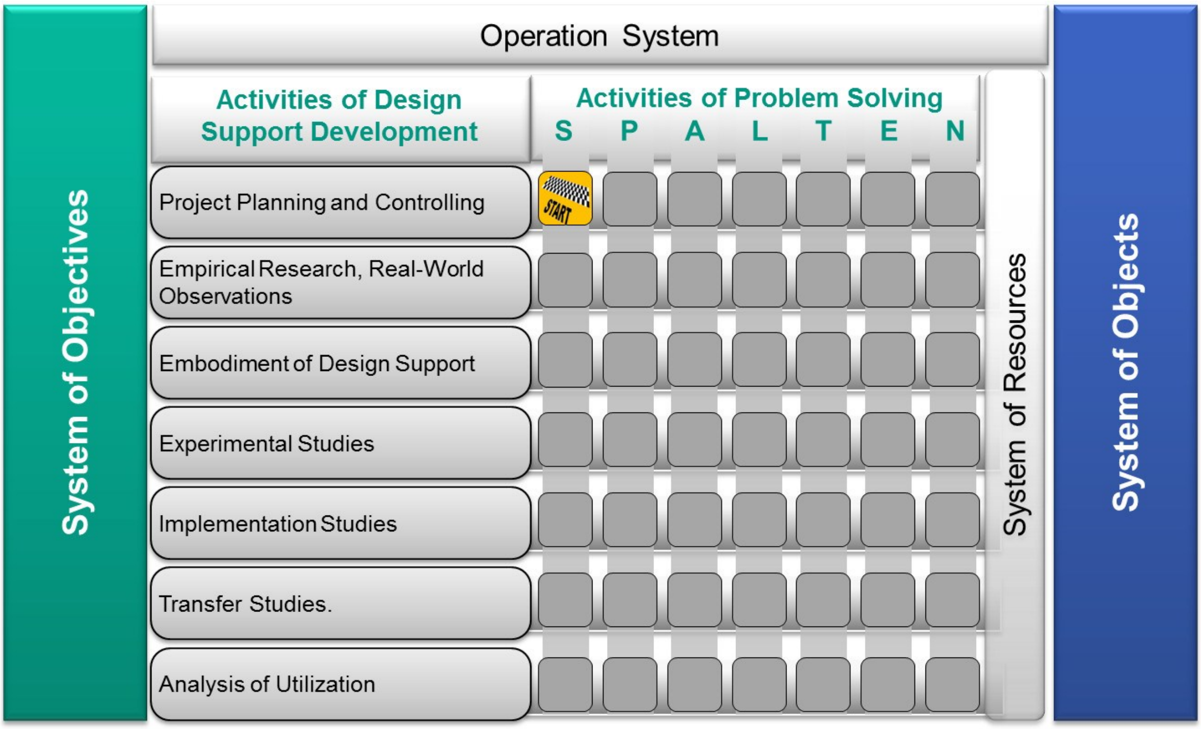


Figure 29 - The integrated Design Support Development Modell iDSDM, (source: own illustration)

In the author's point of view, the most suitable activities for the here specified context are the universal problem solving cycle SPALTEN (cp. 5.1.2.2) and the activities of design support development as developed in chapter 5.2.1. They are based on CANTAMESSA and further modified from MARXEN AND ALBERS with an added, initial layer of project planning activities (as reasoned in 5.2.1).[475] The resulting matrix is shown in Figure 29. In the following paragraphs, the activities are described and explained, according to the activities of design support development.[476]

5.2.2.1 Project Planning and Controlling

Project Planning activities initially kick off the design support development project. The key deliverables are the initial system of objectives and the project plan. They are both developed and aligned with the available resources under the guiding questions:

- *What* is to be done,
- *When and in which order should it be done*
- By *whom* and *how* will he/she do it?

A typical outcome can be e.g. a research proposal to be reviewed by peers, a customer, or a sponsor. However, it is quite possible that the project planning activities are not conducted by the same person / team as later research activities. In research groups if e.g. team managers, professors or senior scientists do the planning or write the proposals while the doctoral students in large parts execute the research and

---

[475] Cantamessa (2003); Marxen / Albers (2012)

[476] Both dimensions can principally be realized with different suitable sets of concrete activities, if the researcher can logically argue why he/she is using different activities.

development projects.[477] So, for the individual's perspective, the initial step can be to receive the order/proposal from a researcher one's senior.

Controlling in this context refers to the continuous comparison between the conducted research activities and the planned research activities. Changes may be necessary as the design support development project matures, and new information is added to the system of objectives or if changes are made in the resource system – e.g. a company initially willing to participate in an observational study announces that it will no longer participate. The reasoning for the changes takes place during controlling activities. Consequently, during design support development, regular iterations are included always looping back to controlling. The planning and controlling activities can be supported with the dynamic phase model.[478]

Any changes to the initial proposal, the reason for the changes and so on should be documented, hence, the outcome of all controlling activities at the end of the project is its full documentation - in funded research, a final report is usually a mandatory deliverable. Another central object that should generated in project planning and controlling is the System of Objectives for the design support. Empirical studies have been conducted, investigating such questions as "What are the demands from industry and science regarding design support?" (see also chapter 2.1) The targets and requirements identified in those studies have been collected and summed up in a checklist (

Worksheet 3). Some very general aspects are applicable as targets that should always be fulfilled such as "A method should not need too much effort!" Equally, there things to be avoided when developing design support, as well as dangerous false promises one should not make when presenting a design support. Those general aspects are listed in

Worksheet 3. How all this is actually achieved on a concrete level, depends on the characteristics of the design support and cannot be easily answered. Instead, for any design support development project, its individual system of objectives has to be developed.

Worksheet 4 can serve as a starting point for this development. Its application will help researchers clarify their objectives and make them explicit for their initial system of objectives.

5.2.2.2 Empirical Research, Analysis of Real-World Design Processes.

Large parts of what BLESSING AND CHAKRABARTI describe in their "Research Clarification" in DRM happen in empirical observation of real-world design processes and/or the review of corresponding literature. Difficulties and/or good practices are

---

[477] Whether or not this separation of writing proposals and executing research is advisable or not depends on culture and preferences within the research groups. It may well be subject to general discussion but is not to be discussed here.

[478] A full explanation of the dynamic phase model will be given in 5.2.4

identified. DRM demands very explicitly to compare one's observations to the current state of literature. This is also true in this framework, since any empirical research can and should be accompanied by a review of literature.[479] Only through comparison with literature, one can ensure the developed design support is *new*, and only through careful observation of practice, one can ensure that it is *relevant*.

So, in the analysis of real-world processes, the design support developer can be in search of difficulties, designers encounter, and later suggest design support, which reduces these difficulties. Alternatively, he/she can be searching for noticeably successful designers and their practices to try and resemble those successful practices, with a design support. Of course, a combination of both aspects is possible as well as a non-directed, explorative research in which the observation is initiated with the goal to reveal any type of curiosities, which then lead to the posing of further research questions.

The empirical research stage is passed (as all other activities of design support development) as a SPALTEN process beginning with a situation analysis. Typically, this would be a first literature review with the goal to identify relevant literature, previous empirical studies and so on. In problem containment, the researcher sorts out, which statements he can build on from literature and which observations literature still lacks. In doing so, he/she focuses the empiric research. Alternative empirical studies are designed. For generating the different potential empirical research designs, the pool of promising research methods (chapter 4) now becomes relevant. Filtering all available methods for those appropriate for empirical observations, the researcher gets a list of methods to choose from. The corresponding profiles are also documented in chapter 4. They support the quick and easy assessment of the alternative solutions. The researcher can weigh advantages and disadvantages, check for which of the different approaches sufficient resources are available, and decide which alternatives must be discarded, and so on.

After assessing the pros and contras of the alternatives, one of them is chosen.[480] The consequences of the choice are analysed and finally, a decision is made. This process is recapitulated, and the gained knowledge is stored (e.g. if one of the discarded alternative solutions shall be conducted in the future, when sufficient resources are available).

5.2.2.3 Embodiment of Design Support

The term 'embodiment' is not limited to physical, materialized embodiment. It is the Operation System's core activity of synthesizing the System of Objects, which can be tangible and intangible. While this may be the case in some exceptions, in most cases, embodiment of design support will take place by externalizing and synthesizing the

---

[479] cp. e.g. Yin (1994), or Eisenhardt (1989)

[480] One alternative can always be not to do a study at all, e.g. when literature already contains all the observations necessary to derive the later design support.

activities, tools, methods, and processes, necessary for the utilization of the design support. The researcher creates something artificial. This can be software tools supporting a certain workflow or graphical representations of the activities that are part of a design support, such as flow charts, or worksheets, and so on. In the Embodiment of design support stage, the information obtained from the observation before, is synthesized into a design support that helps overcome difficulties or resembles success factors. The initial synthesis is based on assumptions – the interpretation of what has been observed.

In the situation analysis, therefore, one reviews all the gathered information thus far. This includes both literature findings, and the data collected in the empirical research (if applicable). An additional literature review of findings publicized in the meantime should be done, to ensure that one's work is still new and relevant.

The information is interpreted in problem containment leading to assumptions about success factors or difficulties in engineering design practice (cp. Figure 28).

Alternative concepts for the design support based on the assumptions derived from the findings. The alternatives are analyzed, and the best alternative is chosen for further implementation after a thorough analysis of consequences, e.g. by careful comparison with the initial system of objectives. In the analysis of the consequences, BLESSING'S impact model can be a useful tool.[481]

Implementation in this context means the actual synthesis of the design support. This can be a complex and lengthy procedure possibly worthwhile modelling as its own problem-solving process, making use of the fractal character of SPALTEN. Finally, lessons learned from this process are documented before we go into the next stage.

5.2.2.4 Experimental Studies, Evaluation in Controlled Environment

The first step in building credibility for a design support is its evaluation in an experiment or a series of experiments. The degree to which the researcher can control the test situation may vary according to the available resources. Most important, this is a fundamental activity that ensures the difference between an educated guess and scientific research.

Therefore, in Situation Analysis, the researcher needs to determine: Which aspects can be controlled, which can be observed (directly and indirectly), and what must the experimental setup look like in order to evaluate the design support and its underlying assumptions. When the experimental design is developed, it is most important to maintain a high level of validity (cp. Chapter 4.3.3).

Filtering the methods provided in chapter 4 for those, suitable for experimental studies gives a selection of potential experimental setups thus supporting problem containment. Alternative experimental designs are generated, and with the provided

---

[481] 4.1.1.1

method profiles (chapter 4), they can be evaluated. Different considerations for the evaluation are given also in Worksheet 5. In the analysis of consequences, the researcher must search for potential sources of bias and select a final experimental design with minimum bias.

Implementing will lead to actually conducting the experimental study, including data analysis for which again, methods and the necessary statistics have been provided. Recapitulating and learning will include reasoning about which methods worked and, which should be altered in future studies. For the implementation, suitable design tasks need to be found, to test the design support's performance. Task design is a method dedicated to this step and has been presented in chapter 4. Questionnaires have been derived to support evaluation based task design according to BENDER and SCHRODA (Worksheet 6 and Worksheet 7).

The outcome of the experimental study reveals whether or not the design support achieves the intended results, i.e. whether it helps overcome observed difficulties or successfully resembles the previously discovered, good practice.

If the experimental results do not support the concept of the suggested design support, the researcher has to iterate back and revise. Several things might have "gone wrong": The suggested support might simply be inadequate. However, it is also possible that the experimental setup has flaws that lead to unintended outcomes. Furthermore, the basic assumptions from the empirical analysis need to be revised. It is possible that observed difficulties or success factors are closely linked to a very specific situation and do not occur or work under different conditions as in the experiment. In this case, the relevance of the intended design support is very much at question, and the researcher has to decide carefully whether he/she wants to alter assumptions and the resulting design support, or to discard the support and go back to observation, where he/she can discover more relevant difficulties or best practices.

5.2.2.5 Implementation studies, real-world deployment of design support.

While a successful experimental evaluation is necessary in order to build credibility, controlled experiments alone will not be sufficient. Stopping the design support development, here would stoke the criticism explained in chapter 2.1. No matter how impressive results from experiments in university laboratories might be, eventually they cannot show how well a design support performs in a realistic environment with experienced engineers who deal with multiple projects, multiple conflicting goals, and usually a lot of pressure regarding time and money. This is why implementation studies are dedicated their own stage in the framework and not sub-summarized under experimental studies. Another important issue is the problem solving team for the implementation studies. Such studies should always be accompanied by researchers with industrial experience to increase acceptance (compare chapter 2.1).

In Situation Analysis, the researcher needs to gather information about the main differences between the conditions in so far conducted experimental studies and real design environments. What could be shown thus far and what couldn't? Problem containment will deal with the subsequent question of what are the key factors that still need to be shown in realistic conditions and where and how to find suitable conditions in practice that resemble the application, for which the support is intended. The resource system can be an extremely limiting factor. Finding the right companies willing to participate in implementation studies can be a challenge. Alternative study designs with different partners must be set up from which the most suitable is/are chosen.

Methodological support is given just like for experimental studies. A set of methods can be filtered and assessed to support efficiently going through the problem-solving process SPALTEN.

If the implementation study reveals flaws in the design support that makes it insufficient for practice, either alternatives from the embodiment stage can be tested hoping for better results, or alterations can be made to the design support. Depending on the degree to which it is changed, retesting in an implementation study might be risky and a repetition of an experimental study might be advisable.

5.2.2.6 Transfer studies dedicated to industry and education.

If however, the results from the implementation study show that the support is feasible, the final steps can be tackled, and education concepts are suggested. In situation analysis and problem containment, it must be clarified who should be educated and how the researcher gets access to those that are supposed to learn about the newly developed design support. Can it be self-taught by those who will apply it or does the support require a moderator? Should it be taught to students or will it only be applied by a small group of professionals? In that situation, industrial lectures or trainings might be the approach of choice. With the actual problem specified, alternative solutions (e.g. e-learning courses, book or classic lecture material) can be conceptualized and assessed in the selection step. After thorough analysis of the resulting consequences, the concept is implemented, i.e. the lecture material is put together, the e-learning course set up, or whatever the chosen concept turned out to be. The lessons learned are finally stored.

The knowledge gathered from the transfer studies is fed back into the system of objectives and may lead to changes that are then executed by the operation system. Such may be the case if elements of the design support turn out to be too difficult to teach. "If it cannot be taught, it is not a design support as it will not help anyone."[482]

---

[482] This statement is one of the key messages Albers gives to his students in his Product Engineering Curse.

5.2.2.7 Analysis of Utilization

Once it is known how to transfer the design support to industry, it will develop its own dynamics. The researcher is not necessarily present at all times it is being used. Only now will it become evident whether the documentation and embodiment of the design support are sufficient for the support to be utilized in practice without its developers promoting it.

Knowledge and experience with the support are being generated by the users. This can be the ideal starting point for the development of a new, improved design support, or an adaption of the design support for certain situations. Those opportunities only exist if dedicated activities are included in a design support development project, hence the inclusion in the framework.

5.2.3 Navigating the Activities Matrix of Design Support Development

The matrix provides a detailed meta model with 42 single steps (Figure 29).[483] Although at first sight this may seem a lot, there is one important rule that should be followed at all times:

Do not skip any of the steps!

In the early stages of iPeM, common perception was that navigating through the SPALTEN-Matrix should be done pragmatically. ALBERS and MEBOLDT actually stated that "the procedure is not to be applied dogmatically but pragmatically depending on the constraints."[484] This statement has been repeatedly misunderstood as an excuse to simply skip single steps, which was not ALBERS' intention. The idea was to allow for different degrees of accuracy, time and effort for the steps. If e.g. problem containment in the idea generation reveals that an initial idea has been given to the company from the outside, and the project aims at testing this idea, generating alternative solutions and selecting from them can be consciously reduced to only the one single alternative. Deliberately deciding against the generation of further alternatives is very different from not thinking about it (which is equal to just skipping the step).

In design support development, it is therefore equally important to systematically and most of all consciously go through every single step. However, this does not mean that equal effort has to be spent on each step. Every support development project will be individually characterized by different priorities. Hence, it is quite well possible that only a few minutes will be spent in certain steps and then a conscious decision regarding the execution of the step is documented after which the researcher proceeds to the next step or leaps back to a previous step.[485] Iterations are possible and often

---

[483] As explained earlier, SPALTEN can be substituted by any problem solving process. This might change the resolution of the process-model's activities.

[484] Albers / Meboldt (2007a), 4

[485] Compare Albers' first Hypothesis of Product Engineering: "Every product engineering process is unique" Albers (2010b)

necessary as newly generated information makes it necessary to revise decisions from previous steps.

Since no step may be left out, a two further rules can be logically derived:

Always start with a situation analysis in project planning and controlling!

and

Start every single Activity of design support development with a situation analysis!

Figure 30 - Always start with a Situation Analysis in Project Planning, (source: own illustration)

When all 42 activities have been conducted, the system of objects should contain a complete documentation of the design support and its development project. The documentation must be in accordance with the system of objectives, which has been continuously updated and refined. The development of the design support is actually finished. It is, however, in the nature of science that “being finished” is never quite correct. Further studies might be conducted with the design support. In the model presented here, this is absolutely possible if later versions of the design support are regarded as a new design support. In this case, some questions or demand for improvement might trigger the next design support development process. Further studies can be modeled as empirical research with the goal to identify difficulties in practice or success factors that might be integrated into the support. It all starts over again.

5.2.4 Dynamic Modelling of Design Support Development

So far, the Operation System of design support development has been expanded by the Activities Matrix. This matrix is nothing but a *static set of activities*. It contains the elements, which are the same in every single project. They do not change, hence this is also called the *static part of the model*.[486]

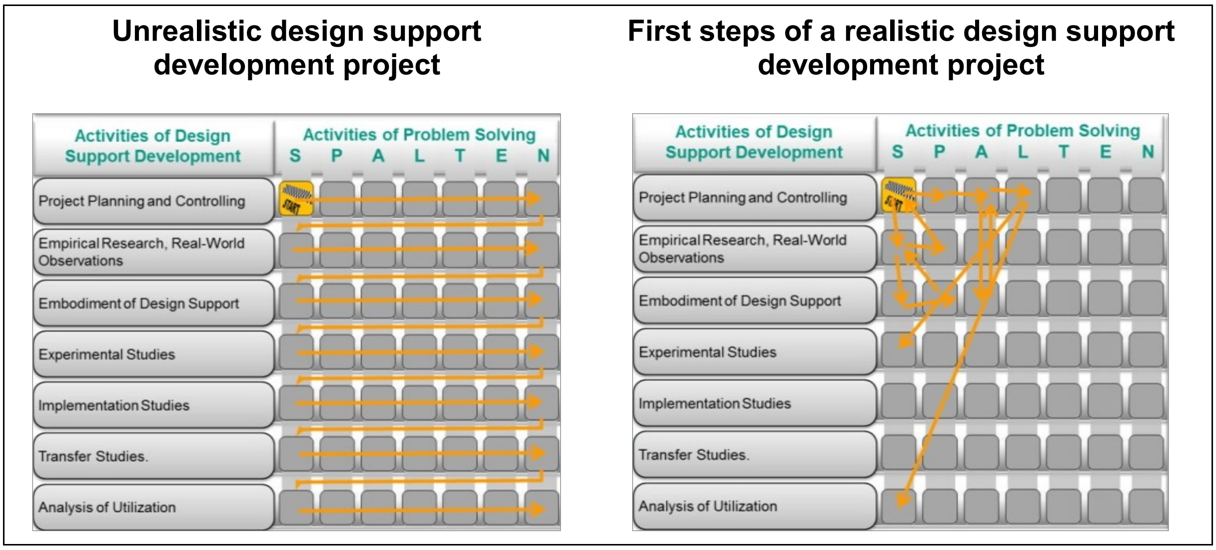
---

[486] This terminology is directly adopted from the iPeM. ,Compare e.g., Albers (2010b), Albers et al. (2011a)

When dealing with an individual design support development process, the questions arise:

- Which activities should be executed?
- In which order should they be executed?
- How much time does each activity take?

It has been defined in the previous section that none of the steps may be skipped. What has not been addressed yet, is the question in which order the activities should be executed? A general answer for this question cannot be given. Real-world development projects are subject to repetitions of activities, iterations, jumping forwards and backwards in the project and so on. Much more, real-world development processes are unique.[487] Merely, in an ideal world, anyone can expect to execute each step sequentially, one after another and produce reliable, new and relevant results.[488] Figure 31 is an attempt to visualize the idealized design support development process (left) and a realistic design support development process. For better readability only the very first steps are visualized. What looks rather chaotic at first sight is the result of consequent information gathering. When e.g. a research proposal is written, with the goal of developing a design support, first pilot experiments or usability studies might be carried out in order to build the proposal (anchored in project planning and controlling) on solid assumptions or even reliable data.[489]



Unrealistic design support development project

First steps of a realistic design support development project

Figure 31 – Idealized vs. realistic Support Development Process

What becomes also apparent from Figure 31 is that a matrix cannot illustrate such dynamic behavior very well. More suitable for showing order and duration is a graphical representation on a timeline, e.g. a Gantt Chart.[490] Figure 32 shows the iPeM with the activities matrix (left) and the corresponding phase model of a specific project on the

---

[487] In Albers' perspective on product engineering the uniqueness of design processes has become the first of the five hypotheses. Compare: Albers (2010b)

[488] Compare Albers et al. (2011a)

[489] Such iterations are described in the exemplary application of the framework in chapter 6

[490] Gantt (1903), cited in Williams (2010), 50 ff.

right. Activities are placed on a timeline in a certain order an they are assigned durations. In this example, the solid grey bar represents the planned duration of the activities whereas the hatched bars represent the actual duration of all past activities.

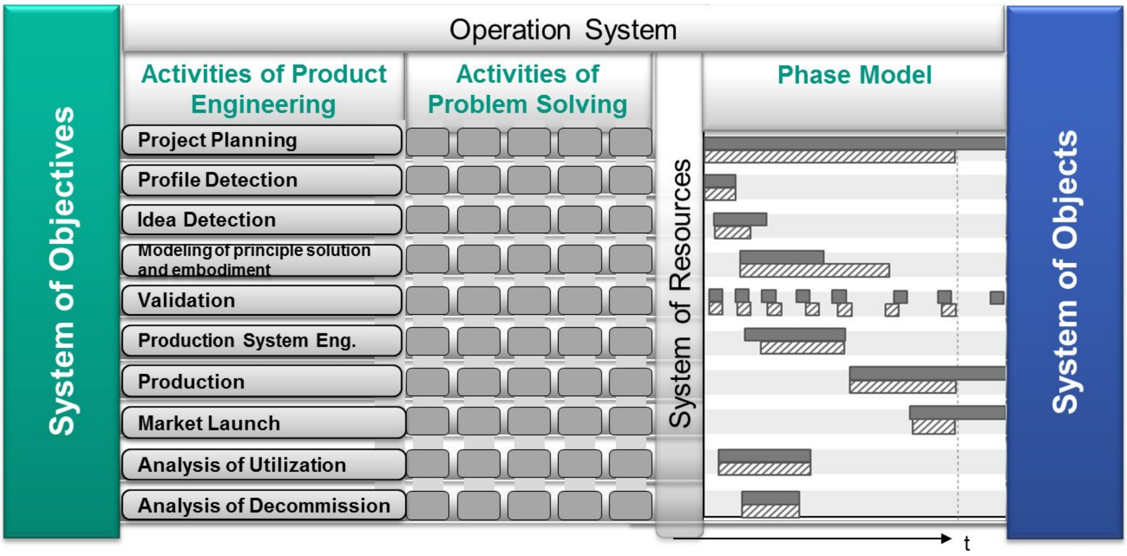


Figure 32 - iPeM with dynamic phase model

Transferring the idea of a dynamic phase model to design support development, a whole new set of possibilities is provided for the Operation System.

The outcomes of ‘Project Planning’ can be visualized and supported. The activities of design support development, as presented in the activities matrix, are put in a planned order and each activity is assigned an estimated duration, based on the available initial information. The result is a timeline for the activities. If visualized with bars it resembles what is also known as Gantt charts. As the project matures, more changes to the initial plan might be necessary. This can be due to additional or more detailed information that becomes available, or it is, in consequence, of changing external conditions. Therefore, the phase model must be updated continuously. The activities to update the model are also part of the Project Planning and Controlling activities.

Expanding the view from one isolated project to many projects, another advantage becomes apparent. Design support with similar Systems of Objectives will be developed conducting similar activities. In consequence, the corresponding phase models will look alike. In reverse, this means that if the framework is applied in a larger number of design support development projects, a very specific question should be integrated in the initial Situation Analysis of Project Planning and Controlling: “Are there any documented phase models from similar projects that could serve as a reference for planning the phase model in my current project?”

This advantage has been pointed in several publications on iPeM.[491] Albers and Meboldt have therefor assigned three distinctive model-layers to the phase model: the '*reference model*', the *'implementation model'* and the *'execution model'*.[492]

In reality, no plan, once put into practice, is fully met, hence the distinction between implementation and execution model. The later the actual activities being conducted in a concrete project, while the *'implementation model'* is the initially planned phase model.[493]

For similar types of projects, typical, reoccurring patterns of phase models can be stored in '*reference models'*. The similarities can be found in surrounding conditions, such as funding. E.g. collaborative projects between industry and science funded by the German Ministry of Science and Education tend to follow a similar plan. The quota between human resources coming from scientific partners and from industrial partners is roughly the same and so forth. The similarities between two projects can equally be found in the System of Objectives. E.g. two projects that aim at analyzing and improving certain work flows in the design departments might follow a similar pattern.

Finally, similarities can be found in the operation system. Three-year-projects conducted by PhD students within a company will resemble one another, but will not be comparable to large-scale ten-year collaborative research centers.

Therefore, using reference models, not only makes Project Planning and Controlling easier, it is also an opportunity to discover imbalances between available resources and the system of objectives in a specific project. If a similar task could *not* be done in the past by one single PhD student, it is unlikely to be successful if one tries again without additional resources.

---

[491] Albers / Meboldt (2007a), Albers (2010b), Albers et al. (2011a)

[492] For the most detailed definitions of the three models see Meboldt (2008), 200 ff.

[493] See e.g. the alternation in the example project which led to major changes ito the initial plan (page 187)

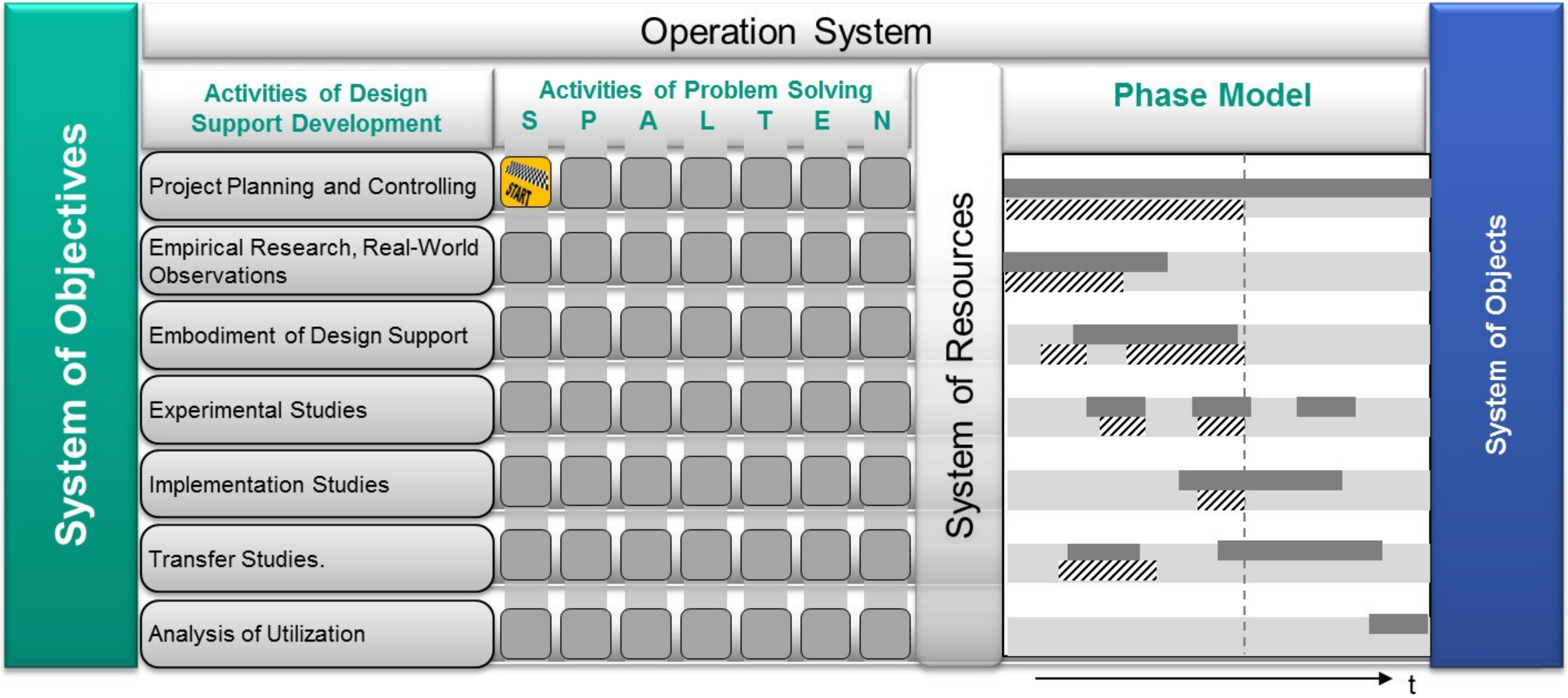


Figure 33 - The Integrated Design Support Development Model with static Activities Matrix and and dynamic Phase Model, derived from the iPeM (Albers (2010b)

6 Exemplary Application of the Framework

In this chapter, the framework is used to systematically model the development of a design support taken from a real research project. The goal is to proof applicability and to identify possible opportunities for improvement. This chapter can also be regarded as an element in the transfer concept for education. The example makes it easier to understand how the elements of the framework, including the IDSDM and also the pool of scientific methods (cp. chapter 4) together with the worksheets provided in this thesis.

The example will be the *Piracy Risk and Measure Analysis (PRMA)*. It is a method that has been developed in a 30-months, government-funded research project on product piracy in cooperation with four industrial partners and one scientific partner apart from IPEK.[494] At IPEK, who also who also had the role of project coordinator, the project was anchored in the design methods and management group.[495]

Note that the starting point of every design support development activity will be its situation analysis leading to a dogmatic navigation pattern throughout the description of the project. Smaller iterations and backwards loops will not be modeled in the upcoming sections for the sake of readability. This is not to imply that no loops or iterations occurred.

6.1 Project Planning and Controlling

In project planning and controlling, the design support development is initiated and kicked-off. The following paragraphs will recapitulate how the development of PRMA was initiated. The name for the support was not fixed then. In the following sections, it will be referred to as "*anti-piracy support.*"

Situation Analysis

The initiation of the project was a call for proposals by the German Ministry of Science and Education.[496] Some basic targets to be included in the system of objectives were initially delivered with the call for proposals.[497] The overall governmental goal was to equip German companies with tools and methods as well as products and systems that help protect companies against product piracy with a strong focus on small and medium enterprises (SME).

Problem Containment

Both security systems and infrastructure as well as design methods were legitimate content for proposals. A preliminary government study had revealed an increasing threat through product piracy and a lack of both countermeasures and design methods

---

[494] www.ipek.kit.edu

[495] Detailed information about the project structure can be found in the final report: Abele et al. (2010)

[496] The German abbreviation is BMBF; http://www.bmbf.de/en/furtherance/6669.php; 5/5/13, 15:03

[497] A translation of the proposal's key elements is available in the appendix: 9.5

to prevent such piracy. It was clarified that the BMBF mainly expected high-tech solutions that would mark original products such as new printing technologies or RFID tags from proposals that would focus on developing “products." Proposals addressing tools and methods were expected to develop design support, which preferably helps to assess piracy risks and handle the already available countermeasures as well as future anti-piracy products.

Regarding the project plan, a quick research in the institute’s database showed that no such project had thus far been handled. Therefore, for the further project planning, no reference model was available. A new implementation model had to be developed.

Detection of alternative Solutions

The alternatives were to either propose the development of a high tech solution or a design support. The proposal could be prepared within one of IPEK’s research groups or by a team composed from different groups. If no suitable constellation could be found, a further alternative would have been not to participate at all and ignore the call for proposals.

Selection of Solutions

Comparing IPEK’s resources and expertise within the different research groups, it was decided that developing RFID technologies, printing technologies and so on would not fit well into IPEK’s research portfolio. However, developing a design support against product piracy would ideally suit the Design Methods and Management Group. Therefore, researchers working in this group should prepare such a proposal and conduct the research.

Analysis of Consequences

This determined the focus for the initial system of objectives. Some kind of design support would be developed rather than a product to reduce piracy risks. Taking into account all running research projects within the group at the time it was decided that if the proposal was accepted and the actual research project was kicked-off, a new PhD-scientist would be employed.

Deciding and Implementing

The decision was made, and a team of two experienced PhD scientists was assigned the task to find potential partners and prepare a proposal within six weeks due to the approaching deadline of the call. The team manager of the design methods group started to look for prospective future team members well fitted for the project.

In parallel, they planned the necessary activities once the project started, and derived a Gantt chart which was also a deliverable of the research proposal. In the terminology of the framework presented here, they were creating and continuously updating the implementation model with the potential industrial partners.

Recapitulation and Learning

Lessons learned about the process of writing the proposal, including the gathering of relevant information and how to find potential partners were documented and presented to younger PhDs from all research groups for future calls.

6.2 Empirical Research and Analysis of real-world Processes

The proposal was elected by a jury and funding was assigned to IPEK to implement the research project. A young PhD was hired to ensure sufficient resources for the project. The participants finally were four companies and two research groups. The second research group (apart from IPEK's design methods group) was an economics group specialized in risk assessment.

Situation Analysis

The two research groups covered two research fields: *Design support development* from a product engineering perspective, on the one hand, and *risk assessment* and *monetary quantification of risks* on the other hand. The participating companies covered a range of different engineering branches, company sizes and piracy problems:

- A manufacturer of production lines for medical devices, about to enter the market with a new product, afraid to be copied as an innovation leader.
- A manufacturer for packaging machinery that had already been copied and in consequence, lost all turnovers in a foreign market.
- An automotive engineering and consulting company seeking ways to engineer copy-proof drive train components.[498]
- A plasma welding and cutting equipment manufacturer, losing turnover in their consumables.

Detailed information about the piracy-related problems at the participating companies was collected as well as literature about piracy for review.

Problem Containment

The setup of the project team allowed taking into account different perspectives regarding *branch*, *company size* and *type of* potential *piracy*. The design support development could be split into two equally important portions:

---

[498] Later in the project, this partner resigned from the project as it did no longer have sufficient resources to participate. The partner was replaced by an engineering consultant specialized in moderating creativity-workshops and quality management support such as FMEA (see also chapter 6.3)

  - Support addressing the generation of ideas on how to design products that are harder to imitate.
  - Support that deals with monetary quantification of potential piracy risks.

From IPEK's perspective[499], the empirical research, therefore, needed to reveal:

- What methods and approaches are documented in literature?
- What types of products are generally more likely to be copied?
- What are successful approaches to protect products?
- What methods are known to identify countermeasures?
- What methods do companies use to identify countermeasures?
- What are the specifics of the participating companies?

Detection of alternative Solutions

Alternatives for the setup of the empirical research were discussed. Going to the pool of methods summarized in chapter 4 of this thesis, *potentially useful* methods to gather the desired information would have been: Ethnography, Case Study Research, Retrospective Protocol, Interviews, and Surveys

Selection of Solutions

Literature review was selected, to identify other companies' successful approaches on how to protect themselves and integrated into a case study approach.

The possibility to conduct an *ethnographic study* was *discarded*, as the resources, the project was equipped with did not allow placing scientists within all four companies. The project funding was limited to 30 months. Hence a sequential placement within the companies would not have been practical either. It would not have left enough time for thorough *ethnographic* observations. However, three visits with each company throughout the project were planned. Each visit should last at least one week, allowing for a *case study* approach instead.

Retrospective protocols by the engineers of the participating companies would have been useful to identify successful techniques. However, those were impossible to acquire, since the participating companies had not dedicated any specific design activities leading to piracy protection, yet. When the project started, they were still looking for ways to protect themselves. *In-depth interviews* with representatives from different departments within the participating companies were planned as part of the case study. An Internet based survey to include insights from further companies was thought of, but the idea was discarded. Preliminary interviews with representatives from companies from outside the project team showed that they were reluctant to reveal how they protected their products or even to admit being "threatened" by potential piracy. Smaller surveys within the project team were incorporated into the case study research.

---

[499] Following, I will concentrate on IPEK's perspective on the project and leave out the monetary assessment. However, the framework could well be used to describe the partner's procedure in the project.

Table 61 - Selection of methods for the project

Methods Integrated into the project	Discarded methods
▪ Literature review ▪ Case Study approach ▪ Interviews ▪ Small surveys within the project team	▪ Ethnography ▪ Retrospective Protocol ▪ comprehensive surveys with external companies

Analysis of Consequences

The analysis of consequences with a research setup as described in the selection of alternative solutions revealed that from an empirical point of view, only data of exploratory character with very limited *statistical relevance* would be gathered. Generalization of the collected data should be done carefully or even avoided. Consequently, the design support could be developed with a strong focus on relevance and applicability. Generalizability on the other hand, would need to be ensured in later steps, i.e. by testing it with companies outside the research team in further implementation studies. If the literature review revealed that such exploratory studies exist, and the data could be transferred to the companies in the project team, an altered research setup would be advisable.

Deciding and Implementing

Very little research on product piracy and its countermeasures was documented in literature. Although product piracy had been a long known problem for industrialized nations, the call for proposals was the first major research action against product piracy, consequently, not many prior results were available. It was decided that the limitations noted in the analysis of consequences were natural for such an early stage of a research area and had to be accepted.

The researchers from the two participating research groups made their first visits with the participating companies. They conducted in-depth interviews with four to nine representatives in each of the companies. Each interview took between 60 and 90 minutes. The interviews were analysed, and preliminary results were presented to all participants at the end of each week to eliminate possible misinterpretations.

The findings from the interviews and literature review contained detailed information about the characteristics of the products targeted by product-pirates. Resulting estimated losses in turnovers along with the design processes applied in the development of the corresponding products were assessed. Finally, possibilities on how to integrate additional anti-piracy design steps into those processes were analysed.

Recapitulation and Learning

What had been learned about the interrelation between the research topic, the setup of the project team and the resulting consequences was documented and discussed with other research groups to be taken into account for future calls, especially those by BMBF. The decision to involve companies from different branches and different sizes determined the limited statistical relevance of the data collected. At the same time this ensured a broad spectrum of aspects to be taken into account. This avoided the design support to be tailored too specifically to a certain branch, type of product or company-size.

6.3 Embodiment of Design Support

The core objective of the project was to develop a design support that would help companies find solutions that make their products harder to copy. The support should be integrated into existing design processes and allow for an analysis of the cost to protection ratio. Protection at all costs had to be strictly avoided.

Situation Analysis

The researchers developing the support had become familiar with the participating companies, their problems at hand and the characteristics of their design departments and design processes. All these findings could now be taken into account, and a first draft of a design support could be suggested.

Problem Containment

A literature review revealed that only little preliminary work on the topic had been published, so the field was wide open and the system of objectives needed to be described more precisely. At this point, a checklist as presented in

Worksheet 3 and

Worksheet 4 would have been helpful, both for the decision-making process and the first draft of the written proposal. Applying the worksheets for the Meta System of Objectives from the perspective of the Design Methods and Management group could have turned out as follows:

Table 62 - Refining the system of objectives using the provided worksheet with meta-targets

Questions for self-check :	**Answer in the context of anti-piracy support**
General questions	
In which aspects is the support new?	Preliminary Studies have revealed a series of security systems in terms of concrete products.[500] These have been listed in tables and databases.[501] No support could be identified that systematically helps to choose the right countermeasures (in terms of effectiveness and cost efficiency) and integrate them in the design process.
Which existing methods does it compete with/replace?	
Why does its developer believe it will be more successful?	
To what extend can improvements be excepted if the method is applied?	Reduction of cost for ineffective countermeasures or unnecessarily expansive piracy protection as well as time-efficient development because the countermeasures are integrated in the product development instead of trying to add them after the product is already developed.
What is the result one may expect/desire when applying the support?	A product that is harder to counterfeit as well as an improved estimate of the likelihood of the product being copied.
For which types of problems is the support intended?	Designing products that carry the risk of being copied by illegitimate competitors.
For which activities is the support intended?	Mainly for modelling of principle solution and embodiment. Partly, for idea detection.
To what extend is the support heuristic / algorithmic?	Unclear, however, it should leave room for creative solutions and only be algorithmic in reoccurring activities.
Who is to benefit from the support?	Engineering teams responsible for the development of a new product and decision makers who need to decide on the budget for countermeasures in a new product.
Interaction - Is the support supposed to improve:	
the speed of communication?	-
the effectiveness of communication?	0
designer's competent and objective presentation and discussion of their ideas?	-
help in reaching agreements?	0
Planning & organization - Is the support supposed to:	
help in planning, organizing and controlling projects or processes?	-
support analysis of the process?	-
ensure sustainability of actions and measures?	-
support individual time- and project-management?	-
help prioritize work quotas?	-
help improve processes?	-
help reduce iterative loops?	++
help visualize existing knowledge?	+
help reach targeted cost?	+
help reach targeted deadlines?	-
save time and/or cost?	++
help regarding technical and organizational decisions?	++
support reaching customer- and goal-oriented decisions?	0
support in accessing linked information?	-

<table>
<tr><td colspan="3">suit human problem solving behavior?</td><td>+</td></tr>
<tr><td colspan="3">**Documentation - Is the support supposed to:**</td><td></td></tr>
<tr><td colspan="3">help with documentation?</td><td>0</td></tr>
<tr><td colspan="3">include information about other methods/ processes it can ideally be combined with?</td><td>+</td></tr>
<tr><td colspan="3">belong to a larger/holistic framework of methods it can be combined with?</td><td>-</td></tr>
<tr><td colspan="3">include information about the intended industrial sector/ type(s) of product it was developed for?</td><td>+</td></tr>
<tr><td colspan="3">come with hints and advice for its proper application?</td><td>++</td></tr>
<tr><td colspan="3">come with a description of the expected benefit of its usage?</td><td>++</td></tr>
<tr><td colspan="3">focus on practicability?</td><td>++</td></tr>
<tr><td colspan="3">Inform about required knowledge/expertise?</td><td>+</td></tr>
<tr><td colspan="3">usage of identical and homogenous terms</td><td>-</td></tr>
<tr><td>++: Very important</td><td>+: Important</td><td>0: Neutral, nice to have but not important</td><td>-: unimportant</td></tr>
</table>

Initial System of Objectives

With the above checklist, a system of objectives can be directly formulated as a result of the problem containment process. In the example of the anti-piracy support project, the refined system of objectives was documented as follows:[502]

Preliminary Studies have revealed a series of security systems in terms of concrete products. These have been listed in tables and databases. No support could be identified that systematically helps to choose the right countermeasures in terms of effectiveness and cost efficiency and integrate them in the design process. The support's goal is to reduce the cost for ineffective countermeasures or unnecessarily expansive piracy protection and at the same time protect small and medium enterprises from loss of turnover due to product piracy. The support will increase time-efficiency in development projects in need of anti-piracy-measures. Implementing the countermeasures is being integrated in the activities of modelling principle solution and embodiment within the product development process rather than adding protection after the product has already been developed.

Since no "recipe" for piracy-safe products exists, an heuristic approach that stimulates the designers' creativity and leaves room for their individual problem solving style, and ideas is aimed for. Future users shall be engineering design teams and decision makers who need to decide on the budget for and realization of countermeasures in new products. General aspects to be taken into account for any support development are listed in the table below.

---

[500] e.g. Wildemann et al. (2007);

[501] e.g. http://www.produktpiraterie.org/out.php?idart=17, 5/5/13, 15:04

[502] A rough initial system of objectives was available right after interpreting the call for proposals.

Table 63 - General aspects for Anti-piracy support's system of objectives[503]

Anti-piracy support should	**Anti-piracy support should not**
▪ be simple ▪ be flexibly applicable ▪ focus on the output ▪ be better integrated in the process ▪ be adaptable to the wishes of the users	▪ need too much effort ▪ carry too much “theoretical ballast” ▪ be presented with lack of preparation and support for its application of the methods ▪ miss computer assistance if available ▪ use heterogeneous terms ▪ mix / assume different paradigms ▪ be too theoretical ▪ have inconsistent representation ▪ have an unnecessary high level of complexity ▪ have an unnecessary high level of abstraction

Detection of Alternative Solutions

After the system of objectives had been verbalized, the next step was to identify possible alternative concepts for the support and analyze how promising the different approaches would fulfill the targets documented in the system of objectives. For the sake of readability, not all ideas ever produced during the project will be described.

Failure Mode and Risk Analysis

One idea was to develop a support using the logic of *Failure Mode and Risk Analysis (FMEA),* a well-established method applied in quality management. It systematically searches for possible ways of a product’s or its subsystems’ failures and quantifies their risk. This is done by assessing the likelihood of the failure to occur, the severity of its consequences, and the possibility of detecting the advancing failure before it occurs in all its consequences.[504] Transferring the logic to the context of product piracy, possible “failures” would be the occurrence of product piracy, or counterfeits of parts of the product, severity would be represented by potential loss of turnovers, and detecting the 'failure', i.e. the occurrence of piracy would be treated as in a regular FMEA.

Business War Gaming

Business War Gaming is an approach that originates from strategic military consulting. War games in a military context can be traced back many hundred years. The term describes any type of war simulation, used to determine alternative moves and strategies for both one’s own perspective as well as the enemy’s perspective. The goal is to uncover moves one has not thought of before and have an action-reaction plan ready at hand for possibly all alternative actions of the enemy in order to reduce reaction time and be prepared for the most likely scenarios.

---

[503] Cp. Worksheet 4: Meta Targets for the System of Objectives. It includes original sources for the meta targets mentioned in literature

[504] Comprehensive literature is available on FMEA and many of its variations, e.g. Stamatis (2003) or McDermott et al. (2008) amongst many.

In a business context, the changes of market conditions, including potential actions of one's competitors such as price reduction are simulated in a "game" of two or more.[505] Transferring this concept to product piracy, possible moves of illegitimate competitors could be simulated and necessary reactions thought through. The simulation would then be a game with two or more teams that attack and defend each other with different actions.

Game Theory

More theoretical than War Gaming is "Game Theory" sometimes also referred to as "Interactive decision making" or "The mathematical theory of games." As a mathematical model, it was originally developed and proved by VON NEUMANN & MORGENSTERN in the 1940s.[506] Many people, however, associate a different name with the theory of games: JOHN NASH, who won the Nobel Price together with JOHN HARSANYI and REINHARD SELTEN in 1994.[507] Game Theory includes theorems that allow modeling and hence predicting win-win, lose-lose win-lose situations. Modelling games between a company and a product pirate, might allow predicting settings in which piracy is unfruitful for the potential pirate. These could be integrated in a company's market strategy.

Database Approach

Another idea was to collect and classify countermeasures in a type of database and categorize the constraints and settings in which they can be or have been successfully applied.

Selection of Solutions

For the selection of solutions, it needs to be determined, which concept is most promising regarding the fulfillment of the system of objectives. For this cause, we revisit Table 62 and Table 63 which were produced in problem containment. The suitability of the different ideas is assessed against the system of objectives as shown in Table 64.

The analysis shows that both War Gaming and FMEA based anti-piracy support would show the highest potential to fulfill the important objectives, with FMEA showing slight advantages if fewer relevant objectives (marked as neutral) are also considered.

---

[505] For literature on business war gaming refer to Bracken (2001); Herman / Frost (2008); Gilad (2009)

[506] Neumann / Morgenstern (1944)

[507] The Nobel Price in Economics was assigned for Game Theory related works also in 1996, 2005, 2007 and 2012. A comprehensive summary of the development of Game Theory can be found at: http://www.stratgaming.com/game_theory.html, 9/5/2013, 15:15
For further reading: Neumann / Morgenstern (1944); Nash (1950); Binmore / Blackwell (1991); Osborne / Rubinstein (1994); Nash (1996)

Table 64 - Comparison between alternative concepts

		FMEA	War Gaming	Game Theory	Database
Interaction - Is the support supposed to improve:		**How well does the support improve:**			
the speed of communication?	-	+	-	-	-
the effectiveness of communication?	0	+	-	-	-
designer's competent and objective presentation and discussion of their ideas?	-	0	0	0	0
help in reaching agreements?	0	+	0	+	0
Planning & organization - Is the support supposed to:		**How well does the support improve:**			
help in planning, organizing and controlling projects or processes?	-	+	0	0	0
support analysis of the process?	-	+	+	0	0
ensure sustainability of actions and measures?	-	+	0	0	0
support individual time- and project-management?	-	0	0	0	0
help prioritize work quotas?	-	+	0	0	0
help improve processes?	-	0	0	0	0
help reduce iterative loops?	++	+	+	+	0
help visualize existing knowledge?	+	0	+	0	0
help reach targeted cost?	+	+	0	+	0
help reach targeted deadlines?	-	+	0	+	0
save time and/or cost?	++	+	+	+	+
help regarding technical and organizational decisions?	++	+	+	+	+
support reaching customer- and goal-oriented decisions?	0	0	0	0	0
support in accessing linked information?	-	0	0	0	+
suit human problem solving behavior?	+	+	+	0	-
Documentation - Is the support supposed to:		**How well does the support improve:**			
help with documentation?	0	+	0	0	+
include information about other methods it can ideally be combined with?	+	0	0	0	0
belong to a larger/holistic framework of methods it can be combined with?	-	+	+	+	+
include information about intended industrial sector/ type(s) of product it was developed for?	+	0	0	0	0
come with hints and advice for its proper application?	++	+	+	+	+
come with a description of the expected benefit of its usage?	++	+	+	+	+

focus on practicability?	++	+	+	-	-
inform about required knowledge/expertise?	+	+	+	+	+
usage of identical and homogenous terms	-	+	+	+	+
Anti-piracy support should					
be simple		-	-	-	+
be flexibly applicable		+	+	-	-
focus on the output		+	+	-	-
be better integrated in the process		+	+	0	+
be adaptable to the wishes of the users		+	+	-	-
Anti-piracy support should not					
need too much effort		0	0	0	0
carry too much “theoretical ballast”		+	+	-	0
be presented with lack of preparation and support for its application of the methods		+	+	+	+
miss computer assistance if available		+	+	+	+
use heterogeneous terms		0	0	0	0
mix / assume different paradigms		+	+	+	+
be too theoretical		0	+	-	+
have inconsistent representation		+	+	+	+
have an unnecessary high level of complexity		+	+	+	+
have an unnecessary high level of abstraction		+	+	0	+
		+: Possibility to fulfill the objective 0: Neutral, will not affect the objective -: Will hinder the fulfillment of the objective			

Analysis of Consequences

For the analysis of consequences, rough versions of War Gaming and an FMEA like approach were drafted and presented to the participating companies in a workshop were both methods were applied for a fictive example. Group interviews were conducted afterwards. The results of which can be summed up as follows:

While a war gaming approach was generally deemed “more fun," consensus was that an FMEA based approach would probably be taken more seriously and have a greater chance for acceptance. It was also credited advantages for it produces a systematic documentation of the possible measures, making it much easier to transfer ideas to other products.

Deciding and Implementing

It was therefore decided that the FMEA based approach would be developed further.

For the implementation, a new partner was included in the project.[508] A small engineering consultant specialized in moderating and monitoring FMEAs for engineering companies. The consultant was interested in developing an anti-piracy support that would allow offering anti-piracy design as a service in their portfolio.

The support was developed over a period of eleven months. It included the development of MS Excel based software tools as well as an alteration of special FMEA software. The application, necessary preparations and resources were documented in a manual on how to conduct the “Piracy Risk and Measure Analysis (PRMA)”, the final title for the support.[509]

Loop Back to Project Planning and Controlling

With the decision, to develop and adapt an FMEA approach, the System of Objectives was concretized. Substituting one of the project’s partners meant, altering the operation system. These changes, consequently, led to major revisions in the project plan. Hence, the core project managers had to revise and adapt the initial implementation model as the system of objectives was concretized and the Operation System was altered, and an updated implementation model was generated.[510]

Recapitulation and Learning

It was i.e. documented that the later introduction of a new partner is quite possible in projects funded by the BMBF, which was not clear for the participating institutions from the beginning. This lesson learned was also communicated throughout all other anti-piracy research projects launched and funded parallel to the here described project.

---

[508] The original partner, the automotive engineering consultant resigned from the project. The change of partners was to some extend a lucky leap of faith. Some scientist claim that such a leap of faith is “unscientific”. In the author’s point of view, it is an excellent example for real design support development projects. Things happen although planned differently. A well-planned project, however, allows for quick, yet thorough analysis and a reaction to such a change of parameters.

[509] It was later revised and published within the final project report: Abele et al. (2010)

[510] Cp. 5.2.4

6.4 Experimental Studies, Evaluation in controlled Environment

The first preliminary experiments had been conducted in the embodiment stages. Their goal was to help decide whether FMEA or War Gaming should serve as the basic concept for the support. An FMEA based approach was found to be advantageous the PRMA was developed. Next, the experimental evaluation had to be tackled.

Situation Analysis

The development of a preliminary version of PRMA had been finished. The experimental study would be conducted with the participants of the research project first, i.e. the industrial partners. If further companies could be motivated to take part in test runs, they would be included. Funding allowed for traveling expenses to extra companies, but the companies could not be reimbursed for their effort.

Problem Containment

The guiding questions for the problem containment were: What should be tested? With whom could it be tested? How extensive could be tested?

What should be tested could easily be looked up in the system of objectives (cp. Table 65) previously developed from

Worksheet 3 and

Worksheet 4. The participants would primarily have to be drawn from the project team. However, this would not allow producing data of statistical relevance in a quantitative approach. Outside companies were therefore approached. They generally claimed to be interested in the results. The support had not yet been established, so no company was willing to participate with a design team of about five engineers for a full day, since the results were unsure. Other companies explained that they did not want to discuss potential piracy risks with “outsiders”. Eventually, none but one company apart from the project partners would agree to participate in a test run, actively.

Detection of alternative Solutions

In order to detect principally possible alternatives on how to evaluate the aspects listed in Table 65, the method profiles in chapter 4 were filtered, for methods suitable for experimental evaluation. The results are listed below. The analysis of the results could be done in a quantitative or qualitative approach, hence determining the way in which the data should be produced.

▪ Evaluation based Task Design ▪ Consensual Assessment Technique ▪ Metric for assessment of design outcomes ▪ Observation of sketching activities ▪ Observation of sketching ▪ Myers-Briggs Type Indicator ▪ Kirton-Adaption-Innovation inventory ▪ Coding Schemes ▪ Interview studies ▪ Survey	▪ Three alternatives were generally possible considering the involvement of participants: ▪ The already participating companies of the research project could server as participants, ▪ further companies could be involved, and ▪ Students could be involved.

Table 65 - Checklist for experimental evaluation

Interaction – Does PRMA improve:	
the effectiveness of communication?	0
help in reaching agreements?	0
Planning & organization – Does PRMA:	
help reduce iterative loops?	++
help visualize existing knowledge?	+
help reach targeted cost?	+
save time and/or cost?	++
help regarding technical and organizational decisions?	++
support reaching customer- and goal-oriented decisions?	0
suit human problem solving behavior?	+
Documentation – Does PRMA improve:	
help with documentation?	0
include information about other methods it can ideally be combined with?	+
include information about intended industrial sector/ type(s) of product it was developed for?	+
come with hints and advice for its proper application?	++
come with a description of the expected benefit of its usage?	++
focus on practicability?	++
inform about required knowledge/expertise?	+
Is PRMA	
Simple?	
flexibly applicable?	
focused on the output ?	
Easy to integrate in the design process?	
adaptable to the wishes of the users?	
Is PRMA design to avoid	
too much effort?	
too much “theoretical ballast”?	
to be presented with lack of preparation and support for its application of the methods?	
missing computer assistance?	
use of heterogeneous terms?	
mixing different paradigms?	
be too theoretical	
have inconsistent representation	
have an unnecessary high level of complexity	
have an unnecessary high level of abstraction	

Selection of Solutions

For the selection of the study-design, advantages and disadvantages of the different aspects needed to be weighed against one another.

The output of the PRMA is eventually supposed to be a product that faces less piracy. Output oriented evaluation is therefore difficult to apply, since the quality and quantity of ideas against piracy alone are not an intended variable. The “symbiosis” between the product itself and feasible ideas integrated into the product to reduce piracy risk are the relevant success factors. Additionally, if PRMA reveals that not taking any measures is economically sensible, than ‘no measure' is the ideal output. Considering the system of objectives, it became apparent that process-oriented evaluation seemed to be the better choice (cp. Table 65 - Checklist for experimental evaluation). The process of applying PRMA should be given more attention than the outputs it produces. Concerning the involvement of participants, it was decided to plan with the project members and in parallel try to acquire some outside companies if possible. In those companies, test runs would be conducted with examples from real, yet finished development projects. So hypothetical PRMA sessions were conducted with products already in the marked under the guiding question: “What could have been designed differently leading to lower piracy risk today?”

The PRMA should be evaluated by the participants in the end of each workshop in a structured group interview with a questionnaire based on Table 65. It should also be evaluated by the moderators – one researcher and on representative from the engineering consultant – from the perspective of applicability of PRMA as a consulting service under the guiding question: “How well does PRMA fit into a consulting portfolio? Is it possible to offer it as a professional service?” Since easy access was available, a parallel yet smaller study with student designers was decided on. The goal was to also do some output-oriented evaluation of PRMA. Student design teams should therefore compare common techniques for idea creation (Brainstorming, 6-3-5, …) to PRMA. The results should be judged through consensual assessment technique.

Analysis of Consequences

The process-oriented evaluation selected was qualitative and would allow gathering information from both the users’ perspective and a moderator’s perspective. The moderator and scientist would be able to compare the different experiments in a cross-sectional perspective. However, no directly transferable data would be produced that would allow any statements about the degree to which e.g. certain branches found PRMA useful in comparison, since this would demand quantitative data gathered in longitudinal studies.

The output oriented evaluation principally showed the potential for quantitative data analysis. However, the amount of data to be gathered would hardly allow for statistically relevant findings. They would not be transferable to industrial settings either and should be taken as an indicator for PRMA’s potential.

Deciding and Implementing

A workshop-based approach was designed. The representative of the engineering consultant in the project would act as a moderator, and the researcher as his assistant. After the sessions, retrospective protocols from the perspective of the consultant were written for later analysis and group interviews with the participants were also conducted to gather information about their perception of PRMA's applicability and consensual assessment was conducted, discussing the quality and feasibility of the ideas produced through the PRMA session.

The setup was discussed with all project members and agreed upon. Dates were set for the different test runs, and companies outside the project team were contacted and asked if they would participate in a test run. The follow-up interviews and protocols were prepared with the checklist shown in Table 65. The results of the retrospective protocols and the interviews conducted after the PRMA-tests, were summed up and presented to the project team.

The output-oriented evaluation with students was set up as follows: A group of students from a product development management course were trained in different methods for creating ideas. PRMA was included in the two-week intensive training course. All methods taught were applied in example projects followed by a reflection discussion session in which they should argue which methods they would use in certain situations. The majority agreed that in order to develop ideas against product piracy, PRMA was more suitable than the more general idea generation methods like 6-3-5 or brainstorming.

A similar test was designed with a group of 42 mechanical engineering students in their senior year. They were given an example product and told to apply brainstorming and 6-3-5 technique to generate ideas to reduce piracy risks. Later assessment of the ideas by the researchers indicated that although brainstorming and similar methods produced more ideas, these methods eventually produced fewer feasible ideas.

Recapitulation and Learning

6.5 Implementation Studies

After the test runs in the hypothetical PRMA workshops, the guiding question now was: "How can the PRMA be integrated into the ongoing processes of the participating companies?" In a next step, the question would be: "How can the PRMA be integrated into the running operation systems of the further companies?"

This must be done, assuring PRMA acts as expected, i.e. that it does not absorb more effort than was planned for and promised and also meaning that it does what it is supposed to do, generating the output promised to the users. The challenge is to find and avoid aspects that lead to different outputs than were observed in experimental studies, although the researchers that have developed the support are not necessarily present any more.

Situation Analysis

The PRMA had been developed, presented to the project partners and suggestions for improvement had been collected. The changes were implemented and controlled test runs with follow-up interviews had been conducted. Overall, the project partners found that the developed support fulfilled their expectations, and it was recommended by the users to proceed with the development. What was also apparent from the test runs that PRMA could only be conducted successfully with a dedicated moderator who preferably had some experience with similar methods, such as FMEA.

A detailed description on how and when to prepare and conduct a PRMA workshop, therefore, was documented and provided to the project partners.

A problem resulted from the fact that no further companies could be convinced to test the method. The impact of the 2007 banking crises had driven almost every company into short-time working hours, making it impossible for them to participate – especially without funding. Anti-piracy activities had lost importance for those as they were coping with the crisis.

Problem Containment

The challenges for the final implementation into practice had two aspects:

Firstly, it had to be verified that PRMA works as intended. Secondly, the PRMA had to be connected to the “Cost Analysis Support” that had been developed in parallel by the economics and risk assessment group in the project. The goal of the project had been from the very beginning to provide both a way to selected proper countermeasures against piracy and at the same time have a way to base the decision for or against a certain countermeasure on an economic basis (cp. 6.1).

Detection of alternative Solutions

The researchers now had to come up with different setups for further implementation into industrial practice. One option would have been to keep looking for companies that would engage in the implementation study and proceed according to the original plan. Since this had been unsuccessful up to then, it would eventually have meant moving the project deadline.

The other option was to stick to the deadlines and adapt the original plan. This meant to try to ensure the possibility for successful implementation without further test runs in companies that had not yet been exposed to PRMA.

Different alternatives were collected for such a change of plans. An attempt for a self-explanatory version of PRMA was one of the options, or in other words: Making the application of the support really easy and extinguish any ambiguity in its documentation. Another option would have been to finalize the support in a way that it uses a set of trained moderators who always accompany the application hence superseding the self-explanatory character. A compromise of the previous options would have been to establish training for those who want to apply the method.

Selection of Solutions

Waiting for further companies contradicted with the project deadline, since the project was coordinated to run parallel with nine other anti-piracy projects. Extending the deadline was impossible. The goal to develop a self-explanatory design support was – while quite desirable –very challenging. In scientific terms, this meant: Achieve a very high inter-rater reliability (cp. 4.3.4.3)! Different users applying the support need to produce comparable results under equal conditions. If a support can be shown to have a high inter-rater reliability, the procedure is described without (or at least very little) ambiguity. Transferring the statistical procedures for the assessment of test reliability to the problem at hand ("Is PRMA self-explanatory and have its developers documented it without ambiguity?"), it became clear that this could only be done with adequate data coming from a sufficiently large number of probands. This brought the PRMA developers back to the original problem: their limited resources for the test runs. A series of tests for inter-operator-reliability assessment was discarded. Instead, interviews were conducted with the project's participants. The goal was to determine whether they felt confident to apply the method without the researchers present. These were mostly academics who had followed the development of the support. If they would not apply the method without a dedicated and trained moderator present, no-one else would.

If inter-operator reliability could not be shown, the alternative was to come up with a concept independent from that form of reliability. Implementing a moderator or a defined group of moderators makes it unnecessary to prove inter-operator reliability and shifts the focus on intra-operator reliability. This means that one and the same moderator should act comparably in comparable situations. The given constraints were very suitable for such a study. The engineering consultant in the team had employees with a strong background in FMEA moderation and was interested in offering PRMA as a consulting service. Repeated tests could be conducted using both standardized training problems as well as realistic problems provided within the project. This option was eventually preferred over the last option to develop a training-concept, which would have had to be validated for reliability. Again this would have demanded for sufficient data, which was impossible to provide within the scope of the project.

Analysis of Consequences

The decision to provide the method with a moderator had two main sets of consequences. On the one hand, it made the finalization of the PRMA much easier as moderators and developers could directly communicate. Inter-operator reliability was unnecessary, and the acceptance of PRMA by the actual users could be increased since the necessary preparations for PRMA workshops would be done as a service by the engineering consultant, reducing the barrier for the users. However, it meant running the risk of limiting acceptance on the side of decision makers. If a method is only available as a paid consulting service, it might put decision makers off assuming they would prefer to develop their solutions completely in-house and save the money spent on consultants. A series of short interviews was conducted to clarify this aspect leading to the conclusion that managers much rather pay a trained moderator who knows how to lead such a workshop efficiently than pay their employees to learn a new method and possibly waste a lot of time conducting the workshops due to lack of experience. Overall improvement of acceptance could therefore be expected if a consulting-type implementation was chosen. Nevertheless, if only a few trained moderators could conduct PRMA workshops, its application was going to be limited regionally to the catchment area of the moderators, hence limiting acceptance on a global perspective.

Deciding and Implementing

Although the regional limitations posed a severe problem, it was decided to proceed, implementing the moderator concept. However, it was decided to improve the documentation and address it to a target group familiar with FMEA, preferably professional FMEA moderators in other regions. If future resources allowed to do so, and if demand was detected in industry, a training concept for larger companies could still be developed. Additionally, an MS-Excel based software tool was developed. On demand, the moderators would use it and teach the participants of the PRMA workshop on the project, making the users autonomous from the engineering consultant.

Recapitulation and Learning

The lessons learned from the implementation study was mainly that company's commitment to such studies, although promised before the beginning of a project is something dangerous to rely on and must be planned very intensively. One might go as far as to claim that the planned resources should have been split equally between the embodiment, the implementation studies, while the experimental study turned out to have been easier than expected.

Another interesting finding that can be transferred to future design support development projects was the fact that companies’ decision makers actually welcomed the concept of delivering a new design support together with a moderator. The cost for paying a moderator was estimated to be much smaller than the prospective cost for wasted time spent on training employees in a method that might only be practiced a few times, as well as potential cost for inefficient workshops, applying the method.

6.6 Transfer Studies

The final step in the development of a design support in the here-presented framework is to provide a transfer concept addressing those who want to learn how to apply the support.

Situation Analysis

The PRMA had been designed in a way that it can be conducted by one single moderator leading a team of designers through a PRMA Workshop. IT had been based on FMEA. Both the experimental study as well as the implementation study had produced feedback from potential users. PRMA was acknowledged as a useful way to incorporate anti-piracy measures into the design process of new products. This meant that it would not be applied on a daily basis but only in early stages of product engineering projects.

Problem Containment

Due to its rather specialized character, it was found unpractical to teach it to engineering students in general. However, most engineering students are exposed to the concept of FMEA during their education. As PRMA is based on FMEA, it was found safe to assume that any designer who had been part of one or more FMEA workshops could also claim to have some experience applicable to PRMA. The moderator concept made it unnecessary to train every participant. A quick introduction directly before starting the PRMA turned out to be sufficient. The main task for delivering a transfer concept therefore was reduced to addressing it to potential moderators, not the users. Those moderators were assumed to be professional FMEA moderators. Additionally, decision makers in companies had to be informed that an FMEA-like method was available that could help protect their companies' products against product piracy.

Detection of alternative Solutions

The alternatives for training future moderators were to offer instruction courses for potential moderators, to train them “on the project," or to encourage self-teaching with sufficient reading material.

Selection of Solutions

Offering training lectures at the university was found to be impractical since a university is not necessarily set up to promote commercial training. Providing such resources would make it too expensive. The engineering consultant could have offered the trainings. However, they too were specialized in providing services and did not have experience in training others. Therefore, the PRMA design team approached one of the leading FMEA software and database specialists on the market. Their business model is to train moderators with their software and offer licenses, mainly to engineering consultants who then offer their FMEA service and database to customers. The company was found to be a suitable partner by the PRMA team. Nevertheless, the software provider only offered to allow a presentation of the PRMA at one of their yearly exhibitions and was not interested in designing an extra training, neither in cooperation with the PRMA developers, nor on their own.

Providing reading material was eventually found to be a feasible solution. Since a comprehensive documentation had to be prepared in any event, it could as well be written in a way that it addressed moderators rather than users/participants. The engineering consultant on the team had several employees who worked as moderators and could review the material from a target group's perspective. For training future moderators within their companies, a case-based training was approached. For this, however, the PRMA developers could not rely on a commercial FMEA-software. It was very unlikely that many potential users would possess the relevant software license or would be willing to purchase it just for conducting PRMA. The concept for such a training course was to conduct one PRMA workshop with a team of engineers at the customer and have the customer's future PRMA moderator present in the team. He / she could observe and learn and have the chance to ask questions.

Analysis of Consequences

Providing reading material, no active advertisement for the design support could be made as would have been possible in trainings. The responsibility of promoting the newly developed design support to the users (not the moderators) would be shifted towards the moderators. However, at the same time it was found likely that a (financially) motivated moderator would advertise for the method and increase the chance of dissemination. Furthermore, the credibility of the support from the perspective of engineering companies can be increased if its application is promoted and supported by professional moderators/consultants. For the case-based trainings, commonly available tools needed to be provided and some extra time for explaining, training and reviewing with the future moderator would need to be planned.

Deciding and Implementing

It was decided to focus on the provision of reading material for interested moderators, and a comprehensive documentation was prepared. Additionally, an MS Excel-based software tool was derived from an existing FMEA Excel-Sheet. This provided the option to offer trainings on a concrete anti-piracy project at companies.

As the analysis of consequences showed that the advertisement / promotion of PRMA was somewhat handed to potential moderators, it was decided to enforce the promotion of the concept at exhibitions, and an anti-piracy working committee was established in cooperation with the "Association of German Engineers"[511], to support dissemination of PRMA.

Recapitulation and Learning

Transfer can be easily underestimated, especially for a university-institute used to teaching students. A method which is rather specialized, such as the PRMA, can of course be integrated into the curriculum of the design methods course. There might already be a chapter on FMEA. Nonetheless, it is to be expected that only a minority of the students will be in need of such a design method in their future career, so new target groups must be approached. In this case, the ideal target group was found to be professional engineering consultants that specialize in moderating FMEA workshops. University lectures do not work on such a target group, since the researchers don't

---

[511] German: Verband Deutscher Ingenieure (VDI)

have direct access to them as they do to students. However, opposite to students, consultants have a monetary interest in learning, adapting and eventually offering special methods – provided they are relevant for practitioners. This offers the chance of self-propelled dissemination in practice – an ideal constellation for the developers of a design support.

6.7 Analysis of Utilization

Upon finalization of the PRMA, it was marketed in three ways:

- The engineering consultant offered and promoted it as a unique service for their customers.
- The two scientific partners in the project joined to offer a combined consulting concept consisting of PRMA and cost-to-benefit assessment.
- The anti-piracy working committee established in cooperation with the “Association of German Engineers” planned to hold regular meetings with speakers on the topic and discussion groups to exchange experiences with successful countermeasures.

Situation Analysis

Situation Analysis of the PRMA’s utilization revealed that the cost for marketing and advertisement of the PRMA exceeded the money the engineering consultant made with the PRMA. However, it led to further acquisition of regular FMEA projects as an unintended indirect effect. The combined workshops, including PRMA and cost to benefit analysis were not once requested. The working committee of the Association of German Engineers kept planning meetings every six months of which about every second meeting got canceled as the number of participants was too small.

Problem Containment

Discussions with potential users revealed that the demand from industry that had been detected prior to the project had in the meantime declined. A worldwide financial crisis had led restructuring within companies and rearrangement of priorities. Whether this explains the effect, or if the demand had been somewhat over exaggerated remains unclear. What has become obvious though is the PRMA exceeds the effort, companies find reasonable for anti-piracy measures, although during the project, interviews had suggested otherwise.

Detection of alternative Solutions

The options were to either invest further effort and improve PRMA in order to make it more efficient, or to accept its limitations.

Selection of Solutions

Potential users were confronted with the discrepancy between the initially formulated demand from industry and the disappointing utilization of the support. The results showed that industry was hoping for much simpler methods to fight product piracy.

The necessary reductions concerning the effort when applying the PRMA were found to be unrealistic. No further funding was available, and no-one from the original project team believed that an improved, easier to apply PRMA was going to be worth the effort.

Analysis of Consequences

If no improved version of the PRMA would be developed, it was only going to be a matter of time until it completely vanished, since the researchers who developed it would eventually move on to different research questions.

Deciding and Implementing

Although discarding the method was – from a design support development perspective – a setback, it was decided to end the development of PRMA and invest the manpower in more promising research.

Recapitulation and Learning

What was learned from this setback was that an identified demand for a design support can also be the product of trends. Around the time of the initial call for proposals, product piracy was a frequent topic in the media. The presence of reports about dangerous counterfeits has drastically declined. In future design support development projects, a once identified demand will be continuously questioned and reassessed, making it one of the key activities in Project Planning and Controlling.

6.8 Summary of the Sample Project

It was shown in a sample project that the framework can be used to retrospectively describe a design support development project, as it provides a detailed structure that allows the summarizing of decisions made along the way. Consequently, the framework developed in this thesis can also be used for planning such research projects. The activities matrix for design support development can even serve as a template for the formulation of research proposals, as it provides a logical structure and standardized vocabulary and supports the selection of appropriate methods, hence also allowing better estimation of necessary resources.

Applying the framework has shown that it is perfectly possible to structure a research project that has the goal of developing design support. The activities provided in the static part of the framework – the activities matrix – were in this example sufficient to model all important steps. This indicates that the activities matrix is comprehensive for research projects up to a comparable degree of complexity. For a comparison, the example project involved two research groups and four companies from different industries. The project itself lasted 30 months. Including the application for funding and final documentation, it lasted for just over three years.

It cannot be concluded from the sample application whether the framework and the activities matrix would be ample if applied in a larger-scale research project such as collaborative research centre.[512] However such large research projects are usually conducted to answer a complex set of questions and investigate completely new fields of technology or knowledge. It is rather questionable whether such large projects will ever be launched with the goal of developing an heuristic design support. Therefore, the author concludes that the activities matrix is sufficiently detailed for the needs of realistic heuristic design support development projects.

Furthermore, the framework proved to be flexible. Although a major change in the boundary conditions occurred during the project, it was possible to document and model all activities without the necessity to rearrange previous steps after one of the project partners had left the project and was replaced by a new partner with very different goals within the project. After this experience, it is safe to assume that the framework will meet the demand for flexibility in other real-life design support development projects, where such unforeseeable changes of some of the boundary conditions are rather common. However, the above-named conclusions are based on a single application of the framework along with the experience the underlying framework iPeM. Only the consequent application of the framework by different researchers in different projects can conclusively reveal whether the framework meets all the demands design support development poses.

---

[512] German: „Sonderforschungsbereich“. A German type of research program that usually includes several research institutes within one, in some cases two, universities.

7 Conclusion and Outlook

Design as a science has produced a large number of methods and processes to support engineering design work. It is a young field of science, and as it is getting more mature, the community keeps providing its own methods and is integrating scientific methods and procedures from other disciplines. Each one of them addresses a different specific problem, and most of them are literally scattered throughout the different publications that have emerged from the design science community. While there are a few more general approaches available, none of them offers a satisfactory way bringing these specific methods together. This lack of methodology is the main justification of this thesis. A new framework was suggested that aims to aid scientists in their effort to produce design support that is relevant, reliable and credible. The approach taken by the author of this work is based on two fundamental theoretical concepts:

- The framework is based on the assumption that, in engineering design research, certain aspects cannot be fully proven. It may be impossible and more important, it is not useful either. Therfore, developing design support is about building a case for credibility.
- Design support development can be regarded as design in itself.

The second assumption is the justification for why the elementary practical construct, the framework, should be based on a model identified in the world of engineering design – ALBERS' integrated Product Engineering Model iPeM. iPeM is itself based on a series of theoretical constructs most important the five hypotheses of product engineering according to ALBERS.[513] Clearly, the framework presented here inherits those theoretical constructs from its "parent" iPeM. The most important of them are:

- The systems engineering approach to describing product engineering. A System of Objectives is transferred into a corresponding System of Objects. This is done by the Operation System.514
- Regarding every single process as individual and unique, meaning that no two engineering processes or design support development projects will exactly be the same.
- Regarding validation as a result of continuous activities throughout the process rather than as isolated activities in the end of a project.

and last,

---

[513] Albers (2010b)
[514] Albers / Meboldt (2007a)See Albers / Braun (2011), Braun et al. (2013a)

- The activity-based approach to modelling design projects rather than sequential process modelling.

For the application of the framework in practice, it is important to keep these theoretical and practical constructs in mind. Apart from that, users should also remember the purpose of the framework. It is a model that has been intended for developing heuristic design support, meaning design support in which human decision making and creativity play an important role. This is also true for the framework itself. It is an abstract guide to design support development and a collection of methods. Applying the framework will demand decision making and reasoning by the user. It does not automate any of the steps necessary to build relevant, reliable and credible design support.

Finally, users must be reminded that this work is the first draft of the framework. The author has made the case that it is important to rely on scientific methods and procedure that have proven useful and valid. This puts this work in a dilemma. It is young and new. Although it is put together from elements that appear as valid in the reviewed literature, this does not automatically make this recombination of the elements valid. Only the application of the framework in several different projects, the scientific discussion about the framework and the elimination of flaws found in future projects can drive this framework towards true validity.

Accomplishments:
This thesis is the first of its kind to draw together and structure such a vast collection of specific methods available to the developers of design support. The collection includes abstract and rather general approaches that can help plan and navigate research projects as well as those very particular methods addressing isolated single aspects of design support development. Where applicable, the methods have been categorized with a standardized profile. The short descriptions and concise information given aim to provide the necessary details a researcher needs in order to compare the methods and decide individually which methods suit his or her research project. The thesis does not claim to provide all information available for the different methods, which would by far exceed the perimeters of a thesis. However, for every method presented here, references are given for further reading. For some of the methods, worksheets have been provided, thus assisting their application.

The second major achievement of this thesis is the framework for design support development. It is a structured process model based on the Integrated Product Engineering Model iPeM. It allows for an activity-based view on design support development, addressing all necessary actions to develop design support that is relevant, reliable and credible. The framework and the methods are linked through the method profiles provided: They contain the information indicating for which activities a method is potentially suitable. Navigating through the framework and weighing the advantages and disadvantages of the potentially useful methods, allows to develop a selection strategy applicable to any design support development project.

In short, this framework provides the following key features for those developing design support, namely:

- It assists in building a case for credibility from the very first step;
- It provides assistance for the initial planning;
- It helps in managing and controlling a research project;
- It provides structure and vocabulary for the documentation of research outcomes and the steps taken to achieve them;
- Within a design support development project, it helps navigate through the process itself, providing information for the researcher necessary to reason and decide on what the next steps should be;
- Finally, it contains a large selection of concrete methods that help in actually conducting the activities previously decided on.

All in all, the operation system of design science has been expanded with the necessary tools to tackle the challenges which design support development poses. At the core of the operation system of design science are the design researchers – who are working on their PhD in many cases. The approaches provided in this thesis complement the *Karlsruhe Education Model for Product Development “KaLeP” In higher Education*, broadening its radius to the education of researchers in design science.

It was shown in a sample project that the framework can be used to retrospectively describe a design support development project, as it provides a detailed structure that allows the summarising of decisions made along the way. Consequently, the framework developed in this thesis can also be used for planning such research projects. The activities matrix for design support development can even serve as a template for the formulation of research proposals, as it provides a logical structure and standardized vocabulary and supports the selection of appropriate methods, hence also allowing better estimation of necessary resources.

In consequence, if the framework is applied and design support development projects are documented accordingly, it will be possible to compare research projects thus opening a whole new field for meta studies on design support development processes. Such meta studies will allow the identification of patterns, which again can be used to derive success factors as well as actions to be avoided.

Limitations

“Doing science” is a profession that can be learned. This thesis is intended to offer an easier start for researchers commencing to get familiar with their profession by reducing the time they spend on finding literature dealing with scientific procedure instead of focusing on their research topic. However, other researchers are very likely to be working on new methods (or will do so in the future) that should be integrated into the collection presented here. Therefore, a PhD's efforts on the topic of research methodology cannot be limited to studying this thesis alone. Furthermore, experience is what makes someone a good professional. Experience requires practice, time, and endurance, which is something this thesis cannot substitute.

The framework was developed within a research team very much influenced by the Karlsruhe Education Model. The author has discussed this work with design scientists from other research institutions and received the feedback that the topic is of high relevance. However, no conclusions can be drawn about the individual acceptance of the framework from this feedback alone. Further analysis is necessary to assess the acceptance and perceived benefit, especially from researchers of other educational "schools".

Finally, the advantages mentioned before that occur from patterns and resulting reference models can only be realized if the framework receives a high level of acceptance in the scientific community. As a prerequisite, a number of researchers actually must plan and document their design support development projects according to the model. Furthermore, those implementation models need to be collected in a database in order to recognize patterns and derive reference models across the community.

The framework is one option offered to the design science community. The author does not claim it to be the only way, but neither is it one of many. It is one of very few. Therefore, other authors should feel encouraged to provide competing alternatives.

Outlook

Further improvement, specifically to the work presented in this thesis could be achieved if additional scientific methods could be integrated into the collection and also categorized by the same logic. Although a vast collection has been presented here, it is in the nature of science that it doesn't stop producing new methods.

Future work on the topic of research methodology for design science could include comparative studies of this framework and other high-level approaches such as e.g. DRM or the Spiral of Applied Research. It would be valuable to assess under which conditions which elements prove to help researchers the most. From those findings, advice can be derived as to which framework to choose or a new, improved framework integrating the most successful elements of the existing frameworks might be developed.

If a sufficiently large number of research projects is planned using the framework, including the dynamic phase model, implementation models, and their corresponding execution models along with the necessary resources (plan and actual) could be stored in databases for analysis. This would allow to derive reference models from typical patterns but also enable the scientific community to conduct a variety of meta-studies dealing with questions such as:

- Which are the success factors that lead to execution models, which meet the initial implementation models?
- Which patterns of design support development projects lead to outcomes with high acceptance?
- Identification of cost to benefit ratios in design support development project comparing the used resources to the results of the projects.
- Comparison between typical approaches from different scientific groups or regional differences.

The design and implementation of a database to collect research project's information across the design science community as a necessary requirement to allow for such meta studies provides enough challenge and interesting questions for a PhD project alone, opening the field for a number of follow-up PhD projects like the meta studies mentioned above.

After all, the overall goal of this thesis was to provide researchers with a framework that helps them develop design support better. It is supposed to support documentation, and communication of the results, and help researchers discuss what they have developed with others. This should lead to better results produced by design science as a whole and eventually increase the acceptance of heuristic methods in engineering design and the credibility of the design science community as a whole.

If this thesis leads to nothing but provoke others to develop a better framework, then I feel I have reached this goal.

Mannheim, 27.9.2013

8 References

Abele et al. (2010) Abele, E.; Albers, Albert; Aurich, J. C.; Günthner, A.: *Wirksamer Schutz gegen Produktpiraterie im Unternehmen*. 1. Aufl., VDMA Verlag (2010)

Abercrombie et al. (1984) Abercrombie, N.; Hills, S.; Turner, B.S.: *Dictionary of Sociology*. Penguin Books (1984)

acatech (2012) acatech: *Faszination Konstruktion - Berufsbild und Tätigkeitsfeld im Wandel: Empfehlungen zur Ausbildung qualifizierter Fachkräfte in Deutschland*. Springer Vieweg (2012)

Albers et al. (2002) Albers, Albert; Burkardt, Norbert; Saak, Marcus: *Gezielte Problemlösung bei der Produktentwicklung mit Hilfe der SPALTEN-Methode*. In: Internationales Wissenschaftliches Kolloquium, Technische Universität Illmenau (2002)

Albers et al. (2004) Albers, Albert; Ohmer, Manfred; Eckert, Claudia: *Engineering Design in a different Way: Cognitive Perspective on the Contact & Channel Model Approach*. In: Proceedings of the 3rd International Conference on Visual and Spatial Reasoning in Design (2004)

Albers et al. (2005) Albers, Albert; Burkardt, Norbert; Meboldt, Mirko; Saak, Marcus: *SPALTEN problem solving methodology in the product development*. In: Proceedings of the 15th International Conference on Engineering Design (ICED’05). Melbourne, Australia (2005)

Albers et al. (2006) Albers, Albert; Burkardt, Norbert; Meboldt, Mirko: *The Karlsruhe Education Model For Product Development KaLeP in higher education*. In: Proceedings of DESIGN 2006. Dubrovnik, Croatia (2006)

Albers et al. (2008a) Albers, Albert; Burkardt, Norbert; Deigendesch, Tobias; Meboldt, Mirko: *Enabling Key Competencies by Educational Project Work Exemplified by Teamwork and Cooperation.* In: EPDE 2008, Barcelona (2008)

Albers et al. (2008b) Albers, Albert; Deigendesch, Tobias; Alink, Thomas: *Support of design engineering activity - The Contact and Channel Model (C&CM) in the context of problem solving and the role of modelling*. In: Proceedings of the DESIGN 2008. Dubrovnik, Croatia (2008)

Albers et al. (2009) Albers, A.; Merkel, P.; Geier, M.; Ott, S.: *Validation of powertrain systems on the example of real and virtual investigations of a dual mass flywheel in the X-in-the-Loop (XiL) environment*. In: Proceedings of the international CTI Symposium, Transmission 2009. Berlin (2009)

Albers (2010a) Albers, Albert: *Using design theories in engineering design*. In: Industrial Tutorial of TMCE 2010. Ancona, Italy (2010)

Albers et al. (2010) ALBERS, ALBERT; BRAUN, ANDREAS; MUSCHIK, SABINE: *Uniqueness and the Multiple Fractal Character of Product Engineering Processes*. In: Modelling for Business Improvement (2010)

Albers (2010b) ALBERS, ALBERT: *Five Hypotheses about Engineering Processes and their Consequences*. In: Proceedings of TMCE 2010. Ancona, Italy (2010)

Albers et al. (2011a) ALBERS, ALBERT; SADOWSKI, EIKE; MARXEN, LEIF: *A new Perspective on Product Engineering Overcoming Sequential Process Models*. In: Birkhofer, Herbert (Hrsg.): The Future of Design Methodology : Springer London (2011)

Albers et al. (2011b) ALBERS, ALBERT; LOHMEYER, QUENTIN; EBEL, BJOERN: *Dimensions of Objectives in interdisciplinary Product Development Projects*. In: Proceedings of the 18th International Conference on Engineering Design (ICED'11). Copenhagen, Denmark (2011)

Albers et al. (2012a) ALBERS, ALBERT; LOHMEYER, QUENTIN; RADIMERSKY, ALINE: *Individuelle und organisatorische Akzeptanz von Methoden des Systems Engineering*. In: Tag des Systems Engineering. Paderborn, Germany (2012)

Albers et al. (2012b) ALBERS, ALBERT; EBEL, BJOERN; LOHMEYER, QUENTIN: *Systems of Objectives in complex Product Development*. In: Proceedings of TMCE 2012. Proceedings of TMCE 2012. Karlsruhe, Germany (2012)

Albers / Alink (2007) ALBERS, ALBERT; ALINK, THOMAS: *Support of design engineering activity for a systematic improvement of products*. In: The Future of Product Development (2007)

Albers / Braun (2011) ALBERS, ALBERT; BRAUN, ANDREAS: *A generalised framework to compass and to support complex product engineering processes*. In: International Journal of Product Development 15 (2011)

Albers / Braun (2012) ALBERS, ALBERT; BRAUN, ANDREAS: *Towards Handling Complexity - Testing the iPeM Process Modelling Approach*. In: Proceedings of TMCE 2012. Karlsruhe, Germany (2012)

Albers / Lohmeyer (2012) ALBERS, ALBERT; LOHMEYER, QUENTIN: *Advanced systems engineering - towards a model-based and human-centered methodology*. In: Proceedings of TMCE 2012. Karlsruhe, Germany (2012)

Albers / Marxen (2012) ALBERS, ALBERT; MARXEN, LEIF: *Produktentstehung - ein mentales Modell*. In: Innovation Excellence. Düsseldorf : Symposion Publishing GmbH (2012)

Albers / Meboldt (2007a) ALBERS, ALBERT; MEBOLDT, MIRKO: *IPEMM-Integrated Product Development Process Management Model, Based on Systems Engineering and Systematic Problem Solving*. In: Proceedings of the 16th International Conference on Engineering Design (ICED'07). Paris, France (2007)

Albers / Meboldt (2007b) ALBERS, ALBERT; MEBOLDT, MIRKO: *SPALTEN Matrix - Product Development Process on the Basis of Systems Engineering and Systematic Problem Solving*. In: The Future of Product Development : Springer Berlin Heidelberg (2007)

Alink (2010) ALINK, THOMAS: *Bedeutung, Darstellung und Formulierung von Funktion für das Lösen von Gestaltungsproblemen mit dem C&C-Ansatz*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2010)

Alink et al. (2011) ALINK, THOMAS; ECKERT, CLAUDIA; RUCKPAUL, ANNE; ALBERS, ALBERT: *Different function breakdowns for one existing product: experimental results*. In: Design Computing and Cognition'10 (2011)

Amabile (1983) AMABILE, T. M.: *The social psychology of creativity: A componential conceptualization.* In: Journal of personality and social psychology 45 (1983), Nr. 2

Amabile (1996a) AMABILE, T. M.: *Creativity in context: Update to" the social psychology of creativity."* Westview press (1996)

Amabile (1996b) AMABILE, T. M.: *Creativity and innovation in organizations*. Harvard Business School (1996)

Amabile (1998) AMABILE, T. M.: *How to kill creativity*. In: Harvard business review 76 (1998), Nr. 5

Anderson (2013) ANDERSON, DAVID R.: *Statistics for Business & Economics, 12th ed.* Cengage Learning (2013)

Anderson / Finn (1996) ANDERSON, T. W.; FINN, JEREMY D.: *The New Statistical Analysis of Data*. Corrected. Aufl., Springer (1996)

Andreasen (2010) ANDREASEN, MOGENS MYRUP: *Design Research Consolidation as a Design Society Crusade*. In: Proceedings of CEPHAD 2010 Conference. Copenhagen, Denmark (2010)

Archer (2004) ARCHER, BRUCE: *Transcript of DRS lifetime achievement award acceptance speech*. (2004)

Argyris (1980) ARGYRIS, C.: *Inner contradictions of rigorous research*. Academic Press New York (1980)

Atkinson et al. (2000) ATKINSON, PAUL; BAUER, MARTIN W.; GASKELL, GEORGE: *Qualitative Researching with Text, Image and Sound: A Practical Handbook for Social Research*. SAGE (2000)

Atkinson / Hammersley (1994) ATKINSON, P.; HAMMERSLEY, M.: *Ethnography and participant observation*. In: Handbook of qualitative research 1 (1994)

Atteslander (2008) ATTESLANDER, PETER: *Methoden der empirischen Sozialforschung*. 12. Aufl., Berlin : Schmidt Erich Verlag (2008)

Atteslander / Kneubühler (1982) ATTESLANDER, PETER; KNEUBÜHLER, HANS-ULRICH: *Verzerrungen in Interviews. Zu einer Fehlertheorie der Befragung*. VS Verlag für Sozialwissenschaften (1982)

Audia / Goncalo (2007) AUDIA, P. G.; GONCALO, J. A.: *Past success and creativity over time: A study of inventors in the hard disk drive industry*. In: Management Science 53 (2007), Nr. 1

Badke-Schaub et al. (2011) BADKE-SCHAUB, P.; DAALHUIZEN, J.; ROOZENBURG, N.: *Towards a Designer-Centred Methodology: Descriptive Considerations and Prescriptive Reflections*. In: Birkhofer, H. (Hrsg.): The Future of Design Methodology : Springer London (2011)

Baer (1994a) BAER, JOHN: *Why You Shouldn't Trust Creativity Tests.* In: Educational Leadership 51 (1994), Nr. 4

Baer (1994b) BAER, JOHN: *Why You Still Shouldn't Trust Creativity Tests.* In: Educational Leadership 52 (1994), Nr. 2

Baer (1994c) BAER, J.: *Divergent thinking is not a general trait: A multidomain training experiment*. In: Creativity Research Journal 7 (1994), Nr. 1

Baer (1994d) BAER, J.: *Performance assessments of creativity: Do they have long-term stability?* In: Roeper Review 17 (1994), Nr. 1

Baer (1996) BAER, J.: *The effects of task-specific divergent-thinking training*. In: The Journal of Creative Behavior 30 (1996), Nr. 3

Baer et al. (2004) BAER, J.; KAUFMAN, J. C.; GENTILE, C. A.: *Extension of the consensual assessment technique to nonparallel creative products*. In: Creativity research journal 16 (2004), Nr. 1

Baer / McKool (2009) BAER, J.; MCKOOL, S. S.: *Assessing Creativity Using the Consensual Assessment Technique*. In: Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education. Hershey, Pennsylvania : IGI Global (2009)

Baird et al. (2000) BAIRD, F.; MOORE, C. J.; JAGODZINSKI, A. P.: *An ethnographic study of engineering design teams at Rolls-Royce Aerospace*. In: Design Studies 21 (2000), Nr. 4

Ball et al. (1994) BALL, L. J.; EVANS, J. S. B. T.; DENNIS, I.: *Cognitive processes in engineering design: A longitudinal study*. In: Ergonomics 37 (1994), Nr. 11

Barabasz / Barabasz (1995) BARABASZ, ARREED; BARABASZ, MARIANNE: *Attention Deficit Hyperactivity Disorder:* In: Journal of Neurotherapy 1 (1995), Nr. 1

Basadur / Thompson (1986) BASADUR, M.; THOMPSON, R.: *Usefulness of the Ideation Principle of Extended Effort in Real World Professional and Managerial Creative Problem Solving*. In: The Journal of Creative Behavior 20 (1986), Nr. 1

Basili / Rombach (1988) BASILI, V. R.; ROMBACH, H. D.: *The TAME project: Towards improvement-oriented software environments*. In: Software Engineering, IEEE Transactions on 14 (1988), Nr. 6

Beitz et al. (1992) BEITZ, W.; BIRKHOFER, H.; PAHL, G.: *Konstruktionsmethodik in der Praxis*. In: Konstruktion 44 (1992), Nr. 12

Bender et al. (2002) BENDER, B.; REINICKE, T.; WÜNSCHE, T.; BLESSING, LUCIENNE T.M.: *Application of Methods from Social Sciences in Design Research*. In: Proceedings of the 7th International Design Conference - DESIGN 2002. Dubrovnik, Croatia (2002)

Bender (2003) BENDER, BERND: *Task design and task analysis for empirical studies into design activity*. In: Journal of Engineering Design 14 (2003), Nr. 4

Bender (2004) BENDER, BERND: *Erfolgreiche individuelle Vorgehensstrategien in frühen Phasen der Produktentwicklung*. Technische Universität Berlin, (2004)

Berndt (2011) BERNDT, JÜRGEN: *Qualitative Sozialforschung*. GRIN Verlag (2011)

Binmore / Blackwell (1991) BINMORE, KENNETH GEORGE; BLACKWELL, BASIL, K.: *Essays on the foundations of game theory*. In: Managerial and Decision Economics 12 (1991), Nr. 4

Birkhofer (1991) BIRKHOFER, H.: *Methodik in der Konstruktionspraxis–Erfolge, Grenzen und Perspektiven*. In: Proceedings of the International Conference on Engineering Design (ICED'91). 91. Zurich, Switzerland (1991)

Birkhofer et al. (2001) BIRKHOFER, HERBERT; LINDEMANN, UDO; ALBERS, ALBERT; MEIER, M.: *Product development as a structured and interactive network of knowledge - a revolutionary approach*. In: Proceedings of the 13th International Conference on Engineering Design (ICED'01). Glasgow (2001)

Birkhofer et al. (2002) BIRKHOFER, H.; KLOBERDANZ, H.; BERGER, B.; SAUER, T.: *Cleaning up Design Methods - Describing Methods Completely and Standardised*. In: Proceedings of the 7th International Design Conference - DESIGN 2002. Dubrovnik, Croatia (2002)

Birkhofer (2004) BIRKHOFER, H.: *There is nothing as practical as a good Theory–An Attempt to deal with the Gap between Design Research and Design Practice*. In: Proceedings of DESIGN 2004. Dubrovnik, Croatia (2004)

Björk / Ottosson (2007) BJÖRK, EVASTINA; OTTOSSON, STIG: *Aspects of consideration in product development research*. In: Journal of Engineering Design 18 (2007), Nr. 3

Blessing et al. (1998) BLESSING, LUCIENNE T.M.; CHAKRABARTI, AMARESH; WALLACE, KEN M.: *An Overview of Descriptive Studies in Relation to a General Design Research Methodology*. In: Frankenberger, Dr-Ing Eckart ; Birkhofer, Herbert ; Badke-Schaub, Petra (Hrsg.): Designers: The Key to Successful Product Development : Springer London (1998)

Blessing (2002) BLESSING, LUCIENNE T.M.: *What is this thing called Design Research?* In: Annals of the 2002 International CIRP Design Seminar. Hong Kong (2002)

Blessing / Baumgaertner (2001) BLESSING, LUCIENNE T.M.; BAUMGAERTNER, C. E.: *Establishing casual links between success characteristics and project outcome*. In: Proceedings of the 13th International Conference on Engineering Design (ICED'01) (2001)

Blessing / Chakrabarti (2002) BLESSING, LUCIENNE T.M.; CHAKRABARTI, AMARESH: *DRM: A Design Research Methodology*. In: Proceedings of Les Sciences de la Conception. Lyon, France (2002)

Blessing / Chakrabarti (2009) BLESSING, LUCIENNE T.M.; CHAKRABARTI, AMARESH: *DRM, a Design Research Methodology*. Springer London (2009)

Boden (1990) BODEN, M.: *The Creative Mind: Myths and Mechanisms*. George Weidenfel & Nicolson Ltd. (1990)

Boersting et al. (2008) BOERSTING, PETER; KELLER, R.; ECKERT, CLAUDIA M.; ALBERS, ALBERT; CLARKSON, P. J.: *The relationship between functions and requirements for an improved detection of component linkages*. In: Proceedings of DESIGN 2008. Dubrovnik, Croatia (2008)

Boersting (2012) BOERSTING, PETER: *Konstruktionsbarrieren in der Mikrosystemtechnik und deren Überwindung durch einen Funktionskatalog*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2012)

Boring (1953) BORING, EDWIN G.: *A history of introspection*. In: Psychological Bulletin 50 (1953), Nr. 3

Bortz et al. (2000) BORTZ, JÜRGEN; LIENERT, GUSTAV ADOLF; BOEHNKE, KLAUS: *Verteilungsfreie Methoden in der Biostatistik*. Springer DE (2000)

Bracken (2001) BRACKEN, P.: *Business war gaming*. In: Scenario and Strategy Planning 3 (2001), Nr. 2

Braun et al. (2013a) BRAUN, ANDREAS; EBEL, BJÖRN; ALBERS, ALBERT: *Activity-Based Modeling and Analysis of Product Engineering Processes*. In: Abramovici, Michael ; Stark, Rainer (Hrsg.): Smart Product Engineering : Springer Berlin Heidelberg (2013)

Braun et al. (2013b) BRAUN, ANDREAS; EBEL, BJOERN; ALBERS, ALBERT: *Aktivitätenbasierte Analyse von Produktentstehungsprozessen*. In: Tag des Systems Engineering : Carl Hanser Verlag GmbH & Co. KG (2013)

Breach (2009) BREACH, MARK: *Dissertation Writing for Engineers and Scientists*. Pearson Education (2009)

Briand et al. (1998) BRIAND, L.; KIM, Y. M.; MELO, W.; SEAMAN, C.; BASILI, V. R.: *Q-MOPP: qualitative evaluation of maintenance organizations, processes and products*. In: Journal of Software Maintenance 10 (1998), Nr. 4

Bucciarelli (1988) BUCCIARELLI, L. L.: *An ethnographic perspective on engineering design*. In: Design studies 9 (1988), Nr. 3

Buffinton et al. (2002) BUFFINTON, K. W.; JABLOKOW, K. W.; MARTIN, K. A.: *Project team dynamics and cognitive style*. In: Engineering Management Journal 14 (2002), Nr. 3

Buhler / Guirl (1963) BUHLER, E. O.; GUIRL, E. N.: *The more able student: Described and rated*. In: Educating the academically able (1963)

Bunge (1983) BUNGE, M.: *Treatise on Basic Philosophy: Volume 5: Epistemology & Methodology I: Exploring the World*. Springer (1983)

Buttner / Gryskiewicz (1993) BUTTNER, E. H.; GRYSKIEWICZ, N.: *Entrepreneurs' Problem-solving styles: An empirical study using the Kirton adaption/innovation theory*. In: Journal of Small Business Management 31 (1993)

Campbell / Stanley (1966) CAMPBELL, DONALD T.; STANLEY, JULIAN C.: *Experimental and Quasi-Experimental Designs for Research*. Cengage Learning (1966)

Candy (1996) CANDY, L.: *Understanding creativity: an empirical approach*. In: Proc. Creativity and Cognition (1996)

Cantamessa (2003) CANTAMESSA, MARCO: *An empirical perspective upon design research*. In: Journal of Engineering Design 14 (2003), Nr. 1

Capraro / Capraro (2002) CAPRARO, R. M.; CAPRARO, M. M.: *Myers-Briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study*. In: Educational and Psychological Measurement 62 (2002), Nr. 4

Carlyn (1977) CARLYN, M.: *An assessment of the Myers-Briggs type indicator*. In: Journal of Personality Assessment 41 (1977), Nr. 5

Carmines / Zeller (1979) CARMINES, EDWARD G.; ZELLER, RICHARD A.: *Reliability and Validity Assessment*. SAGE (1979)

Carskadon / Cook (1982) CARSKADON, T. G.; COOK, D. D.: *Validity of MBTI type descriptions as perceived by recipients unfamiliar with type*. In: Research in Psychological Type 8 (1982)

Chan (2000) CHAN, D.: *Detection of differential item functioning on the Kirton Adaption-Innovation Inventory using multiple-group mean and covariance structure analyses*. In: Multivariate Behavioral Research 35 (2000), Nr. 2

Clandinin / Connelly (2000) CLANDININ, D. JEAN; CONNELLY, F. MICHAEL: *Narrative inquiry: experience and story in qualitative research*. San Francisco : Jossey-Bass Publishers (2000)

Clarkson / Eckert (2004) : *Design Process Improvement: A review of current practice*. Springer (2004)

Coley et al. (2007) COLEY, FIONA; HOUSEMAN, OLIVER; ROY, RAJKUMAR: *An introduction to capturing and understanding the cognitive behaviour of design engineers*. In: Journal of Engineering Design 18 (2007), Nr. 4

Cooper (1999) COOPER, R. G.: *The invisible success factors in product innovation*. In: Journal of Product Innovation Management 16 (1999), Nr. 2

Cooper / Kleinschmidt (1987) COOPER, R. G.; KLEINSCHMIDT, E. J.: *New products: what separates winners from losers?* In: Journal of product innovation management 4 (1987), Nr. 3

Crawford (1977) CRAWFORD, C. MERLE: *Marketing Research and the New Product Failure Rate*. In: Journal of Marketing 41 (1977), Nr. 2

Cross (1996) CROSS, N.: *Creativity in design: not leaping but bridging*. In: Proceedings of the 2nd International Symposium on Creativity and Cognition (1996)

Csikszentmihalyi (1988) CSIKSZENTMIHALYI, M.: *Society, culture, and person: A systems view of creativity*. In: The nature of creativity: Contemporary psychological perspectives : Cambridge University Press (1988)

Dacey (1989) DACEY, JOHN S.: *Fundamentals of Creative Thinking*. Lexington Books (1989)

Danmin et al. (2000) DANMIN, M.; EN, H.; JIANJUN, R. C. C. R.: *The validity analysis of the Chinese version MBTI*. In: JOURNAL OF CHINESE PSYCHOLOGY ACTA PSYCHOLOGICA SINICA (2000)

Darwin (1877) DARWIN, C.: *A Biographical Sketch of an Infant*. In: Mind (1877)

Deigendesch (2009) DEIGENDESCH, TOBIAS: *Kreativität in der Produktentwicklung und Muster als methodisches Hilfsmittel / von Tobias Deigendesch*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2009)

Descartes (1931) DESCARTES, R.: *Mediations on first philosophy. 1641*. In: Haldane, ES; Ross, GRT Trans.: The philosophical works of Rene Descartes. Cambridge University Press, Cambridge (1931)

DIN EN ISO 9000 (2005) DIN EN ISO 9000: *Quality management systems - Fundamentals and vocabulary (ISO 9000:2005); Trilingual version EN ISO 9000:2005*. , Beuth (2005)

DIN EN ISO 9001 (2008) DIN EN ISO 9001: *Quality management systems - Requirements (ISO 9001:2008); Trilingual version EN ISO 9001:2008*. , Beuth (2008)

Dodge (2008) DODGE, YADOLAH: *The Concise Encyclopedia of Statistics*. Springer (2008)

Dorst (2003) DORST, K.: *The problem of design problems*. In: Expertise in design (2003)

Dresler / Baudson (2008) DRESLER, MARTIN; BAUDSON, TANJA GABRIELE: *Kreativität: Beiträge aus den Natur- und Geisteswissenschaften*. 1. Aufl., Hirzel, Stuttgart (2008)

Drever / Education (1995) DREVER, ERIC; EDUCATION, SCOTTISH COUNCIL FOR RESEARCH IN: *Using semi-structured interviews in small-scale research: a teacher's guide*. Scottish Council for Research in Education (1995)

Duffy (1997) DUFFY, A. H. B.: *The design productivity debate*. Springer-Verlag New York, Inc. (1997)

Eastman (1968) EASTMAN, C. M.: *On the analysis of intuitive design processes*. Carnegie-Mellon University (1968)

Eckert et al. (2003) ECKERT, C. M; STACEY, M. K; CLARKSON, P. J: *The spiral of applied research: A methodological view on integrated design research*. In: Proceedings of the 14th International Conference on Engineering Design (ICED'03). Stockholm, Sweden (2003)

Eckert et al. (2009) ECKERT, CLAUDIA M.; WYATT, DAVID F.; CLARKSON, P. JOHN: *The elusive act of synthesis: creativity in the conceptual design of complex engineering products*. In: Proceedings of the 7th ACM conference on Creativity and Cognition. New York, NY, USA : ACM (2009)

Eckert et al. (2011) ECKERT, CLAUDIA; ALINK, THOMAS; RUCKPAUL, ANNE; ALBERS, ALBERT: *Different notions of function: results from an experiment on the analysis of an existing product*. In: Journal of Engineering Design 22 (2011), Nr. 11-12

Ehrlenspiel et al. (2005) EHRLENSPIEL, K.; KIEWERT, A.; LINDEMANN, U.: *Kostengünstig entwickeln und konstruieren*. Springer DE (2005)

Ehrlenspiel (2007) EHRLENSPIEL, K.: *Integrierte Produktentwicklung*. Hanser Verlag (2007)

Eisenhardt (1989) EISENHARDT, K. M.: *Building theories from case study research*. In: Academy of management review (1989)

Engels (1987) ENGELS, F.: *The condition of the working class in England*. Penguin (1987)

Ericsson / Simon (1993) ERICSSON, K. A.; SIMON, H. A.: *Protocol analysis: Verbal reports as data (rev. ed.) MIT Press*. In: Cambridge, MA (1993)

Ernst (2002) ERNST, HOLGER: *Success Factors of New Product Development: A Review of the Empirical Literature*. In: International Journal of Management Reviews 4 (2002), Nr. 1

Everitt / Skrondal (2010) EVERITT, BRIAN; SKRONDAL, ANDERS: *The Cambridge dictionary of statistics*. Cambridge; New York : Cambridge University Press (2010)

Eysenck (1994) EYSENCK, H. J.: *The measurement of creativity.* The MIT Press (1994)

Feyerabend (1980) FEYERABEND, PAUL K: *Erkenntnis für freie Menschen*. Frankfurt am Main : Suhrkamp (1980)

Feyerabend (1984) FEYERABEND, PAUL K: *Wissenschaft als Kunst*. Frankfurt am Main : Suhrkamp (1984)

Flick (1997) FLICK, UWE: *The episodic interview*. Methodology Institute (1997)

Flick (2009) FLICK, UWE: *An Introduction to qualitative research*. London [etc.] : SAGE (2009)

Flyvbjerg (2006) FLYVBJERG, B.: *Five misunderstandings about case-study research*. In: Qualitative inquiry 12 (2006), Nr. 2

Foxall / Hackett (1994) FOXALL, G. R.; HACKETT, P. M. W.: *Styles of Managerial Creativity: A Comparison of Adaption-Innovation in the United Kingdom, Australia and the Unites States1*. In: British Journal of Management 5 (1994), Nr. 2

Frankenberger / Badke-Schaub (1998) FRANKENBERGER, E.; BADKE-SCHAUB, P.: *Modelling design processes in industry—empirical investigations of design work in practice*. In: Automation in construction 7 (1998), Nr. 2

Frey (2007) FREY, BRUNO S.: *Evaluierungen, Evaluierungen … Evaluitis*. In: Perspektiven der Wirtschaftspolitik 8 (2007), Nr. 3

Frey / Dym (2006) FREY, DANIEL D.; DYM, CLIVE L.: *Validation of design methods: lessons from medicine*. In: Research in Engineering Design 17 (2006), Nr. 1

Friebertshäuser (2003) FRIEBERTSHÄUSER, BARBARA: *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim; München : Juventa-Verl. (2003)

Fromm (1959) FROMM, E: *The creative attitude*. In: Creativity and its cultivation. New York : Harper & Row (1959)

Fromm / Haist (1989) FROMM, H.; HAIST, F.: *Qualität im Unternehmen: Prinzipien–Methoden–Techniken*. Hanser Verlag, München, Wien (1989)

Frost (1999) FROST, R. B.: *Why Does Industry Ignore Design Science?* In: Journal of Engineering Design 10 (1999), Nr. 4

Fulcher / Hills (1996) FULCHER, A. J.; HILLS, P.: *Towards a Strategic Framework for Design Research*. In: Journal of Engineering Design 7 (1996), Nr. 2

Furnham (1996) FURNHAM, A.: *The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality*. In: Personality and Individual Differences 21 (1996), Nr. 2

Gantt (1903) GANTT, HENRY LAURENCE: *A Graphical Daily Balance in Manufacture*. (1903)

Gardner (1993) GARDNER, H. E.: *Creating Minds: An Anatomy of Creativity As Seen Through the Lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Ghandi*. Basic Books (1993)

Geis et al. (2008) GEIS, C.; BIERHALS, R.; SCHUSTER, I.; BADKE-SCHAUB, P.; BIRKHOFER, H.: *Methods in practice–a study on requirements for development and transfer of design methods*. In: Proceedings of the DESIGN 2008. Dubrovnik, Croatia (2008)

Gero / Kannengiesser (2002) GERO, J. S.; KANNENGIESSER, U.: *The situated function-behaviour-structure framework*. In: Artificial Intelligence in Design. 2 (2002)

Gero / Mc Neill (1998) GERO, J. S.; MC NEILL, T.: *An approach to the analysis of design protocols*. In: Design studies 19 (1998), Nr. 1

Gero / Tang (2001) GERO, J. S.; TANG, H. H.: *The differences between retrospective and concurrent protocols in revealing the process-oriented aspects of the design process*. In: Design Studies 22 (2001), Nr. 3

Getzels / Jackson (1962) GETZELS, J. W.; JACKSON, P. W.: *Creativity and intelligence: Explorations with gifted students.* (1962)

Giapoulis (1998) GIAPOULIS, A.: *Modelle für effektive Konstruktionsprozesse*. Shaker (1998)

Gigerenzer (1998) GIGERENZER, G.: *We need statistical thinking, not statistical rituals*. In: Behavioral and Brain Sciences 21 (1998), Nr. 2

Gigerenzer (2004) GIGERENZER, G.: *Mindless statistics*. In: Journal of Socio-Economics 33 (2004), Nr. 5

Gigerenzer / Goldstein (1996) GIGERENZER, G.; GOLDSTEIN, D. G.: *Reasoning the fast and frugal way: models of bounded rationality.* In: Psychological review 103 (1996), Nr. 4

Gigerenzer / Todd (1999) GIGERENZER, G.; TODD, P. M.: *Simple heuristics that make us smart*. Oxford University Press, USA (1999)

Gilad (2009) GILAD, BENJAMIN: *Business War Games: How large, small, and new companies can vastly improve their strategies and outmaneuver the competition*. Career press (2009)

Gill (1990) GILL, H.: *Adoption of Design Science by Industry - Why so Slow?* In: Journal of Engineering Design 1 (1990), Nr. 3

Goldsmith (1984) GOLDSMITH, R. E.: *Personality characteristics associated with adaption-innovation*. In: The Journal of psychology 117 (1984), Nr. 2

Gordon (1973) GORDON, WILLIAM J. J.: *The metaphorical way of learning & knowing*. Porpoise Books (1973)

Gough / Library (1981) GOUGH, H. G.; LIBRARY, ISABEL BRIGGS MYERS MEMORIAL: *Studies of the Myers-Briggs Type Indicator in a personality assessment research institute*. Isabel Briggs Memorial Library (1981)

Grabowski / Geiger (1997) GRABOWSKI, H.; GEIGER, K.: *Neue Wege zur Produktentwicklung*. Raabe (1997)

Grinnell / Unrau (2010) GRINNELL, RICHARD M. JR.; UNRAU, YVONNE A.: *Social Work Research and Evaluation: Foundations of Evidence-Based Practice*. Oxford University Press (2010)

Guba (1990) GUBA, E.: *G.(1990) The alternative paradigm dialog*. In: The Paradigm Dialog (1990)

Gubrium (2012) GUBRIUM, JABER F: *The Sage handbook of interview research: the complexity of the craft*. Thousand Oaks : SAGE (2012)

Gubrium / Holstein (2002) GUBRIUM, JABER F; HOLSTEIN, JAMES A: *Handbook of interview research: context & method*. Thousand Oaks, Calif. : Sage Publications (2002)

Guilford (1950) GUILFORD, J P: *Creativity*. In: The American psychologist 5 (1950), Nr. 9

Guilford (1959) GUILFORD, J. P.: *Traits of creativity*. In: Creativity and its cultivation (1959)

Guilford (1960) GUILFORD, J. P.: *The psychology of creativity*. In: Creative crafts 1 (1960)

Guilford (1977) GUILFORD, JOY PAUL: *Way Beyond the IQ*. Creative Education Foundation (1977)

Habermas (1972) HABERMAS, JURGEN: *Knowledge and human Interests*. Beacon Press, USA (1972)

Hales (1986) HALES, C.: *Analysis of the engineering design process in an industrial context.* University of Cambridge, (1986)

Hegel (1959) HEGEL, GEORG WILHELM FRIEDRICH: *Encyclopedia of philosophy*. Philosophical Library (1959)

Helfferich (2005) HELFFERICH, CORNELIA: *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. 2. Aufl., VS Verlag für Sozialwissenschaften (2005)

Hennessey (1994) HENNESSEY, B. A.: *The consensual assessment technique: An examination of the relationship between ratings of product and process creativity*. In: Creativity Research Journal 7 (1994), Nr. 2

Hennessey et al. (1999) HENNESSEY, B. A.; AMABILE, T. M.; MUELLER, J. S.: *Consensual assessment*. In: Encyclopedia of creativity 1 (1999)

Herman / Frost (2008) HERMAN, MARK L.; FROST, MARK D.: *Wargaming for leaders: Strategic decision making from the battlefield to the boardroom*. McGraw-Hill (2008)

Heunks (1998) HEUNKS, F. J.: *Innovation, creativity and success*. In: Small Business Economics 10 (1998), Nr. 3

Hocevar / Bachelor (1989) HOCEVAR, D.; BACHELOR, P.: *A taxonomy and critique of measurements used in the study of creativity.* (1989)

Honderich (1995) HONDERICH, TED: *The Oxford Companion to Philosophy*. Oxford University Press (1995)

Hong et al. (1995) HONG, EUNSOOK; MILGRAM, ROBERTA M.; GORSKY, HANA: *Original thinking as a predictor of creative performance in young children*. In: Roeper Review 18 (1995), Nr. 2

Hornbostel / Schelling (2011) : *Evaluation: New Balance of Power?* , iFQ-Working Paper. No. 9Berlin : (2011)

Horváth (2001) HORVÁTH, IMRE: *A contemporary survey of scientific research into engineering design*. In: Proceedings of the 13th International Conference on Engineering Design (ICED'01) (2001)

Horváth (2004) HORVÁTH, IMRE: *A treatise on order in engineering design research*. In: Research in Engineering Design 15 (2004), Nr. 3

Howard et al. (2008) HOWARD, T. J.; CULLEY, S. J.; DEKONINCK, E.: *Describing the creative design process by the integration of engineering design and cognitive psychology literature*. In: Design studies 29 (2008), Nr. 2

Hubka / Eder (1982) HUBKA, V.; EDER, W. E.: *Principles of engineering design*. Butterworth Scientific London (1982)

Hunsley et al. (2003) HUNSLEY, J.; LEE, C. M.; WOOD, J. M.: *Controversial and questionable assessment techniques*. In: Science and pseudoscience in clinical psychology (2003)

Hutchinson / Skinner (2007) HUTCHINSON, L. R.; SKINNER, N. F.: *Self-awareness and cognitive style: Relationships among adaption-innovation, self-monitoring, and self-consciousness*. In: Social Behavior and Personality: an international journal 35 (2007), Nr. 4

Hutterer (2005) HUTTERER, P.: *Reflexive Dialoge und Denkbausteine für die methodische Produktentwicklung*. Dr. Hut (2005)

Isaksen (2007) ISAKSEN, S. G.: *The climate for transformation: Lessons for leaders*. In: Creativity and Innovation Management 16 (2007), Nr. 1

Isaksen / Ekvall (2010) ISAKSEN, S. G.; EKVALL, G.: *Managing for innovation: The two faces of tension in creative climates*. In: Creativity and Innovation Management 19 (2010), Nr. 2

Jabri (1991) JABRI, M. M.: *The development of conceptually independent subscales in the measurement of modes of problem solving*. In: Educational and Psychological Measurement 51 (1991), Nr. 4

Jänsch et al. (2005) JÄNSCH, J.; BIRKHOFER, H.; WALTHER, J.: *The development of design methods and expertise*. In: Proceedings of the 15th International Conference on Engineering Design (ICED'05). Melbourne, Australia (2005)

Jänsch (2007) JÄNSCH, JUDITH: *Akzeptanz und Anwendung von Konstruktionsmethoden im industriellen Einsatz : Analyse und Empfehlungen aus kognitionswissenschaftlicher Sicht*. Technische Universität Darmstadt, (2007)

Jänsch / Birkhofer (2004) JÄNSCH, JUDITH; BIRKHOFER, HERBERT: *The gap between learning and applying design methods*. In: Proceedings of DESIGN 2004. Dubrovnik, Croatia (2004)

Jansson / Smith (1991) JANSSON, D. G.; SMITH, S. M.: *Design fixation*. In: Design Studies 12 (1991), Nr. 1

Johnson et al. (2008) JOHNSON, ROBERT L.; PENNY, JAMES A.; GORDON, BELITA: *Assessing Performance: Designing, Scoring, and Validating Performance Tasks*. Guilford Press (2008)

Jones (1962) JONES, J.C.: *The Conference on design methods: papers presented at the conference on systematic and intuitive methods in engineering, industrial design, architecture and communications.* London : Pergamon Press (1962)

Jorden (1983) JORDEN, W.: *Die Diskrepanz zwischen Konstruktionspraxis und Konstruktionsmethodik*. In: Proceedings of the ICED International Conference on Engineering Design (ICED'83) (1983)

Jung (1921) JUNG, C. G.: *Psychological Types, Vol. 6, The collected works of C. G Jung*. (1921)

Kant (1794) KANT, IMMANUEL: *Metaphysische Anfangsgründe der Naturwissenschaft*. (1794)

Kant (1933) KANT, IMMANUEL: *Critique of pure reason [Reprint 1933; original 1781; translated 1890]*. London : St. Martins Press (1933)

Kanter (1988) KANTER, R. M.: *When a thousand flowers bloom: Structural, collective, and social conditions for innovation in organization*. JAI Press Greenwich, CT (1988)

Kaufman / Baer (2004) KAUFMAN, J. C.; BAER, J.: *Sure, I'm creative - but not inmathematics! Self-reported creativity in diverse domains*. In: Empirical Studies of the Arts 22 (2004), Nr. 2

Kaufmann (2004) KAUFMANN, G.: *Two kinds of creativity–but which ones?* In: Creativity and innovation Management 13 (2004), Nr. 3

Keats (1999) KEATS, DAPHNE: *Interviewing: A Practical Guide for Students and Professionals*. UNSW Press (1999)

Keller et al. (2007) KELLER, R.; ALINK, T.; PFEIFER, C.; ECKERT, C. M; CLARKSON, P. J; ALBERS, A.: *Product models in design: a combined use of two models to assess change risks*. In: Proceedings of the 16th International Conference on Engineering Design (ICED'07). Paris, France (2007)

Keller / Binz (2009) KELLER, A.; BINZ, H.: *Requirements on engineering design methodologies*. In: Proceedings of the 17th International Conference on Engineering Design (ICED'09) (2009)

Khatena / Torrance (1973) KHATENA, J.; TORRANCE, E.P.: *Thinking creatively with sounds and words*. In: Technical Manual (Research Ed.). Lexington, MA : Personnel Press (1973)

King (2010) KING, NIGEL: *Interviews in qualitative research*. SAGE (2010)

Kirton (1976) KIRTON, MICHAEL: *Adaptors and innovators: A description and measure*. In: Journal of Applied Psychology 61 (1976), Nr. 5

Kirton (2003) KIRTON, M. J.: *Adaption-innovation: in the context of change and diversity*. Psychology Press (2003)

Kleinschmidt / Cooper (1991) KLEINSCHMIDT, E. J.; COOPER, R. G.: *The impact of product innovativeness on performance*. In: Journal of product innovation management 8 (1991), Nr. 4

König (2005) KÖNIG, ECKARD: *Systemisch denken und handeln: personale Systemtheorie in Erwachsenenbildung und Organisationsberatung*. Weinheim; Basel : Beltz (2005)

Kridel (2010) KRIDEL, CRAIG: *Encyclopedia of Curriculum Studies*. SAGE (2010)

Kubes (1998) KUBES, M.: *Adaptors and innovators in Slovakia: Cognitive style and social culture*. In: European journal of personality 12 (1998), Nr. 3

Kuhn (1977) KUHN, T. S.: *Objectivity, Value Judgment, and Theory Choice*. (1977)

Kuhn (1996) KUHN, T. S.: *The structure of scientific revolutions*. Chicago : University of Chicago Press (1996)

Kumar et al. (1991) KUMAR, V. K.; HOLMAN, E. R.; RUDEGEAIR, P.: *Creativity Styles of Freshmen Students*. In: The Journal of Creative Behavior 25 (1991), Nr. 4

Kvale (1996) KVALE, STEINA: *InterViews: an introduction to qualitative research interviewing*. Thousand Oaks : SAGE Publications (1996)

Kwang et al. (2005) KWANG, N. A.; ANG, R. P.; OOI, L. B.; SHIN, W. S.; OEI, T. P. S.; LENG, V.: *Do adaptors and innovators subscribe to opposing values?* In: Creativity Research Journal 17 (2005), Nr. 2-3

LeCompte / Schensul (1999) LECOMPTE, M. D.; SCHENSUL, J. J.: *Ethnographer's toolkit*. AltaMira Press (1999)

Lee et al. (2000) LEE, CHENG F.; LEE, JOHN C.; LEE, ALICE C.: *Statistics for Business and Financial Economics*. World Scientific (2000)

Lee (2005) LEE, KYUNG-HWA: *The Relationship between Creative Thinking Ability and Creative Personality of Preschoolers*. In: International Education Journal 6 (2005), Nr. 2

Levenspiel (2007) LEVENSPIEL, OCTAVE: *Rambling Through Science and Technology*. Self-Published © 2007 by Octave Levenspiel (2007)

Lewis-Beck (1995) LEWIS-BECK, MICHAEL S.: *Data Analysis: An Introduction*. SAGE (1995)

Lienert / Raatz (1998) LIENERT, GUSTAV A.; RAATZ, ULRICH: *Testaufbau und Testanalyse*. Weinheim : Beltz (1998)

Lind (1757) LIND, JAMES: *A Treatise on the Scurvy: In Three Parts, Containing an Inquiry Into the Nature, Causes, and Cure, of that Disease*. A. Millar (1757)

Lindemann (2002) LINDEMANN, U.: *Flexible adaptation of methods within the design process*. In: Proceedings of the 7th International Design Conference - DESIGN 2002. Dubrovnik, Croatia (2002)

Lindemann (2009) LINDEMANN, UDO: *Methodische Entwicklung technischer Produkte: Methoden flexibel und situationsgerecht anwenden*. Springer DE (2009)

Linn / Gronlund (1999) LINN, ROBERT L.; GRONLUND, NORMAN E.: *Measurement and Assessment in Teaching*. Pearson College Division (1999)

Loch et al. (1996) LOCH, CHRISTOPH; STEIN, LOTHAR; TERWIESCH, CHRISTIAN: *Measuring Development Performance in the Electronics Industry*. In: Journal of Product Innovation Management 13 (1996), Nr. 1

Lohmeyer (2013) LOHMEYER, QUENTIN: *Menschzentrierte Modellierung von Produktentstehungssystemen unter besonderer Berücksichtigung der Synthese und Analyse dynamischer Zielsysteme*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2013)

Lohse (1993) LOHSE, G.: *Methodical product development-experience of a practical study*. In: Proceedings of the 9th International Conference on Engineering Design (ICED'93) (1993)

Lubart (2001) LUBART, TODD I.: *Models of the Creative Process: Past, Present and Future*. In: Creativity Research Journal 13 (2001), Nr. 3-4

MacKinnon (1978) MACKINNON, DONALD W: *In search of human effectiveness*. Buffalo, N.Y. : Creative Education Foundation (1978)

Maier et al. (2007) MAIER, G. W.; STREICHER, B.; JONAS, E.; FREY, D.: *Innovation und Kreativität*. In: Enzyklopädie der Psychologie-Wirtschaftspsychologie (2007)

Maier (2011) MAIER, THOMAS: *Ein Beitrag zur Erstellung komplexer strukturdynamischer Simulationsmodelle und deren Validierung mittels der 3D-Laservibrometrie*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2011)

Malinowski (1921) MALINOWSKI, B.: *The primitive economics of the Trobriand Islanders*. In: The Economic Journal 31 (1921), Nr. 121

Marxen / Albers (2012) MARXEN, LEIF; ALBERS, ALBERT: *SUPPORTING VALIDATION IN THE DEVELOPMENT OF Design methods*. In: Proceedings of the 12th International Design Conference DESIGN 2012. Dubrovnik, Croatia (2012)

Maslow (1976) MASLOW, A.H.: *Creativity in self-actualizing people*. In: The Creativity Question. Durham, NC : Duke University Press (1976)

Mathis / Jackson (2008) MATHIS, ROBERT L; JACKSON, JOHN HAROLD: *Human resource management*. Mason, OH : Thomson/South-western (2008)

McCrae / Costa Jr. (1989) MCCRAE, R. R.; COSTA JR., P. T.: *Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality*. In: Journal of personality 57 (1989), Nr. 1

McDermott et al. (2008) MCDERMOTT, ROBIN E.; MIKULAK, RAYMOND J.; BEAUREGARD, MICHAEL R.: *The basics of FMEA*. Productivity Press (2008)

McGown et al. (1998) MCGOWN, A.; GREEN, G.; RODGERS, P. A.: *Visible ideas: information patterns of conceptual sketch activity*. In: Design studies 19 (1998), Nr. 4

Meboldt (2008) MEBOLDT, M.: *Mentale und formale Modellbildung in der Produktentstehung-als Beitrag zum integrierten Produktentstehungs-Modell (iPeM)*. IPEK - Institut für Produktentwicklung, Universität Karlsruhe (TH), (2008)

Mednick / Mednick (1965) MEDNICK, SARNOFF A.; MEDNICK, MARTHA T.: *The Associative Basis of the Creative Process*. University of Michigan (1965)

Meneely / Portillo (2005) MENEELY, J.; PORTILLO, M.: *The adaptable mind in design: Relating personality, cognitive style, and creative performance*. In: Creativity Research Journal 17 (2005), Nr. 2-3

Michalewicz / Fogel (2004) MICHALEWICZ, ZBIGNIEW; FOGEL, DAVID B.: *How to Solve It: Modern Heuristics*. Springer (2004)

Milgram / Hong (1993) MILGRAM, R. M.; HONG, E.: *Creative thinking and creative performance in adolescents as predictors of creative attainments in adults: A follow-up study after 18 years*. In: Roeper review 15 (1993), Nr. 3

Mudd (1996) MUDD, S.: *Kirton's A–I theory: Evidence bearing on the style/level and factor composition issues*. In: British Journal of Psychology 87 (1996), Nr. 2

Myers (1962) MYERS, I. B.: *The Myers-Briggs Type Indicator*. Consulting Psychologists Press Palo Alto, CA (1962)

Myers et al. (1985) MYERS, I. B.; MCCAULLEY, M. H.; MOST, R.: *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press Palo Alto, CA (1985)

Myers (1999) MYERS, M.: *Investigating information systems with ethnographic research*. In: Communications of the AIS 2 (1999), Nr. 4es

Myers-Briggs (1962) MYERS-BRIGGS, I.: *The Myers-Briggs type indicator manual*. In: Princeton, NJ: Educational Testing Service (1962)

Naisbitt (1982) NAISBITT, JOHN: *Megatrends: Ten New Directions Transforming Our Lives*. Reissue. Aufl., Grand Central Publishing (1982)

Nash (1950) NASH, J.: *Non-cooperative games. Princeton University Mathematics Department*. Ph. D. thesis. Princeton, NJ, (1950)

Nash (1996) NASH, JOHN F.: *Essays on Game Theory*. Edward Elgar Publishing (1996)

Neumann / Morgenstern (1944) NEUMANN, JOHN VON; MORGENSTERN, OSKAR: *Theory Of Games And Economic Behavior*. Princeton University Press. (1944)

Nunnally et al. (1967) NUNNALLY, JUM C.; BERNSTEIN, IRA H.; BERGE, JOS MF TEN: *Psychometric theory*. . 2 McGraw-Hill New York (1967)

Nystrom (1979) NYSTROM, H.: *Creativity and innovation*. In: R & D Management 10 (1979), Nr. 1

O'Donovan et al. (2004) O'DONOVAN, BRENDAN; ECKERT, CLAUDIA; CLARKSON, P. JOHN: *Simulating Design Processes to Assist Design Process Planning*. In: . 2004 : ASME (2004)

Oerding (2009) OERDING, J.: *Ein Beitrag zum Modellverständnis der Produktentstehung-Strukturierung von Zielsystemen mittels C&CM*. IPEK - Institut für Produktentwicklung, Universität Karlsruhe (TH), (2009)

Okuda et al. (1991) OKUDA, S. M.; RUNCO, M. A.; BERGER, D. E.: *Creativity and the finding and solving of real-world problems*. In: Journal of Psychoeducational Assessment 9 (1991), Nr. 1

Olewnik / Lewis (2005) OLEWNIK, A. T.; LEWIS, K.: *On validating engineering design decision support tools*. In: Concurrent Engineering 13 (2005), Nr. 2

Osborn (1963) OSBORN, A. F.: *Applied Imagination: The principles and procedures of Creative Thinking. 1953*. In: New York: Charles Scribner's Sons (1963)

Osborne / Rubinstein (1994) OSBORNE, MARTIN J.; RUBINSTEIN, ARIEL: *Course in game theory*. The MIT press (1994)

Pahl (1994) PAHL, G.: *Psychologische und pädagogische Fragen beim methodischen Konstruieren*. 1., Aufl. Aufl., TÜV Media GmbH (1994)

Pahl et al. (2005) PAHL, G.; BEITZ, W.; FELDHUSEN, J.; GROTE, K. H.: *Pahl/Beitz Konstruktionslehre: Grundlagen erfolgreicher Produktentwicklung. Methoden und Anwendung*. Springer DE (2005)

Pahl et al. (2006) PAHL, GERHARD; BEITZ, WOLFGANG; FELDHUSEN, JÖRG; GROTE, KARL-HEINRICH: *Engineering Design: A Systematic Approach*. 3rd. ed. Aufl., Springer, Berlin (2006)

Parnes (1961) PARNES, S. J.: *Effects of extended effort in creative problem solving.* In: Journal of Educational psychology 52 (1961), Nr. 3

Peat et al. (2009) PEAT, JENNIFER; BARTON, BELINDA; ELLIOTT, ELIZABETH: *Statistics Workbook for Evidence-based Health Care*. John Wiley & Sons (2009)

Pedersen et al. (2000) PEDERSEN, K.; EMBLEMSVAG, J.; BAILEY, R.; ALLEN, J. K.; MISTREE, F.: *Validating design methods and research: the validation square*. In: ASME Design Theory and Methodology Conference (2000)

Pedgley (2007) PEDGLEY, O.: *Capturing and analysing own design activity*. In: Design Studies 28 (2007), Nr. 5

Plucker (1999) PLUCKER, J. A.: *Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data*. In: Creativity research journal 12 (1999), Nr. 2

Plucker / Renzulli (1999) PLUCKER, J. A.; RENZULLI, J. S.: *Psychometric approaches to the study of human creativity*. In: Handbook of creativity (1999)

Plucker / Runco (1998) PLUCKER, JONATHAN A.; RUNCO, MARK A.: *The Death of Creativity Measurement Has Been Greatly Exaggerated: Current Issues, Recent Advances, and Future Directions in Creativity Assessment.* In: Roeper Review 21 (1998), Nr. 1

Popper (1935) POPPER, KARL: *Logik der Forschung Zur Erkenntnistheorie der Modernen Naturwissenschaft*. Wien : Springer (1935)

Popper (1973) POPPER, KARL R: *Objektive Erkenntnis: ein evolutionärer Entwurf*. Hamburg : Hoffmann, Campe (1973)

Rardin / Uzsoy (2001) RARDIN, RONALD L.; UZSOY, REHA: *Experimental Evaluation of Heuristic Optimization Algorithms: A Tutorial*. In: Journal of Heuristics 7 (2001), Nr. 3

Reason / Rowan (1981) REASON, PETER; ROWAN, JOHN: *Human inquiry: A sourcebook of new paradigm research*. J. Wiley Chichester, UK (1981)

Redelinghuys / Bahill (2006) REDELINGHUYS, CHRISTIAAN; BAHILL, A. TERRY: *A framework for the assessment of the creativity of product design teams*. In: Journal of Engineering Design 17 (2006)

Redtenbacher (1852) REDTENBACHER, FERDINAND: *Prinzipien der Mechanik und des Maschinenbaues*. Bassermann (1852)

Reich (1994) REICH, YORAM: *Layered models of research methodologies*. In: AI EDAM 8 (1994), Nr. 04

Reich (1995) REICH, Y.: *The study of design research methodology*. In: Journal of Mechanical Design 117 (1995)

Reichwald / Conrat (1995) REICHWALD, R.; CONRAT, J. I.: *Integrationslösungen für die Produktentwicklung: eine Wirtschaftlichkeitsbetrachtung auf der Basis von Änderungskosten*. In: VDI-Z 137 (1995), Nr. 5

Rhodes (1961) RHODES, MEL: *An analysis of creativity*. In: The Phi Delta Kappa International 42 (1961), Nr. 7

Rickards / Moger (2000) RICKARDS, T.; MOGER, S.: *Creative leadership processes in project team development: an alternative to Tuckman's stage model*. In: British Journal of Management 11 (2000), Nr. 4

Riessman (1993) RIESSMAN, CATHERINE KOHLER: *Narrative analysis*. Newbury Park, CA : Sage Publications (1993)

Rodgers et al. (2000) RODGERS, P.A; GREEN, G; MCGOWN, A: *Using concept sketches to track design progress*. In: Design Studies 21 (2000), Nr. 5

Rogers (1954) ROGERS, CARL R.: *Toward a theory of creativity*. In: Etc. 11 (1954)

Roozenburg / Cross (1991) ROOZENBURG, N.F.M.; CROSS, N.G.: *Models of the design process: integrating across the disciplines*. In: Design Studies 12 (1991), Nr. 4

Ropohl (1975) ROPOHL, GÜNTER: *Einleitung in die Systemtechnik*. In: Systemtechnik – Grundlagen und Antworten,. München : Carl Hanser Verlag (1975)

Ropohl (2009) ROPOHL, GÜNTER: *Allgemeine Technologie : eine Systemtheorie der Technik*. (2009)

Rosenhan (1972) ROSENHAN, D. L.: *On being sane in insane places*. In: Santa Clara Lawyer 13 (1972)

Rowan (1981) ROWAN, J.: *A dialectical paradigm for research*. In: Human inquiry: A sourcebook of new paradigm research (1981)

Rückert et al. (1997) RÜCKERT, C.; GIESA, H. G.; SCHRODA, F.; BENDER, B.: *Categorizing engineering design problems*. In: Proceedings of the 11th International Conference on Engineering Design ICED'97 (1997)

Rumsey (2011) RUMSEY, DEBORAH: *Statistics for Dummies*. John Wiley and Sons Ltd (2011)

Runco (1989) RUNCO, M. A.: *Parents' and teachers' ratings of the creativity of children.* In: Journal of Social Behavior & Personality; Journal of Social Behavior & Personality (1989)

Runco et al. (2010) RUNCO, MARK A.; MILLAR, GARNET; ACAR, SELCUK; CRAMOND, BONNIE: *Torrance Tests of Creative Thinking as Predictors of Personal and Public Achievement: A Fifty-Year Follow-Up*. In: Creativity Research Journal 22 (2010), Nr. 4

Rutz (1994) RUTZ, A.: *Akzeptanz-und transferfragen beim methodischen konstruieren in der industrie*. In: Psychologische und pädagogische fragen beim methodischen konstruieren, Verlag TÜV Rheinland, Köln (1994)

Sachs (2004) SACHS, LOTHAR: *Angewandte Statistik: Anwendung statistischer Methoden ; mit 317 Tabellen*. Springer DE (2004)

Saunders et al. (2012) SAUNDERS, MARK N. K.; LEWIS, PHILIP; THORNHILL, ADRIAN: *Research Methods for Business Students*. Pearson Education (2012)

Sawyers / Canestaro (1989) SAWYERS, J. K.; CANESTARO, N. C.: *Creativity and achievement in design coursework*. In: Creativity Research Journal 2 (1989), Nr. 1-2

Schensul et al. (1999) SCHENSUL, STEPHEN L.; SCHENSUL, JEAN J.; LECOMPTE, MARGARET DIANE: *Essential Ethnographic Methods: Observations, Interviews, and Questionnaires*. Rowman Altamira (1999)

Schilling (2005) SCHILLING, M. A.: *A" small-world" network model of cognitive insight*. In: Creativity Research Journal 17 (2005), Nr. 2-3

Schlicksupp (1977) SCHLICKSUPP, HELMUT: *Kreative Ideenfindung in Der Unternehmung: Methoden Und Modelle*. Walter de Gruyter (1977)

Schregenberger (1980) SCHREGENBERGER, J. W.: *Methodenbewusstes Problemlösen*. Diss. Techn. Wiss. ETH Zürich, Nr. 6625, 1980, Ref.: Pozzi, A.; Korref.: Büchel, A., (1980)

Schregenberger (1983) SCHREGENBERGER, J. W.: *Probleme als Konstrukte*. In: Proceedings of ICED 83 Copenhagen (1983)

Schroda (2000) SCHRODA, F.: *Uber das Ende wird am Anfang entschieden*. In: Zur Analyse der Anforderungen von Konstruktionsauftragen. Technische Universität Berlin. Digitale Dissertation: http://edocs. tu-berlin. de/diss/2000/schroda_frauke. htm (2000)

Schuler / Görlich (2007) SCHULER, HEINZ; GÖRLICH, YVONNE: *Kreativität. Ursachen, Messung, Förderung und Umsetzung in Innovation*. Göttingen : Hogrefe (2007)

Seepersad et al. (2006) SEEPERSAD, CAROLYN C.; PEDERSEN, KJARTAN; EMBLEMSVAG, JAN; BAILEY, REID; ALLEN, JANET K.; MISTREE, FARROKH: *The validation square: how does one verify and validate a design method*. In: Decision-Based Design: Making Effective Decisions in Product and Systems Design (2006)

Shah / Vargas-Hernandez (2003) SHAH, JAMI J.; VARGAS-HERNANDEZ, NOE: *Metrics for measuring ideation effectiveness*. In: Design Studies 24 (2003), Nr. 2

Sheldon (2006) SHELDON, DEREK F.: *Design Review 2005/2006—The ever increasing maturity of design research papers and case studies*. In: Journal of Engineering Design 17 (2006), Nr. 6

Shiomi / Loo (1999) SHIOMI, K.; LOO, R.: *Cross-cultural response styles on the Kirton Adaption-Innovation Inventory*. In: Social Behavior and Personality: an international journal 27 (1999), Nr. 4

Skinner / Drake (2003) SKINNER, N. F.; DRAKE, J. M.: *Behavioral implications of adaption-innovation: III. Adaption-innovation, achievement motivation, and academic performance*. In: Social Behavior and Personality: an international journal 31 (2003), Nr. 1

Smith (1991) SMITH, N. C.: *The management research handbook*. Taylor & Francis (1991)

Sobek / Jain (2007) SOBEK, D.K., II; JAIN, V.K.: *Process Factors Affecting Design Quality: A Virtual Design of Experiments Approach*. In: Journal of Mechanical Design 129 (2007), Nr. 5

Van Someren et al. (1994) SOMEREN, M. W. VAN; BARNARD, Y. F.; SANDBERG, J. A. C.: *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press London (1994)

Song / Agogino (2004) SONG, S.; AGOGINO, A. M.: *Insights on designers' sketching activities in new product design teams*. In: Proceedings of the ASME Design Theory and Methods Conference (2004)

Spitzer (1975) SPITZER, R. L.: *On pseudoscience in science, logic in remission, and psychiatric diagnosis: A critique of Rosenhan's" On being sane in insane places".* (1975)

Spradley (1979) SPRADLEY, JAMES P.: *The ethnographic interview*. Holt, Rinehart and Winston (1979)

Sprent / Smeeton (2010) SPRENT, PETER; SMEETON, NIGEL C.: *Applied Nonparametric Statistical Methods, Fourth Edition*. CRC Press (2010)

Stachowiak (1973a) STACHOWIAK, H.: *Allgemeine Modelltheorie*. Springer (1973)

Stachowiak (1973b) STACHOWIAK, H.: *General model theory*. Springer (1973)

Stake (1995) STAKE, R. E.: *The art of case study research*. Sage Publications, Incorporated (1995)

Stamatis (2003) STAMATIS, DEAN H.: *Failure mode effect analysis: FMEA from theory to execution*. ASQ Quality Press (2003)

Stauffer et al. (1991) STAUFFER, LARRY A.; DITEMAN, MICHELE; HYDE, REVA: *Eliciting and Analysing Subjective Data about Engineering Design*. In: Journal of Engineering Design 2 (1991), Nr. 4

Sternberg (1995) STERNBERG, R. J.: *In search of the human mind.* Harcourt Brace College Publishers (1995)

Sternberg (1998) STERNBERG, ROBERT J.: *Handbook of Creativity*. Cambridge University Press (1998)

Sternberg (2006) STERNBERG, ROBERT J.: *The Nature of Creativity*. In: Creativity Research Journal 18 (2006), Nr. 1

Sternberg / Lubart (1995) STERNBERG, R. J.; LUBART, T. I.: *Defying the Crowd: Cultivating Creativity in a Culture of Conformity*. New York : NY, Free Press (1995)

Stetter (2000) STETTER, R.: *Method implementation in integrated product development*. (2000)

Stetter / Lindemann (2005) STETTER, R.; LINDEMANN, U.: *The transfer of methods into industry*. In: Design process improvement (2005)

Stevens et al. (1999) STEVENS, GREG; BURLEY, JAMES; DIVINE, RICHARD: *Creativity Business Discipline = Higher Profits Faster from New Product Development*. In: Journal of Product Innovation Management 16 (1999), Nr. 5

Stricker / Ross (1964) STRICKER, L. J.; ROSS, J.: *An assessment of some structural properties of the Jungian personality typology.* In: The Journal of Abnormal and Social Psychology 68 (1964), Nr. 1

Student (1908) STUDENT: *The probable error of a mean*. In: Biometrika (1908)

Stum (2009) STUM, J.: *Kirton's adaption-innovation theory: managing cognitive styles in times of diversity and change*. In: Emerging Leadership Journeys 2 (2009), Nr. 1

Suwa et al. (1998) SUWA, MASAKI; PURCELL, TERRY; GERO, JOHN: *Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions*. In: Design studies 19 (1998), Nr. 4

Suwa / Tversky (1997) SUWA, M.; TVERSKY, B.: *What do architects and students perceive in their design sketches? A protocol analysis*. In: Design studies 18 (1997), Nr. 4

Talbi (2009) TALBI, EL-GHAZALI: *Metaheuristics: From Design to Implementation*. Wiley (2009)

Taylor (1989) TAYLOR, W. G. K.: *The Kirton Adaption-Innovation Inventory: A re-examination of the factor structure*. In: Journal of Organizational Behavior 10 (1989), Nr. 4

Teegavarapu (2009) TEEGAVARAPU, SUDHAKAR: *Foundations of Design Method Development*. Clemson University. Mechanical Engineering, (2009)

Thompson / Borrello (1986a) THOMPSON, B.; BORRELLO, G. M.: *Construct validity of the Myers-Briggs type indicator*. In: Educational and Psychological Measurement 46 (1986), Nr. 3

Thompson / Borrello (1986b) THOMPSON, B.; BORRELLO, G. M.: *Second-Order Factor Structure of the MBTI: A Construct Validity Assessment.* In: Measurement and evaluation in counseling and development 18 (1986), Nr. 4

Torrance (1962) TORRANCE, E. P.: *Guiding creative talent.* (1962)

Torrance (1964) TORRANCE, E. P.: *ROLE OF EVALUATION IN CREATIVE THINKING.* (1964)

Torrance (1965) TORRANCE, E. PAUL: *Gifted Children in the Classroom.* , The MacMillan Company (1965)

Torrance (1979) TORRANCE, E. P.: *Unique needs of the creative child and adult*. In: The gifted and talented: Their education and development. 78th NSSE Yearbook (1979)

Torrance (1981) TORRANCE, E. PAUL: *Predicting the Creativity of Elementary School Children (1958-80)--and the Teacher Who "Made a Difference."* In: Gifted Child Quarterly 25 (1981), Nr. 2

Torrance (1995) TORRANCE, E.: *Insights about creativity: Questioned, rejected, ridiculed, ignored*. In: Educational Psychology Review 7 (1995), Nr. 3

Tovey et al. (2003) TOVEY, M.; PORTER, S.; NEWMAN, R.: *Sketching, concept development and automotive design*. In: Design Studies 24 (2003), Nr. 2

Treffinger et al. (2001) TREFFINGER, D.J.; FELDHUSEN, J.F.; RENZULLI, J.S.: *Statement for the public workshop on the Florida draft gifted rule.* , Unpublished Manuscript (2001)

Treffinger et al. (2002) TREFFINGER, D. J.; YOUNG, G. C.; SELBY, E. C.; SHEPARDSON, C.: *Assessing Creativity: A Guide for Educators.* In: National Research Center on the Gifted and Talented (2002)

Tröhler (2003) TRÖHLER, U: *James Lind and scurvy: 1747 to 1795*. , JLL Bulletin: Commentaries on the history of treatment evaluation (2003)

Tullett (2011) TULLETT, A. D.: *The adaptive—innovative (A—I) cognitive styles of male and female project managers: Some implications for the management of change*. In: Journal of Occupational and Organizational Psychology 68 (2011), Nr. 4

Turberville (2006) TURBERVILLE, A. S.: *Johnson's England: An Account of the Life & Manners of His Age*. Read Books (2006)

Usher (1996) USHER, J. M.: *Implementing concurrent engineering in small manufacturing enterprises*. In: Engineering Management Journal 8 (1996)

VDI (1993) VDI, FACHBEREICH PRODUKTENTWICKLUNG UND MECHATRONIK: *VDI 2221, Methodik zum Entwickeln und Konstruieren Technischer Systeme und Produkte*. In: Düsseldorf: VDI-GPP (1993)

VDI (1998) VDI: *2225 Blatt 3: Design engineering methodics - Engineering design at optimum cost - Valuation of costs*. In: : VDI-Handbuch Konstruktion (1998)

VDI (2003) VDI: *2230 Blatt 1: Systematische Berechnung hochbeanspruchter Schraubenverbindungen–Zylindrische Einschraubenverbindungen*. In: Verein Deutscher Ingenieure, Düsseldorf (2003)

Verein Deutscher Ingenieure (2004) VEREIN DEUTSCHER INGENIEURE: *VDI 2206 Design methodology for mechatronic systems*. , Beuth Verlag GmbH (2004)

Wallas (1926) WALLAS, GRAHAM: *The Art of Thought*. Harcourt, Brace (1926)

Weichselbauer (1996) WEICHSELBAUER, J.: *Erfolg von Reorganisationsprozessen – Leitfaden zur strategieorientierten Bewertung*. In: Stuttgart: Schäffer-Poeschel (1996)

Weimer (1979) WEIMER, W. B.: *Notes on the methodology of scientific research*. L. Erlbaum Associates (1979)

Wiedner (2013) WIEDNER, AARON JONAS: *"Feldstudie zur Identifikation der von Konstrukteuren praktizierten Handlungsmuster bei der Funktion-Gestalt-Synthese"*. IPEK - Institut für Produktentwicklung IPEK am Karlsruher Institut für Technologie, (2013)

Wilcoxon (1945) WILCOXON, FRANK: *Individual Comparisons by Ranking Methods*. In: Biometrics Bulletin 1 (1945), Nr. 6

Wildemann (1993) WILDEMANN, H.: *Optimierung von Entwicklungszeiten: Just-In-Time in Forschung& Entwicklung und Konstruktion*. Transfer-Centrum-Verlag (1993)

Wildemann et al. (2007) WILDEMANN, HORST; ANN, CHRISTOPH; BROY, MANFRED; GÜNTHNER, W.; LINDEMANN, U.: *Plagiatschutz–Handlungsspielräume der produzierenden Industrie gegen Produktpiraterie*. In: Konstruktion 10 (2007)

Williams (2010) WILLIAMS, CHUCK: *Management*. Cengage Learning (2010)

Winker (2006) WINKER, PETER: *Empirische Wirtschaftsforschung und Ökonometrie*. 2. vollständig überarb. Aufl. Aufl., Springer (2006)

Wittgenstein (1921) WITTGENSTEIN, LUDWIG: *Tractatus logico-philosophicus*. DF Pears and BF McGuinness, Trans.). London: Routledge & Kegan Paul (1921)

Witzel / Reiter (2012) WITZEL, ANDREAS; REITER, HERWIG: *The Problem-Centred Interview*. SAGE (2012)

Wögerbauer (1942) WÖGERBAUER, HUGO: *Die Technik des Konstruierens*. München, Oldenbourg Verlag, (1942)

Woodman et al. (1993) WOODMAN, R. W.; SAWYER, J. E.; GRIFFIN, R. W.: *Toward a theory of organizational creativity*. In: Academy of management review (1993)

Wu et al. (2011) WU, S.; HOFMAN, J. M.; MASON, W. A.; WATTS, D. J.: *Who says what to whom on twitter*. In: Proceedings of the 20th international conference on World wide web (2011)

Yang (2009) YANG, M. C.: *Observations on concept generation and sketching in engineering design*. In: Research in Engineering Design 20 (2009), Nr. 1

Yin (1994) YIN, R. K.: *Case study research: Design and methods*. . 2nd edition Sage Publications, Incorporated (1994)

Zanker (1999) ZANKER, WINFRIED: *Situative Anpassung und Neukombination von Entwicklungsmethoden*. Shaker (1999)

Zucchini et al. (2009) ZUCCHINI, WALTER; SCHLEGEL, ANDREAS; NENADIC, OLEG; SPERLICH, STEFAN: *Statistik für Bachelor-und Masterstudenten: Eine Einfuhrung für Wirtschafts-und Sozialwissenschaftler*. 2009. Aufl., Springer (2009)

9 Appendix

9.1 Definitions and Explanations

9.1.1 Algorithmic Methods - Heuristic Methods - Heuristics

This thesis focuses on design science, a field within science that is still very young, as has been pointed out in the introduction. One of the difficulties when dealing with something young and immature is that there are usually many terms that still lack one commonly accepted definition. Instead, many similar yet different definitions occur in different publications. This chapter will clarify my understanding of the term “heuristic”.

9.1.2 Algorithmic Methods

Algorithmic methods are all those methods that provide clear and specific step by step instructions. This does not necessarily mean that when applied, an algorithmic method will lead the user through the exact same steps each time it is executed. Boolean operators such as if-then relations can also be part of the method. The key point is that no type of creative decision making through the user is made, which leads to high reliability. Typical examples are methods for dimensioning parts as can be found in industrial guidelines.[515] In many of the natural sciences, the term method is used with such an understanding since in that field, methods that would include human judgment are extremely uncommon – bio analytics or chemistry are just two examples.

9.1.3 Heuristic Methods

The term “heuristics” is derived from the Greek word „heurískein“, which means “to find” or “to discover”. Its connotation is closely linked to problem-solving activities. Some specify it to problem solving with limited information.[516] An heuristic can be described as a strategy or course of action to find an acceptable solution for a certain class of problems. In design engineering, all methods that place the designer as a creative human being in the centre of the approach and include his or her decisions in the course of action, will not guarantee an optimal solution. However, they aim to provide a certain degree of direction, boundaries and decision support for finding a good solution. Therefore, those methods are referred to as heuristic methods. There is no clear definition for this term to refer to, but only approaches to describe what is meant by the term heuristic method. Therefore, following definition is suggested:

An heuristic method in product design is a set of rules that help designers proceed successfully in their problem-solving activities. Applying an heuristic method will

---

[515] See for example dimensioning guidelines for the calculation of screw joints VDI (2003).

[516] G. Gigerenzer is one of the most popular, modern advocates of this definition. Compare e.g. Gigerenzer / Todd (1999), Gigerenzer / Goldstein (1996)

involve decision making by a human being. Success is achieved when a good solution is found.

This human-centered definition is quite contrary to any definition that occurs within the context of artificial intelligence, where heuristics are used to replace human decision making by computers.[517] Since the goal of artificial intelligence is to mimic human behavior with computers, however, it seems rather suitable.

Examples for typical heuristic methods are all types of creativity encouraging methods like brainstorming and all its derivatives, synectics and other methods based on association and also methods to support selection of solutions, e.g. pro-contra-lists or fast-and-frugal-trees.

9.1.4 Heuristics in other scientific disciplines

Some readers who might have dealt with heuristics, heuristic methods or heuristic optimization might possibly disagree with the description above. In economics, there are a range of optimization problems that cannot be solved in terms of finding the one optimum as in a global optimum. However, there are algorithms, helping one to find a good solution. These are often called heuristics as well. Although optimization is an important part of product design, the economically driven definitions are not suitable for the types of problems addressed in this thesis and are therefore, only briefly presented here.[518] The definition for such local improvement heuristics in this context is that they cannot guarantee to deliver a global optimum but will find a good solution.[519] Prominent optimization problems are e.g. “The traveling salesman”[520], or the “Knapsack Problem”[521]. Other typical types of heuristics in economics are starting point heuristics applied to determine a decent starting point for an optimization algorithm to be efficient; heuristic strategies apply different types of optimization methods or change the parameters of a certain optimization algorithm, in iterations, to make sure it is e.g. doing a broad search for the beginning, not missing any local optima and only towards the end narrowing down the “most promising” optima. These types are also called meta heuristics.

---

[517] J. Weizenbaum, Computer Power and Human Reason, 1976; John McCarthy, WHAT IS ARTIFICIAL INTELLIGENCE?, 2007, digital resource: http://www-formal.stanford.edu/jmc/whatisai/node1.html, 11/2/2012, 1:40 pm

[518] For further reading on heuristics in economic context, see Michalewicz / Fogel (2004); Talbi (2009); Rardin / Uzsoy (2001)

[519] IBM defines heuristics in CPLEX as: “a procedure that tries to produce good or approximate solutions to a problem quickly but which lacks theoretical guarantees. In the context of solving a [mixed integer programming problem], an heuristic is a method that may produce one or more solutions, satisfying all constraints and all integrality conditions, but lacking an indication of whether it has found the best solution possible.” http://publib.boulder.ibm.com/infocenter/cosinfoc/v12r3/topic/ilog.odms.cplex.help/Content/Optimization/Documentation/Optimization_Studio/_pubskel/ps_usrmancplex1844.html; 3/3/2012, 12:58

[520] Popular logistics problem - a salesman trying to visit his customers with minimum travelled distance

[521] How does one pack a limited size backpack with items of different benefit and weight?

9.2 Worksheets

Worksheet 1 : General advice for design researchers (derived from Bender et al. (2002))

- Include application-experience of others in one's own reading
- If there is a lack of theory, refrain from a hypothesis-driven approach. Prefer the formulation of research questions directly, or based on exploratory research.
- Interdisciplinary teams may be helpful in finding hypotheses and explanations because of availability and combination of a large number of theories.
- To deal with the asymmetry of empirical relevance, on the one hand, and statistical significance on the other, extend the quantitative approach to data collection and analysis by qualitative methods. It can be helpful to explicitly aim at answering the question: "What may I conclude from this result with certain likelihood and therefore, have to take into account for real design practice?" rather than "What will definitely happen within design practice as a conclusion of my results?"
- Give detailed descriptions of the setup of the study, your analysis and interpretation methods, and *make all assumptions explicit* to ensure that the study can be understood and the results traced.
- Look at the target group! Who are the potential test persons? Which direct or indirect benefits can they expect from taking part? In particular, experts from industry, who are under severe time-pressure, have to be convinced of the research objectives before taking part.
- Multi-method-approaches increase validity; it is best to combine qualitative and quantitative approaches.
- Increase validity of results by using appropriate analysis methods
- For small sample sizes and strong interconnectivity of variables, *dichotomization* of statistical populations and *non-parametric statistics* are promising.
- Increase validity of study by applying fundamental rules of test design and test analysis.[522]
- To ensure homogeneity and inherent consistency of design tasks (not only) for laboratory studies, the taxonomy developed by SCHRODA[523] might help.
- Establish causality between design success (e.g. in terms of design quality) and co-varying characteristics that have been gathered retrospectively, based on co-variation, time period and the exclusion of spuriousness.[524]
- Use valid methods for evaluating design success / design quality; follow a systematic evaluation process to rank designs;
- Estimate and document evaluation uncertainties

---

[522] Lienert / Raatz (1998), 29 ff.

[523] Schroda (2000)

[524] For details see also Blessing / Baumgaertner (2001)

Worksheet 2: Collected questions to assess credibility (cp. Keller / Binz (2009))

	Question for self-check
Operation System	Have valid procedures been used when developing the support?
	Has potential bias been reduced?
System of Objects	Does the support do things right?
	Does the support do the right things?
	Have the findings the support is based on been validated?
	Is the support set in a way that it does not bias its user towards certain solutions?
	Can the results be reproduced?
	Is the support documented in an understandable way
	Is it presented in a way that acceptance is likely?

Worksheet 3: General objectives for design support development

Design support should…		
	be simple.	Geis et al. (2008)
	be flexibly applicable.	
	focus on the output .	
	be better integrated in the process.	
	be adaptable to the wishes of the users.	
	Be provided with computer assistance if available.	Grabowski / Geiger (1997)
	include information about other methods/ processes it can ideally be combined with.	
	suit human problem solving behavior.	Jänsch (2007)
	include information about the intended industrial sector/ type(s) of product it was developed for.	
	come with hints and advice for its proper application.	
	come with a description of the expected benefit of its usage.	
	focus on practicability.	
	Inform about required knowledge/expertise.	
	Be documented using identical and homogenous terms.	Horváth (2004)
Design support should not…		
	need too much effort.	Grabowski / Geiger (1997)
	carry too much "theoretical ballast".	
	be presented with lack of preparation and support for its application of the methods.	
	use heterogeneous terms.	Jänsch (2007)
	mix / assume different paradigms.	
	be too theoretical.	
	have inconsistent representation.	
	have an unnecessary high level of complexity.	
	have an unnecessary high level of abstraction.	
The lure of dangerous promises		
	need as little effort for learning and training as possible.	Birkhofer (2004)
	Is easy to use.	
	solves problems "in no time".	
	produces convincing results for complex problems.	
	Is integrated in the existing design environment.	

Worksheet 4: Meta Targets for the System of Objectives

General questions for self-check :	**Source**
In which aspects is the support new?	Keller / Binz (2009)
Which existing design support does it compete with/replace?	
Why do you believe it will be more successful?	
To what extend can improvements be expected if the method is applied?	
What is the result one may expect/desire when applying the support?	
For which types of problems is the support intended?	
For which activities is the support intended?	
To what extend is the support heuristic/algorithmic?	
Who is to benefit from the support?	
Interaction - Is the support supposed to improve:	
the speed of communication?	Geis et al. (2008)
the effectiveness of communication?	
designer's competent and objective presentation and discussion of their ideas?	
reaching agreement in groups?	
Planning & organization - Is the support supposed to:	
help in planning, organizing and controlling projects or processes?	Geis et al. (2008)
support analysis of the process?	
ensure sustainability of actions and measures?	
support individual time- and project-management?	
help prioritize work quotas?	
help improve processes?	Grabowski / Geiger (1997)
help reduce iterative loops?	
help visualize existing knowledge?	
help reach targeted cost?	
help reach targeted deadlines?	
save time and/or cost?	
help to make technical and organizational decisions?	
support reaching customer- and goal-oriented decisions?	
support in accessing linked information?	
Proper Documentation – Does the support:	
help with systematic and structured documentation?	Grabowski / Geiger (1997)
include information about other methods/ processes it can ideally be combined with?	
belong to a larger/holistic framework of methods it can be combined with?	

Worksheet 5: Considerations for experimental Design

In the setup of experimental studies, principally, there are two dimensions that can be varied or kept constant: the experimental subjects, and the problem. Each combination will have its own benefits and problems concerning bias. An overview over the main biasing influences is given in the table below. A detailed explanation of benefits and problems of each combination is given thereafter.

	Repeating experiment with one problem	**Varying the problem**
Repeating experiment with the same people	Memory and experience when experiment is conducted the second time	Different degree of experience for the problems and different degree of difficultness
Varying the experimental subjects	Beforehand knowledge about the problem in one group. Comparability of the intelligence / problem solving capabilities of the different groups.	Different degree of experience for the problems and different degree of difficultness. Comparability of the intelligence / problem solving capabilities of the different groups.

Repeating an experiment with the same group of experimental subjects with one single problem

This combination underlies the strongest degree of bias and should be avoided. Experience and learning strongly affect the experimental results. The memory of the participants cannot be erased, so when applied a second time, better results can be expected, or the solution will be given right away. If the participants are forced to apply the method, they might pretend to use it and instead actually act from memory.

Possible countermeasures:

Theoretically, the problem could be reduced, when there is a long time span in between two experiments. However, the definition of the term “long” depends on the personal memory of the participants. There is no means of indicating when one has waited long enough and waiting for a long time is not practical for the researcher.

The amount of time can be reduced, if the subject's memory is put under stress. Making the subjects concentrate on something else might lead them to forget parts of the solution. However, the effect cannot be quantified, making it hard to judge the quality of the experimental results.

Repeating an experiment with the same group of but varying the problem

If the problem is altered, the participants will not have the solution memorized. However, some individuals might have some experience with one of the experimental problems. This can lead to either exaggerating the effect a tested design support has (if the problem tested with the support is already known), or it might reduce the effect (if the problem without the support is known). Even if there is no prior experience with the tasks, it is still possible that one of them is easier to solve for the participants thus biasing the observed effect of the method.

Possible countermeasures:

	I have solved it before.	Heard about it but never worked on it.	I have encountered a similar task before.	It is completely new to me.
How familiar are you with the task?				

The last problem can be reduced, if the degree of difficulty could be quantified beforehand (use e.g. Worksheet 6). The researcher would have to assure that only comparably difficult pro-blems are used for the assessment of the method. Concerning the possible experience, two different tactics can be applied to reduce the problem: One option is to ask the participants how familiar they are with the problems on an ordinal scale as shown in the example and only use the data-sets in which the participants feel equally familiar with both problems. Alternatively, the familiarity assessment could be used for a “Matched Pairs Test” or a “Signed-Rank Test Rank Test for matched Samples” (both statistical procedures are described in chapter 4.3.2.7)
Alternatively, the researcher can try and normalize the prior experience of the group, giving them extensive information about each problem beforehand. Rising the average experience-level of the group for all problems, the effect of the group being acquainted with a single problem is reduced.

Varying the experimental subjects while sticking to one problem
Presenting the same problem to varying experimental subjects, the results will provide for better comparability. However, it confronts the researcher with a new problem. Different individuals will be subject to differing levels of intelligence, design experience, and creativity. Also, some individuals might be familiar with the presented.
Possible countermeasures:
The researcher can reduce the effects by assessing the individuals’ characteristics (intelligence, design experience, and creativity) as well as familiarity with the problem through a pre-test and:

- make sure to put together comparable groups or,
- use Matched Pairs / Signed-Rank Test for matched Samples (see chapter 4.3.2.7)

Another approach is to assure for a large number of data-sets and randomizing the groups. Statistically, individual (dis)advantages within the total of the data-sets are then no longer significant.
Alternatively, the researcher can try to normalize the prior experience of all individuals, giving them extensive information about the problem beforehand. Rising the average familiarity-level of the whole group. The effect of single individuals being acquainted with the problem is reduced. At the same time, however, the experiment becomes less realistic, since under normal circumstances, such additional information might be uncommon.

Varying the experimental subjects and altering the problem
This approach makes it most difficult to compare the experimental results and is only useful, if a large number of experiments can be conducted. The variation of the experimental subjects still provides for the problem that individuals have different degrees of creativity and intelligence, while the variation of the problem may lead to different levels of difficulty. Bias through prior experience is statistically less likely but even harder to quantify.
Possible countermeasures:
The researcher can reduce the effects by assessing the individuals’ intelligence, creativity and familiarity with the problem through a pre-test and

- put together comparable groups or
- use Matched Pairs / Signed-Rank Test for matched Samples (see chapter 4.3.2.7)

The task's degree of difficulty needs to be quantified beforehand (Worksheet 6). The researcher has to make sure that only comparably difficult problems are used for the experiment. Alternatively, the group’s familiarity with the task can be normalized by providing additional information. When varying the problem there is a further difficulty. The researcher has to make sure, to give comparable information and hints for the different problems.

Worksheet 6: Evaluation taxonomy for design tasks (based on Schroda (2000), p 160 ff.)

Date: Task:		very few/ very little	few/ little	medium	many/ strong	very many/ very strong
Conflicting goals	How many goals does the task contain					
	How many conflicting goals does the task contain					
	How strong do the goals conflict?					
Complexity	How many different sub functions must be considered in the task?					
	How many interrelations are between the sub functions?					
	How strong are the sub functions interrelated?					
Transparency	How much information about constraints and conditions is available?					
	How much information about the solution process is available?					
	How strong is the final and desired solution defined?					
Degrees of freedom	How many sensible solution alternatives are possible?					
	How many sensible solution processes are possible?					
Knowledge	How much expert knowledge is necessary?					
	How much methodological knowledge is necessary?					

Compare Schroda (2000), p 160 ff.

Worksheet 7: Quality Checklist for Task design

This checklist is intended for researchers planning a design experiment. Its contents are adopted from Bender (2003) who draws on the ideas of Lienert / Raatz (1998). See chapter 4.2.1 for a more comprehensive explanation.

Cri-terion	Explanation	Self-check	Possible countermeasures
Objectivity task performance evaluation	When conducting a design experiment, the design outcome has to be ranked/ evaluated. This can be influenced by personal preferences of the evaluator(s), hence it is a possible cause of bias.	▪ How am I evaluating the test performance? ▪ Would someone else come to the same results me? ▪ Are several evaluators involved? ▪ Would I come to the same results repeatedly?	▪ Predefined Evaluation metrics ▪ Training of evaluators ▪ Pre-study to check evaluator's performance for conflicting results ▪ Redundant evaluation through two or more evaluators per design outcome. ▪ Group decisions
Reliability of task description	A verbalized task is subject to interpretation by the participant. It must be avoided to have unclear task descriptions that leave room for subjective speculation/ interpretation.	▪ Would the same participant understand the task the same way repeatedly (e.g. different days)? ▪ Would two different individuals understand the task and know what is demanded of them the same way? ▪ If the task is given to the same person more than once with intended variations, are truly only those parameters varied that I wanted to vary?	▪ Precise task description ▪ Avoid free oral task description. Prepare written task description. ▪ Pre study to observe interpretation of the task by different individuals.
Validity of test results	Validity of results refers to the *certainty* with which the researcher can know the quality of a result.	▪ What are the cause and effect that will lead to poor/good test results? ▪ Can multiple causes lead to the same test result? ▪ Can under any circumstances the same cause lead to contradicting test results?	▪ Precise and operational performance criteria ▪ Isolation of the parameters under investigation ▪ Ideally, a reference solution should exist.

Empirical relevance of the task	Refers to the transferability of results. It has to be made sure that the laboratory situation can be compared to real-world product development practice.	▪ What real-world conditions would have a major influence on the task? ▪ Are all relevant real-world influences included into the laboratory situation? ▪ Do the participants start with the same information in the experiment as they would start with in reality? ▪ Do the participants feel comfortable as if in a natural situation, concentrating on the task and not the surroundings?	▪ Pilot descriptive study to ensure that similar situations actually occur in practice. ▪ Consultation of practitioners prior to experiment. ▪ Integration of participants that have practical experience.
Adequate difficulty of the task	It is important to maintain the balance between not being too "trivial" (compare empirical relevance) and not frustrating the participants with a task that is too hard.	▪ Have I made sure the participants know what is expected of them? ▪ Is it possible that participants feel uncomfortable under observation and think they are being tested? ▪ Is the test trivial and will this lead to doubt about the relevance of the results? ▪ Is it possible to finish the task in the given time? ▪ Do the participants really have the necessary qualifications to complete the task? ▪ Have I provided the necessary resources and information to complete the task?	▪ Pretest with a small number of participants in order to determine how they feel about the degree of difficulties. ▪ Use past tasks that have proven to have a good degree of difficulty. ▪ Compare with established standard tasks, if available. ▪ Inform the participants that this is not a test of their personal skills. ▪ Ensure that participants do not interpret the situation as a competition. ▪ If the task is (intended) unsolvable, participants should be informed to avoid the feeling of personal failure.
Efficiency	It has to be ensured that the captured data and the involved participants are kept at a minimum while still providing enough data for the experiments to be effective.	▪ How much data can I analyze? ▪ How many cases do I really need to come to a result? ▪ Is more data going to change the results or is it just making me feel more comfortable trusting the results? ▪ What questions should be addressed with the experiment and what is not part of the experiment?	▪ Calculations prior to the experiments to determine the necessary number of test cases that lead to a desired statistical validity. ▪ Pre-tests and examination of the data to get an estimate value for the total amount of time the data acquisition and analysis will take. ▪ Clear definition of the targets that are to be addressed with the experiment.

Worksheet 8: Guideline for posing Questions (compare Atteslander (2008), 146)[525]

- Composed questions of simple expressions; avoid uncommon technical terms, loanwords, abbreviations and slang.
- Keep them short.
- Be concrete: “How satisfied are you with your current working situation?” is better than the question “How happy are you with your life?” Transform general and abstract expressions into concrete and precise terms.
- Never provoke an answer (leading questions). “Have you seen ‘Forrest Gump’?” is better than “Everybody has seen ‘Forrest Gump’; I am sure you have too, right?”
- Be neutral; avoid terms with a negative / positive connotation (e.g. 'bureaucratic', 'malfunction', 'freedom', 'integrity' and so on)
- Avoid hypothetical questions, such as “Let’s say you won the lottery. Would you spend it all or save some?”
- One question may only address one issue. Avoid multidimensional questions. “Would you use a project management software to plan and coordinate a research project but avoid using it for private projects such as planning a wedding?” The question addresses two subjects and therefore, it must be divided into two questions.
- Avoid double negatives.
- Make answering easy for the interviewee. “Which percentage of your monthly net salary do you spent on rent?” is difficult as the interviewee has to do the math. Instead ask: “How much do you earn?” then “How much do you pay for rent?”
- Questions should be balanced. Verbalize both positive and negative answers in the question to demonstrate that both are equally legitimate. Prefer: “Should engineers from production be included in decision-making processes for product design or should this be done without them.” Avoid: “Would you include the production engineers in design decisions?” It is human nature that “Yes” is easier than “No”.

---

[525] Some of the examples have been translated and adapted to an engineering design context. The original list can be found at: Atteslander (2008), 146

9.2.1 Code of fair Testing Practices in Education

The JOINT COMMITTEE ON TESTING PRACTICES, developed a summary code of fair testing practices to guide test developers and test users. It has its origin in education and tests in this context are actually tests as exams in school. However, the underlying considerations might be useful for those developing tests and task for the evaluation of design support. After all, in many cases, student projects in university setting are used for design support evaluation. The following text is not copyrighted, and its dissemination is encouraged. This is a reproduction of the original text which is also available online at the following source: http://ericae.net/code.txt, or by mail from the American Psychological Association.

Worksheet 9 : Code of fair Testing Practices in Education

Code of Fair Testing Practices in Education. (1988)
Washington, D.C.: Joint Committee on Testing Practices.

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed primarily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The Code is not intended to cover tests made by individual teachers for use in their own classrooms.

The Code addresses the roles of test developers and test users separately. Test users are people who select tests, commission test development services, or make decisions on the basis of test scores. Test developers are people who actually construct tests as well as those who set policies for particular testing programs. The roles may, of course, overlap as when a state education agency commissions test development services, sets policies that control the test development process, and makes decisions on the basis of the test scores.

The Code has been developed by the Joint Committee on Testing Practices, a cooperative effort of several professional organizations, that has as its aim the advancement, in the public interest, of the quality of testing practices. The Joint Committee was initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. In addition to these three groups the American Association for Counseling and Development/Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association are now also sponsors of the Joint Committee.

This is not copyrighted material. Reproduction and dissemination are encouraged. Please cite this document as follows:

Code of Fair Testing Practices in Education. (1988)
Washington, D.C.: Joint Committee on Testing Practices.

(Mailing Address: Joint Committee on Testing Practices,
American Psychological Association, 1200 17th Street, NW,
Washington, D.C. 20036.)

The Code presents standards for educational test developers and users in four areas:

A. Developing/Selecting Tests
B. Interpreting Scores
C. Striving for Fairness
D. Informing Test Takers

Organizations, institutions, and individual professionals who endorse the Code commit themselves to safeguarding the rights of test takers by following the principles listed. The Code is intended to be consistent with the relevant parts of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1985). However, the Code differs from the Standards in both audience and purpose. The Code is meant to be understood by the general public; it is limited to educational tests; and the primary focus is on those issues that affect the proper use of tests. The Code is not meant to add new principles over and above those in the Standards or to change the meaning of the Standards. The goal is rather to represent the spirit of a selected portion of the Standards in a way that is meaningful to test takers and/or their parents or guardians. It is the hope of the Joint Committee that the Code will also be judged to be consistent with existing codes of conduct and standards of other professional groups who use educational tests.

A. DEVELOPING/SELECTING APPROPRIATE TESTS*

Test developers should provide the information that test users need to select appropriate tests.	Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test taking populations.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
1. Define what each test measures and what the test should be used for. Describe the population(s) for which the test is appropriate.	1. First define the purpose for testing and the population to be tested. Then, select a test for that purpose and that population based on a thorough review of the available information.
2. Accurately represent the useful characteristics, usefulness, and limitations of tests for their intended purposes.	2. Investigate potentially sources of information, in addition to test scores, to corroborate the information provided by tests.
3. Explain relevant measurement concepts by as necessary for clarity at the level of detail that is appropriate for the intended audience(s).	3. Read the materials provided test developers and avoid using tests for which unclear or incomplete information is provided.
4.Describe the process of test development. Explain how the content and skills to be tested were selected.	4. Become familiar with how and when the test was developed and developed and tried out.

5. Provide evidence that the test meets its intended purpose(s).	5. Read independent evaluations of a test and of possible alternative measures. Look for evidence required to support the claims of test developers.
6. Provide either representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.	6. Examine specimen sets, disclosed tests or samples of questions, directions, answer sheets, manuals, and score reports before selecting a test.

*Many of the statements in the Code refer to the selection of existing tests. However, in customized testing programs test developers are engaged to construct new tests. In those situations, the test development process should be designed to help ensure that the completed tests will be in compliance with the Code.

TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
7. Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.	7. Ascertain whether the test content and norm group(s) or comparison group(s) are appropriate for the intended test takers.
8. Identify and publish any specialized and to interpret scores correctly.	8. Select and use only those tests for which the skills needed to administer the test and interpret scores correctly are available.

B. INTERPRETING SCORES

Test developers should help users interpret scores correctly	Test users should interpret scores correctly.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
9. Provide timely and easily understood score reports that describe test performance clearly and accurately. Also, explain the meaning and limitations of reported scores.	9. Obtain information about the scale used for reporting scores, the characteristics of any norms or comparison group(s), and the limitations of the scores.
10. Describe the population(s) represented by any norms or comparison group(s), the process used to select the samples of	10. Interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers.

dates the data were gathered, and the takers.	Also take into account any differences in test administration practices or familiarity with the specific questions in the test.
11. Warn users to avoid specific, reasonably anticipated misuses of test scores.	11. Avoid using tests for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use.
12. Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.	12. Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.
13. Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).	13. Obtain evidence to help show that the test is meeting its intended purpose(s).

C. STRIVING FOR FAIRNESS

Test developers should strive to make tests that are as fair as possible for test takers of different races, gender, ethnic backgrounds, or different handicapping conditions.	Test users should select tests that have been developed in ways that attempt to make them as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
insensitive content or language	

and related materials to avoid potentially used 14. Review and revise test questions	14. Evaluate the procedures by test developers to avoid . potentially insensitive content or language.
15. Investigate the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.	15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused of the test.
16. When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in non-comparable scores.	16. When necessary and feasible, use appropriately modified forms or administration procedures for test takers with handicapping conditions. Interpret standard norms with care in the light of the modifications that were made.

D. INFORMING TEST TAKERS

Under some circumstances, test developers have direct communication with test takers. Under other circumstances, test users communicate directly with test takers. Whichever group communicates directly with test takers should provide the information described below.

TEST DEVELOPERS OR TEST USERS SHOULD:

17. When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.

18. Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.

Under some circumstances, test developers have direct control of tests and test scores. Under other circumstances, test users have such control. Whichever group has direct control of tests and test scores should take the steps described below.

TEST DEVELOPERS OR TEST USERS SHOULD:

19. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.

20. Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.

21. Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.

Note: The membership of the Working Group that developed the Code of Fair Testing Practices in Education and of the Joint Committee on Testing Practices that guided the Working Group was as follows:

Theodore P. Bartell
John R. Bergan
Esther E. Diamond
Richard P. Duran
Lorraine D. Eyde
Raymond D. Fowler
John J. Fremer
(Co-chair, JCTP and Chair, Code Working Group)
Edmund W. Gordon
Jo-Ida C. Hansen
James B. Lingwall
George F. Madaus
(Co-chair , JCTP)
Kevin L. Moreland
Jo-Ellen V. Perez
Robert J. Solomon
John T. Stewart
Carol Kehr Tittle
(Co-chair, JCTP)
Nicholas A. Vacc
Michael J. Zieky

(Debra Boltas and Wayne Camara of the American Psychological Association served as staff liaisons)

Additional copies of the Code may be obtained from the National Council on Measurement in Education, 1230 Seventeenth Street, NW, Washington, D.C. 20036.

Single copies are free.

9.3 Tables

9.3.1 Table of the Standard Normal Distribution

This table was generated with Microsoft Excel. Standard Distribution Tables can be found online or in any standard statistics book, e.g. Anderson / Finn (1996).

Negative z (-3,59 ≤ z ≤ 0)

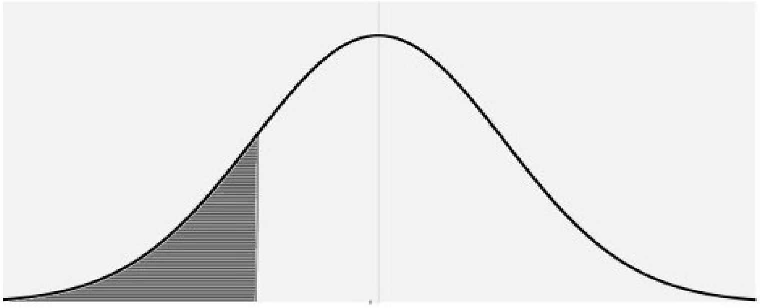


Table 66 - Table of the Standard normal distribution for negative z)

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,5	0,000233	0,000224	0,000216	0,000208	0,0002	0,000193	0,000185	0,000178	0,000172	0,000165
-3,4	0,000337	0,000325	0,000313	0,000302	0,000291	0,00028	0,00027	0,00026	0,000251	0,000242
-3,3	0,000483	0,000466	0,00045	0,000434	0,000419	0,000404	0,00039	0,000376	0,000362	0,000349
-3,2	0,000687	0,000664	0,000641	0,000619	0,000598	0,000577	0,000557	0,000538	0,000519	0,000501
-3,1	0,000968	0,000935	0,000904	0,000874	0,000845	0,000816	0,000789	0,000762	0,000736	0,000711
-3	0,00135	0,001306	0,001264	0,001223	0,001183	0,001144	0,001107	0,00107	0,001035	0,001001
-2,9	0,001866	0,001807	0,00175	0,001695	0,001641	0,001589	0,001538	0,001489	0,001441	0,001395
-2,8	0,002555	0,002477	0,002401	0,002327	0,002256	0,002186	0,002118	0,002052	0,001988	0,001926
-2,7	0,003467	0,003364	0,003264	0,003167	0,003072	0,00298	0,00289	0,002803	0,002718	0,002635
-2,6	0,004661	0,004527	0,004396	0,004269	0,004145	0,004025	0,003907	0,003793	0,003681	0,003573
-2,5	0,00621	0,006037	0,005868	0,005703	0,005543	0,005386	0,005234	0,005085	0,00494	0,004799
-2,4	0,008198	0,007976	0,00776	0,007549	0,007344	0,007143	0,006947	0,006756	0,006569	0,006387
-2,3	0,010724	0,010444	0,01017	0,009903	0,009642	0,009387	0,009137	0,008894	0,008656	0,008424
-2,2	0,013903	0,013553	0,013209	0,012874	0,012545	0,012224	0,011911	0,011604	0,011304	0,011011
-2,1	0,017864	0,017429	0,017003	0,016586	0,016177	0,015778	0,015386	0,015003	0,014629	0,014262
-2	0,02275	0,022216	0,021692	0,021178	0,020675	0,020182	0,019699	0,019226	0,018763	0,018309
-1,9	0,028717	0,028067	0,027429	0,026803	0,02619	0,025588	0,024998	0,024419	0,023852	0,023295
-1,8	0,03593	0,035148	0,03438	0,033625	0,032884	0,032157	0,031443	0,030742	0,030054	0,029379
-1,7	0,044565	0,043633	0,042716	0,041815	0,04093	0,040059	0,039204	0,038364	0,037538	0,036727
-1,6	0,054799	0,053699	0,052616	0,051551	0,050503	0,049471	0,048457	0,04746	0,046479	0,045514
-1,5	0,066807	0,065522	0,064255	0,063008	0,06178	0,060571	0,05938	0,058208	0,057053	0,055917
-1,4	0,080757	0,07927	0,077804	0,076359	0,074934	0,073529	0,072145	0,070781	0,069437	0,068112
-1,3	0,0968	0,095098	0,093418	0,091759	0,090123	0,088508	0,086915	0,085343	0,083793	0,082264
-1,2	0,11507	0,113139	0,111232	0,109349	0,107488	0,10565	0,103835	0,102042	0,100273	0,098525
-1,1	0,135666	0,1335	0,131357	0,129238	0,127143	0,125072	0,123024	0,121	0,119	0,117023
-1	0,158655	0,156248	0,153864	0,151505	0,14917	0,146859	0,144572	0,14231	0,140071	0,137857
-0,9	0,18406	0,181411	0,178786	0,176186	0,173609	0,171056	0,168528	0,166023	0,163543	0,161087
-0,8	0,211855	0,20897	0,206108	0,203269	0,200454	0,197663	0,194895	0,19215	0,18943	0,186733
-0,7	0,241964	0,238852	0,235762	0,232695	0,22965	0,226627	0,223627	0,22065	0,217695	0,214764
-0,6	0,274253	0,270931	0,267629	0,264347	0,261086	0,257846	0,254627	0,251429	0,248252	0,245097
-0,5	0,308538	0,305026	0,301532	0,298056	0,294599	0,29116	0,28774	0,284339	0,280957	0,277595
-0,4	0,344578	0,340903	0,337243	0,333598	0,329969	0,326355	0,322758	0,319178	0,315614	0,312067
-0,3	0,382089	0,37828	0,374484	0,3707	0,366928	0,363169	0,359424	0,355691	0,351973	0,348268
-0,2	0,42074	0,416834	0,412936	0,409046	0,405165	0,401294	0,397432	0,39358	0,389739	0,385908
-0,1	0,460172	0,456205	0,452242	0,448283	0,44433	0,440382	0,436441	0,432505	0,428576	0,424655
0	0,5	0,496011	0,492022	0,488034	0,484047	0,480061	0,476078	0,472097	0,468119	0,464144

Positive z (0 ≤ z ≤ 3,59)

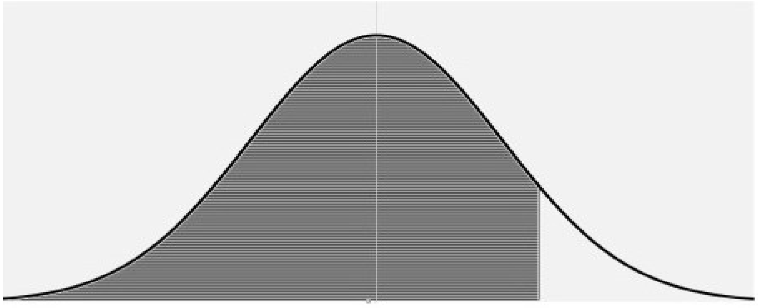


Table 67 - Table of the Standard normal distribution for positive z)

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,503989	0,507978	0,511966	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,55567	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,57926	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,60642	0,610261	0,614092
0,3	0,617911	0,62172	0,625516	0,6293	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705401	0,70884	0,71226	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,77035	0,773373	0,776373	0,77935	0,782305	0,785236
0,8	0,788145	0,79103	0,793892	0,796731	0,799546	0,802337	0,805105	0,80785	0,81057	0,813267
0,9	0,81594	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1	0,841345	0,843752	0,846136	0,848495	0,85083	0,853141	0,855428	0,85769	0,859929	0,862143
1,1	0,864334	0,8665	0,868643	0,870762	0,872857	0,874928	0,876976	0,879	0,881	0,882977
1,2	0,88493	0,886861	0,888768	0,890651	0,892512	0,89435	0,896165	0,897958	0,899727	0,901475
1,3	0,9032	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914657	0,916207	0,917736
1,4	0,919243	0,92073	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935745	0,936992	0,93822	0,939429	0,94062	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,95254	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,95907	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,96407	0,964852	0,96562	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971283	0,971933	0,972571	0,973197	0,97381	0,974412	0,975002	0,975581	0,976148	0,976705
2	0,97725	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,98983	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,99224	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,99379	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,99506	0,995201
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,99702	0,99711	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,99825	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3	0,99865	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,99893	0,998965	0,998999
3,1	0,999032	0,999065	0,999096	0,999126	0,999155	0,999184	0,999211	0,999238	0,999264	0,999289
3,2	0,999313	0,999336	0,999359	0,999381	0,999402	0,999423	0,999443	0,999462	0,999481	0,999499
3,3	0,999517	0,999534	0,99955	0,999566	0,999581	0,999596	0,99961	0,999624	0,999638	0,999651
3,4	0,999663	0,999675	0,999687	0,999698	0,999709	0,99972	0,99973	0,99974	0,999749	0,999758
3,5	0,999767	0,999776	0,999784	0,999792	0,9998	0,999807	0,999815	0,999822	0,999828	0,999835

9.3.2 Student’s Distribution

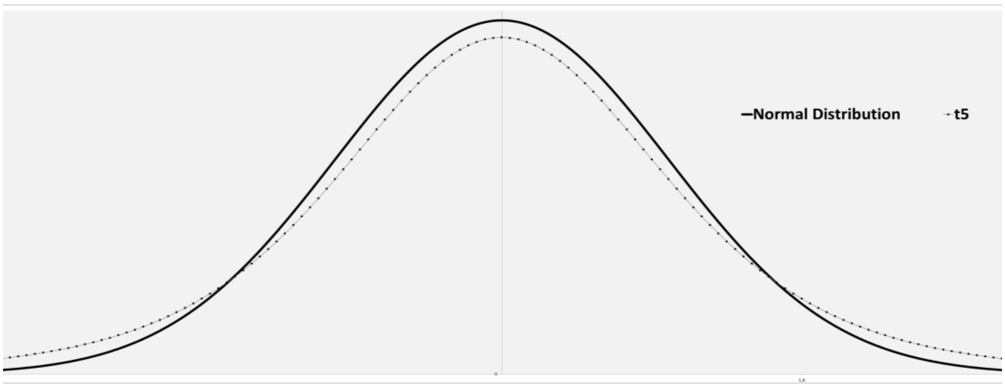


Table 68 - Student's T-Distribution, left tail area for 1 ≤ (n-1) ≤ 30

α Degrees of freedom	0,005	0,01	0,05	0,1	0,15	0,2	0,5	0,8	0,85	0,9	0,95	0,99	0,995
1	-63,657	-31,821	-6,314	-3,078	-1,963	-1,376	0,0	1,376	1,963	3,078	6,314	31,821	63,657
2	-9,925	-6,965	-2,920	-1,886	-1,386	-1,061	0,0	1,061	1,386	1,886	2,920	6,965	9,925
3	-5,841	-4,541	-2,353	-1,638	-1,250	-0,978	0,0	0,978	1,250	1,638	2,353	4,541	5,841
4	-4,604	-3,747	-2,132	-1,533	-1,190	-0,941	0,0	0,941	1,190	1,533	2,132	3,747	4,604
5	-4,032	-3,365	-2,015	-1,476	-1,156	-0,920	0,0	0,920	1,156	1,476	2,015	3,365	4,032
6	-3,707	-3,143	-1,943	-1,440	-1,134	-0,906	0,0	0,906	1,134	1,440	1,943	3,143	3,707
7	-3,499	-2,998	-1,895	-1,415	-1,119	-0,896	0,0	0,896	1,119	1,415	1,895	2,998	3,499
8	-3,355	-2,896	-1,860	-1,397	-1,108	-0,889	0,0	0,889	1,108	1,397	1,860	2,896	3,355
9	-3,250	-2,821	-1,833	-1,383	-1,100	-0,883	0,0	0,883	1,100	1,383	1,833	2,821	3,250
10	-3,169	-2,764	-1,812	-1,372	-1,093	-0,879	0,0	0,879	1,093	1,372	1,812	2,764	3,169
11	-3,106	-2,718	-1,796	-1,363	-1,088	-0,876	0,0	0,876	1,088	1,363	1,796	2,718	3,106
12	-3,055	-2,681	-1,782	-1,356	-1,083	-0,873	0,0	0,873	1,083	1,356	1,782	2,681	3,055
13	-3,012	-2,650	-1,771	-1,350	-1,079	-0,870	0,0	0,870	1,079	1,350	1,771	2,650	3,012
14	-2,977	-2,624	-1,761	-1,345	-1,076	-0,868	0,0	0,868	1,076	1,345	1,761	2,624	2,977
15	-2,947	-2,602	-1,753	-1,341	-1,074	-0,866	0,0	0,866	1,074	1,341	1,753	2,602	2,947
16	-2,921	-2,583	-1,746	-1,337	-1,071	-0,865	0,0	0,865	1,071	1,337	1,746	2,583	2,921
17	-2,898	-2,567	-1,740	-1,333	-1,069	-0,863	0,0	0,863	1,069	1,333	1,740	2,567	2,898
18	-2,878	-2,552	-1,734	-1,330	-1,067	-0,862	0,0	0,862	1,067	1,330	1,734	2,552	2,878
19	-2,861	-2,539	-1,729	-1,328	-1,066	-0,861	0,0	0,861	1,066	1,328	1,729	2,539	2,861
20	-2,845	-2,528	-1,725	-1,325	-1,064	-0,860	0,0	0,860	1,064	1,325	1,725	2,528	2,845
21	-2,831	-2,518	-1,721	-1,323	-1,063	-0,859	0,0	0,859	1,063	1,323	1,721	2,518	2,831
22	-2,819	-2,508	-1,717	-1,321	-1,061	-0,858	0,0	0,858	1,061	1,321	1,717	2,508	2,819
23	-2,807	-2,500	-1,714	-1,319	-1,060	-0,858	0,0	0,858	1,060	1,319	1,714	2,500	2,807
24	-2,797	-2,492	-1,711	-1,318	-1,059	-0,857	0,0	0,857	1,059	1,318	1,711	2,492	2,797
25	-2,787	-2,485	-1,708	-1,316	-1,058	-0,856	0,0	0,856	1,058	1,316	1,708	2,485	2,787
26	-2,779	-2,479	-1,706	-1,315	-1,058	-0,856	0,0	0,856	1,058	1,315	1,706	2,479	2,779
27	-2,771	-2,473	-1,703	-1,314	-1,057	-0,855	0,0	0,855	1,057	1,314	1,703	2,473	2,771
28	-2,763	-2,467	-1,701	-1,313	-1,056	-0,855	0,0	0,855	1,056	1,313	1,701	2,467	2,763
29	-2,756	-2,462	-1,699	-1,311	-1,055	-0,854	0,0	0,854	1,055	1,311	1,699	2,462	2,756
30	-2,750	-2,457	-1,697	-1,310	-1,055	-0,854	0,0	0,854	1,055	1,310	1,697	2,457	2,750

Reading instructions:

For one-tailed tests with a given significance level α:

Refer to the column with $t=a$ for testing if the sample’s mean is *smaller than* population’s mean ($\bar{x}<\mu_0$) and to the column with $t=1-a$ for testing if the sample’s mean is *greater than* the population’s mean ($\bar{x}>\mu_0$)!

For two-tailed tests with a given significance level α: Refer to the column with t=(1-α)/2!

9.3.3 Critical Values for Wilcoxon’s signed-rank-sum-test

Table 69 - Critical values for Wilcoxon's Rank Test for matched pairs

		Significance Level				
	one-tailed	**0,05**	**0,025**	**0,01**	**0,005**	**0,0025**
	two-tailed	**0,1**	**0,05**	**0,02**	**0,01**	**0,005**
n (number of pairs)	**4**	-	-	-	-	-
	5	0	-	-	-	-
	6	2	0	-	-	-
	7	3	2	0	-	-
	8	5	3	1	0	-
	9	8	5	3	1	0
	10	10	8	5	3	1
	11	13	10	7	5	3
	12	17	13	9	7	5
	13	21	17	12	9	7
	14	25	21	15	12	9
	15	30	25	19	15	12
	16	35	29	23	19	15
	17	41	34	27	23	19
	18	47	40	32	27	23
	19	53	46	37	32	27
	20	60	52	43	37	32
	25	100	89	76	68	60
	30	151	137	120	109	98
	35	213	195	173	159	146
	40	286	264	238	220	204
	45	371	343	312	291	272
	50	466	434	397	373	350

Compare Bortz et al. (2000), 729. F

9.3.4 Chi-Square Table

	χ^2										
df	**0,995**	**0,99**	**0,975**	**0,95**	**0,9**	**0,5**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
1	0	0	0	0	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,1	0,21	1,39	4,61	5,99	7,38	9,21	10,6
3	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,3	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,2	5,35	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,7	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19
11	2,6	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,4	5,23	6,3	11,34	18,55	21,03	23,34	26,22	28,3
13	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	4,6	5,23	6,26	7,26	8,55	14,34	22,31	25	27,49	30,58	32,8
16	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,3	28,85	32	34,27
17	5,7	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	18,34	27,2	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40

Larger tables are available online and can also usually be found in the appendix of statistics-books.[526]

Microsoft Excel includes a function called “CHITEST” that can be used instead of the table. SPSS Software is another common tool used to do the test.

---

[526] E.g. Anderson / Finn (1996), 687

9.4 Recommendations for the Evaluation of Tests

The *Education Resources Information Center* clearinghouse on assessment and evaluation (ERICAE) provides a large selection of reading material, worksheets and example tests on their website: http://www.eric.ed.gov/

While their target groups are especially teachers and scientists who do research in education, some of their guidelines are worth reading for design scientists, developing tests and tasks. The ERICAE collection also includes a summary of suggestions and important considerations in *evaluating tests.* The full version is available online at: http://ericae.net/seltips.txt; Worksheet 10 is a reduced copy

Worksheet 10 : Considerations for Test Evaluation (cp. http://ericae.net/seltips.txt)

TEST EVALUATION by Lawrence M. Rudner, ERIC/AE 12/93

You should gather the information you need to evaluate a test.

1) Be sure you have a good idea what you want a test to measure and how you are going to use it.
2) Get a specimen set from the publisher. Be sure it includes technical documentation.
3) Look at reviews prepared by others. The Buros and Pro-Ed Test Locators should help you identify some existing reviews. The MMY also contains references in the professional literature concerning cited tests. The ERIC database can also be used to identify existing reviews.
4) Read the materials and determine for yourself whether the publisher has made a compelling case that the test is valid and appropriate for your intended use.

There are several guidelines to help you evaluate tests.

- The Code of Fair Testing Practices, which is available through this gopher site.
- American Psychological Association (1986) Standards for Educational and Psychological Tests and Manuals. Washington, DC: author
- Equal Employment Opportunity Commission (1978) Uniform Guidelines on Employee Selection Procedures, Federal Register 43, 116, 38295 - 38309.
- Society for Industrial and Organizational Psychology (1987) Principles for the validation and use of personnel selection procedures, Third edition, College Park, MD: author.

In this brief, we identify key standards from the Standards for Educational and Psychological Testing established by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. We describe these standards and questions you may want to raise to evaluate whether the standard has been met.

We discuss standards concerning

A. Test coverage and use

B. Appropriate samples for test validation and norming

C. Reliability

D. Predictive validity

E. Content validity
F. Construct validity
G. Test administration
H. Test reporting
I. Test and item bias

A. Test coverage and use

There must be a clear statement of recommended uses and a description of the population for which the test is intended. The principal question to be asked in evaluating a test is whether it is appropriate for your intended purposes and your students. The use intended by the test developer must be justified by the publisher on technical grounds. You then need to evaluate your intended use against the publisher's intended use and the characteristics of the test.

Questions to ask are:

1. What are the intended uses of the test? What types of interpretations does the publisher feel are appropriate? Are foreseeable inappropriate applications identified?
2. Who is the test designed for? What is the basis for considering whether the test is applicable to your students?

B. Appropriate samples for test validation and norming.

The samples used for test validation and norming must be of adequate size and must be sufficiently representative to substantiate validity statements, to establish appropriate norms, and to support conclusions regarding the use of the instrument for the intended purpose.

The individuals in the norming and validation samples should be representative of the group for which the test is intended in terms of age, experience and background.

Questions to ask are:

1. How were the samples used in pilot testing, validation and norming chosen? Are they representative of the population for which the test is intended? How is this sample related to the your population of students? Were participation rates appropriate? Can you draw meaningful comparisons of your students and these students?
2. Was the number of test-takers large enough to develop stable estimates with minimal fluctuation due to sampling errors? Where statements are made concerning subgroups, is the number of test-takers in each subgroup adequate?
3. Do the difficulty levels of the test and criterion measures (if any) provide an adequate basis for validating and norming the instrument? Are there sufficient variations in test scores?
4. How recent was the norming?

C. Reliability

The test is sufficiently reliable to permit stable estimates of individual ability.

Fundamental to the evaluation of any instrument is the degree to which test scores are free from various sources of measurement error and are consistent from one occasion to another. Sources of measurement error, which include fatigue, nervousness, content sampling, answering mistakes, misinterpretation of instructions , and guessing, will always contribute to an individual's score and lower the reliability of the test.

Different types of reliability estimates should be used to estimate the contributions of different sources of measurement error. Inter-rater reliability coefficients provide estimates of errors dues to inconsistencies in judgement between raters.

Alternate-form reliability coefficients provide estimates of the extent to which individuals can be expected to rank the same on alternate forms of a test. Of primary interest are estimates of internal consistency which account for error due to content sampling, usually the largest single component of measurement error.

Questions to ask are:

1. Have appropriate types of reliability estimates have been computed? Have appropriate statistics been used to compute these estimates? (Split half-reliability coefficients, for example, should not be used with speeded tests as they will produce artificially high estimates.)
2. What are the reliabilities of the test for different groups of test-takers? How were they computed?
3. Is the reliability sufficiently high to warrant the use of the test as a basis for making decisions concerning individual students?

D. Predictive validity

The test adequately predicts academic performance.

In terms of an achievement test, predictive validity refers to the extent to which a test can be appropriately used to draw inferences regarding achievement. Empirical evidence in support of predictive validity must include a comparison of performance on the test being validated against performance on outside criteria.

A variety of measures are available as outside criteria. Grades, class rank, other tests, teacher ratings, and other criteria have been used. Each of these measures, however, have their own limitations.

There are also a variety of ways to demonstrate the relationship between the test being validated and subsequent performance.

Scatterplots, regression equations, and expectancy tables should be provided in addition to correlation coefficients.

Questions to ask are:

1. What criterion measure(s) have been used in evaluating validity? What is the rationale for choosing this measure? Is this criterion measure appropriate?
2. Is the distribution of scores on the criterion measure adequate?
3. What is the basis for the statistics used to demonstrate predictive validity?
4. What is the overall predictive accuracy of the test? How accurate are predictions for individuals whose scores are close to cut-points of interest?

E. Content validity

The test measures content of interest. Content validity refers to the extent to which the test questions are representative of the skills in the specified domain.

Content validity will often be evaluated by an examination of the plan and procedures used in test construction. Did the test development procedure follow a rational approach that ensures appropriate content? Did the process ensure that the collection of items would be representative of appropriate skills?

Questions to ask are:

1. Is there a clear statement of the universe of skills represented by the test? What is the basis for selecting this set of skills? What research was conducted to determine desired test content and/or evaluate it once selected?
2. Were the procedures used to generate test content and items consistent with the test specifications?
3. What was the composition of expert panels used in content validation? What process was used to elicit their judgments?
4. How similar is this content to the content you are interested in testing?

F. Construct validity

The test measures the right psychological constructs.

Construct validity refers to the extent to which a test measures a trait derived from research or experience that have been constructed to explain observable behavior. Intelligence, self-esteem, and creativity are examples of such psychological traits.

Evidence in support of construct validity can take many forms.

One approach is to demonstrate that the items within a measure are inter-related and therefore measure a single construct. Inter-item correlation and factor analysis are often used to demonstrate relationships among the items.

Another approach is to demonstrate that the test behaves as one would expect a measure of the construct to behave. One might expect a measure of creativity to show a greater correlation with a measure of artistic ability than a measure of scholastic achievement would show.

Questions to ask are:

1. Is the conceptual framework for each tested construct clear and well founded? What is the basis for concluding that the construct is related to the purposes of the test?
2. Does the framework provide a basis for testable hypotheses concerning the construct? Are these hypotheses supported by empirical data?

G. Test administration

Detailed and clear instructions outlining appropriate test administration procedures are provided.

Statements concerning the validity of a test for an intended purpose and the accuracy of the norms associated with a test can only generalize to testing situations which replicate the conditions used to establish validity and obtain normative data. Test administrators need detailed and clear instructions in order to replicate these conditions.

All test administration specifications, such as instructions to test takers, time limits, use of reference materials, use of calculators, lighting, equipment, assigning seats, monitoring, room requirements, testing sequence, and time of day, should be fully described.

Questions to ask are:

1. Will test administrators understand precisely what is expected of them?
2. Do the test administration procedures replicate the conditions under which the test was validated and normed? Are these procedures standardized?

H. Test reporting

The methods used to report test results, including scaled scores, subtests results and combined test results, are described fully along with the rationale for each method.

Test results should be presented in a manner that will help schools, teachers and students to make decisions that are consistent with appropriate uses of the test. Help should be available for interpreting and using the test results.

Questions to ask are:

1. How are test results reported to test-takers? Are they clear and consistent with the intended use of the test? Are the scales used in reporting results conducive to proper test use?
2. What materials and resources are available to aid in interpreting test results?

I. Test and item bias

The test is not biased or offensive with regard to race, sex, native language, ethnic origin, geographic region or other factors.

Test developers are expected to exhibit a sensitivity to the demographic characteristics of test-takers, and steps should be taken during test development, validation, standardization, and documentation to minimize the influence of cultural factors on individual test scores. These steps may include the use of individuals to evaluate items for offensiveness and cultural dependency, the use of statistics to identify differential item difficulty, and an examination of predictive validity for different groups.

Tests are not expected to yield equivalent mean scores across population groups. To do so would be to inappropriately assume that all groups have had the same educational and cultural experiences. Rather, tests should yield the same scores and predict the same likelihood of success for individual test-takers of the same ability, regardless of group membership.

Questions to ask are:

1. Were reviews conducted during the test development and validation process to minimize possible bias and offensiveness? How were these reviews conducted? What criteria were used to evaluate the test specifications and/or test items? What was the basis for these criteria?
2. Were the items analyzed statistically for possible bias? What method or methods were used? How were items selected for inclusion in the final version of the test?
3. Was the test analyzed for differential validity across groups? How was this analysis conducted? Does the test predict the same

likelihood of success for individuals of the same ability, regardless of group membership?

4. Was the test analyzed to determine the English language proficiency required of test-takers? Is the English proficiency requirement excessive? Should the test be used with individuals who are not native speakers of English?

9.5 Call for proposals “Innovations against Product Piracy”

The following is an excerpt from the initial system of objectives as provided by the BMBF.[527] The underlined subsection is the core of the system of objectives the project (as described in chapter 5.2.4) aimed at.

Design of products as well as development-, engineering-, and sales-processes which exacerbate plagiary.

- Future products need to be designed to be difficult to copy. Technical solutions are e.g.: Integration of several functions in a module or its interfaces that cannot be opened without its destruction. Product services bundles are more difficult to copy a s a whole.
- Engineering, production and sales must be organized in a way that neither product- nor process knowledge is available for outsiders.
- Track and trace labeling of products and systems
- Methods and processes for economically efficient, copy proof labeling of products and components as originals or even as one of a kind shall be further developed and improved.
- Recognition and information systems must be combined to provide information for preparatory measures and e.g. enforce legal claims.
- Feasibility of track and trace measures that become possible shall be tested in practice. This can be done e.g. by technical service personnel or in surveillance networks (such as customs, trade fairs and so on)
- In cost-benefit analysis, possibilities for exploring new markets with such product surveillance system shall be taken into account.

Design of protection concepts against product piracy

- Instead of single measures, companies should integrate existing technical, organizational and legal measures to comprehensive protection packages.
- For this, strategies, guidelines and analytical methods and technologies for different product classes and branches must be provided. From such o0verall protection concepts, companies must be able to derive individual protection strategies.

Funding is also available for production technologies and machinery that allow for effective protection against product piracy and are evaluated in a complete process chain.

---

[527] Translated from http://www.bmbf.de/en/furtherance/6669.php; 5/513, 15:03

10 Epilogue - Scurvy, an early Experiment

Whenever I discussed the topic of evaluation and validation with fellow PhDs, some of them tend to be scared that demanding more rigor in design science and advising the research community on validation, might provoke the perception that research which hasn't been validated yet might be invalid or of lesser value. Not always having the chance and resources for a full validation might then lead researchers not to pursue their ideas hence leading to possibly valuable discoveries never being made. I would like to finish this thesis with explaining my opinion on this sensitive topic through a short, interesting yet true story:

It is set shortly after Galileo Galilei was forced to lay aside his telescope and had to withdraw his scientific theory of the earth. He was forced to admit to living on a flat disc in the centre of the universe in order to avoid being burned to death. Other scientists of that time explored not the universe but the world's oceans. They were not so much afraid to be burned at the stake, but feared a mysterious disease they would catch out at sea. Ever since sailors started to set out to sea for long journeys, after about three months, their teeth would start to fall out, their gums would rot, they were permanently tired and their muscles were sore, all accompanied by strange stains on their skin. For several hundred years, captains and on-board doctors would observe the ever repeating, same scary disease that was more fearsome than pirates and hurricanes: The scurvy.[528]

According to PROFESSOR JONATHAN LAMB, it all began when man had started to set sail for such long journeys, penetrating the Indian and the Pacific Oceans. Vasco da Gama lost two thirds of his crew to the disease while making his way to India in 1499. In 1520 Magellan lost more than 80 per cent while crossing the Pacific. It should last for several hundred years. During the 18th century, more British sailors lost their lives to scurvy than to their enemies. Admiral of the Fleet George Anson, in one of his documented voyages around 1740, lost 1300 out of his 2000 sailors within the first ten months.[529] According to the Royal Navy's documentation, it had drafted 184,899 sailors for the Seven Years' War. In the end, 133,708 were reported to have died of disease or were "missing". Scurvy was the principal disease at the time.[530]

The end of scurvy came with Captain James Cook and his doctor and companion Dr. James Lind. Lind came up with the hypothesis that the disease was related to a sailor's diet.[531] They tried to evaluate their theory systematically.

---

[528] In Latin: *scorbutus*

[529] See also: http://www.bbc.co.uk/history/british/empire_seapower/captaincook_scurvy_01.shtml; 7/28/13, 12:38

[530] Cp. Turberville (2006)

[531] It is not proven whether other captains might have had similar hypotheses on scurvy. Some argue that grog, a mixture of rum and water, might have been an attempt to fight scurvy. Others argue that it was simply invented to help sailors drink foul water without too much trouble. However, these speculations are not documented. Lind's hypotheses and test are deemed to be the first clinical trial in history.

LIND describes, a study of twelve sailors, suffering from symptoms of scurvy.[532] He put them in six groups of two individuals each. All of them were put on the exact same diet and were given additional, varying oral treatment: apple wine, diluted sulfuric acid, vinegar, saltwater, oranges and lemons, a gargle solution on an herbal basis.

"After only six days he noticed that with the two men that had been given lemons and oranges their gums started to heal and the stains on their skin started to disappear."[533] Captain James Cook from this day on would carry tons of lemons, oranges and sauerkraut and became the first captain to conduct a great overseas journey without losing a single man. The sour-tasting ingredient that seemed to work so well protecting sailors from „scorbutus“ was given the name „Ascorbic Acid“ – derived from Latin "anti scorbutus“ meaning against scurvy. Today, Vitamin C – well known to be rather important for the human diet – still carries the official name Ascorbic Acid. The captain's diet was later applied to the whole British Navy.[534]

Considering the great success of Dr. Lind's findings and the many thousands of lives he saved with his results one is inclined to critically ask:

Are scientific rigor and thorough verification with lots of empirical evidence based on generally large numbers of data always the right procedure? What would have happened to the British Navy if it had done the math and ignored Lind's findings due to limited confidence? Would it have been right to go on the next journey and repeat the experiment with larger numbers of dying sailors?

Of course, science could only proceed to build the (valid) knowledge we can draw from today through continuous *questioning of the results it has produced*. Our knowledge, the progress of technology and hence our wealth are a result of this foundationalist scientific world. As the physicist and winner of the Nobel Price RICHARD P. FEYNMAN put it:

> *"We absolutely must leave room for doubt or there is no progress and there is no learning. There is no learning without having to pose a question. And a question requires doubt."*
>
> Richard P. Feynman

Maybe – even in a foundationalist world – a study that does not satisfy all criteria of reliability and validity is not *so much wrong* or of no value *as* it might simply be *unfinished*. Nevertheless, researchers have to constantly deal with their critics and peers stating that their work is not sufficiently validated and therefore shouldn't be published yet. The next time this happens to you, maybe you will tell them the story of James Lind.

Leif Marxen, October 2013

---

[532] Lind (1757)

[533] Tröhler (2003)

[534] This took however about 50 years! The Royal Navy they did not want to carry only the lemon's juice which did not work like fresh lemons did due to Vitamin C's fast oxidization

Curriculum Vitae

Personal Data

Name: Leif Marxen
Date of Birth: 5 December 1980
Place of Birth: Heidelberg
Nationality: German
Marital Status: married

Education

1987 – 1991	Kurpfalzschule Dossenheim (primary school)
1991 – 2000	Kurpfalz-Gymnasium Schriesheim (Graduation: Allgemeine Hochschulreife)
2000 – 2007	Studies in mechanical engineering specialized in Product Development and Design at the University of Karlsruhe (TH) (Degree: Diplomingenieur)

Professional Experience

2003 – 2004	Student Assistant at IPEK – Institute of Product Engineering, Karlsruhe Institute of Technology (KIT)
2005 – 2006	Intern at Detroit Diesel Corporation, Detroit, MI, USA
2007 – 2012	Research Associate at IPEK – Institute of Product Engineering, Karlsruhe Institute of Technology (KIT) 01/2008 associate Team Manager of the research group Design Methods and Design Management 01/2009 Team Manager of the research group Design Methods and Design Management
2013 – 2014	Scholarship Holder of the EXIST Business Start-Up Grant, funded by the German Federal Ministry of Economics and Technology (BMWi) In association with IPEK – Institute of Product Engineering, Karlsruhe Institute of Technology (KIT)

Lebenslauf

Persönliche Daten

Name:	Leif Marxen
Geburtsdatum:	5. Dezember 1980
Geburtsort:	Heidelberg
Staatsangehörigkeit:	Deutsch
Familienstand:	verheiratet

Bildungsgang

1987 – 1991	Kurpfalzschule Dossenheim (Grundschule)
1991 – 2000	Kurpfalz-Gymnasium Schriesheim (Abschluss: Allgemeine Hochschulreife)
2000 – 2007	Studium des Maschinenbaus mit Vertiefungsrichtung Produktentwicklung und Konstruktion an der Universität Karlsruhe (TH) (Abschluss: Diplom)

Berufstätigkeit

2003 – 2004	Wissenschaftliche Hilfskraft am IPEK – Institut für Produktentwicklung, Karlsruher Institut für Technologie (KIT)	
2005 – 2006	Praktikant bei der Detroit Diesel Corporation, Detroit, MI, USA	
2007 – 2012	Wissenschaftlicher Mitarbeiter am IPEK – Institut für Produktentwicklung, Karlsruher Institut für Technologie (KIT)	
	01/2008	stellvertretender Forschungsgruppenleiter Entwicklungsmethodik und -management
	01/2009	Forschungsgruppenleiter Entwicklungsmethodik und -management
2013 – 2014	Stipendiat im EXIST Gründerstipendium, des Bundesministeriums für Wirtschaft und Technologie (BMWi) am IPEK - Institut für Produktentwicklung, Karlsruher Institut für Technologie (KIT)	